Generally Shifting Coefficients in Interactive Effects Panel Data Models and the COVID–19 Regime*

Yousef Kaddoura†
Örebro University

Joakim Westerlund
Lund University
and
Deakin University

August 5, 2025

Abstract

The present paper proposes a new approach to panel data regression models with interactive effects that is appropriate when the slope coefficients are not necessarily constant but that they may be shifting over time. A major point about the new methodology—referred to as "SHIFTIE"—is that the type of shifts that can be accommodated is very broad and includes most existing specifications in the literature, such as structural breaks and Markov switching, as special cases. This generality is important from a practical point of view, as existing approaches require empirical researchers to take a stance as to the shift-generating process, which is typically unknown. The asymptotic validity of SHIFTIE is established and verified in small samples using Monte Carlo simulations. The empirical implementation is illustrated using as an example economic growth. The results confirm the common heuristic belief that the COVID–19 pandemic was special in many ways.

*Westerlund would like to thank the Knut and Alice Wallenberg Foundation for financial support through a Wallenberg Academy Fellowship.

†Corresponding author: Örebro University School of Business, SE-701 82 Örebro, Sweden. E-mail address: yousef.kaddoura@oru.se.

Keywords: Panel data; interactive effects; time-varying parameter models; structural break; Markov switching; COVID–19.

1 Introduction

Accounting for temporal shifts in the coefficients of otherwise linear regression models has long been important in economics and elsewhere. The urgency of addressing this issue has grown markedly in recent years due to a succession of extreme events—including wars, pandemics, trade conflicts, financial crises, natural disasters, and major policy changes—that continue to unfold.[1] Panel data are especially vulnerable to such shocks because they contain many individual time series. However, only recently have methods designed to handle shifts in panel data become available to practitioners.

Most existing techniques are limited to models with a single "structural break," that is, two regimes separated by a single breakpoint at which the parameters shift permanently (see, for example, Antoch et al., 2019, Baltagi et al., 2016, Karavias et al., 2022, and Zhu et al., 2020). Such single structural break models are appropriate in some situations, as in the case of a monetary policy regime change or a technological breakthrough, but certainly not in all. The methods proposed by Qian and Su (2016), Boldea et al. (2020), Kaddoura and Westerlund (2023), and Ditzen et al. (2025) allow for multiple structural breaks, and are therefore more general in this regard. However, they still assume that breaks are infrequent and that their effects are permanent. These methods therefore fail to accommodate situations in which events are recurrent—such as business cycles, bushfires, or storms—or where the impact is transitory, as with episodes of extreme weather, or financial crises. In fact, switches between regimes may be frequent, as in the

[1] A non-exhaustive list of recent examples includes the 2007–2008 global financial crisis, the subprime mortgage crisis of 2007–2010, the COVID-19 pandemic, Russia's invasion of Ukraine in 2022, and the 2021–2023 energy crisis.

case of rapidly changing financial market conditions (see Piger, 2009, for a discussion of different types of breaks).

Observations like the ones just made have led to the development of a separate literature on "regime switching." Regime switching models can be broadly categorized into two types; "threshold" models and "Markov-switching" models, both of which generate frequent shifts with transitory effects. The key difference between the two lies in the mechanism governing the shifts. Threshold models assume that regime changes are triggered by the level of an observed variable relative to an unobserved threshold. In contrast, Markov-switching models posit that regime shifts follow a latent Markov chain, often with only two states and time-invariant transition probabilities. Each approach has its strengths and weaknesses. Threshold models are relatively easy to estimate and interpret; however, they rely on the assumption that the relevant state variable is observed—an assumption that is seldom satisfied in practice. Markov-switching models relax this requirement, but do so at the expense of imposing strong parametric assumptions and requiring more complex estimation techniques. Interestingly, while the time series strand of the regime switching literature is huge (see Hamilton, 2016, for a recent overview), the panel strand is almost nonexistent (see Cheng et al., 2010, and Miao et al., 2020, for discussions).

The above discussion suggests that if the panel literature on structural breaks is small, the literature on regime switching is even smaller, and there have been no attempts to bridge the two.[2] This point is important not only in its own right but also because of its implications for empirical work. In particular, since there is currently no holistic approach, researchers are forced to take a stance as to whether shifts are in-

[2]This observation applies to the time series literature as well. In fact, the structural breaks and regime switching literatures appear to have developed largely independently, despite the close relationship between the two phenomena. Reflecting this, the otherwise comprehensive surveys by Hamilton (2016) and Perron (2006) hardly mention this issue.

frequent with permanent effects (structural breaks) or frequent with transitory effects (regime switches). The present paper can be seen as a reaction to this. Its aim is to develop a new approach to panel data models that is general enough to accommodate both phenomena, henceforth referred to as "shifts."

Dealing with these shifts would have been easier were it not for the simultaneous presence of unobserved heterogeneity due to omitted variables. This is a major concern because models in economics tend to be very parsimonious and are, therefore, likely to omit variables that may be correlated with the regressors, rendering them endogenous. To address this problem, researchers commonly assume that the unobserved heterogeneity consists of additive unit- and time-specific constants, or "fixed effects." However, unlikely to be enough in many applications, which has recently led to the consideration of more general interactive effects models. In these models, the time- and unit-specific effects—often called "factors" and "loadings," respectively—enter multiplicatively. The most common approach to handling such effects is "de-factoring," which involves projecting the data onto the space orthogonal to the estimated factors. This procedure eliminates the interactive effects asymptotically; however, it does so in an awkward way, as the shifting slope coefficients in the model for the de-factored data are not the same as those for the raw data.

We address this challenge by "de-loading" the data instead of de-factoring them. Unlike the factors, the loadings are time-invariant, which means that we can project them out without affecting the slopes. With the interactive effects removed, we identify shifts by adapting an existing algorithm from the literature on unknown group structures to estimate shifts along the time dimension. The groups in this context are regimes, and the approach allocates each time period to one of these regimes optimally using a least squares (LS) criterion. Time periods within the same regime share the same coefficients, whereas periods in different regimes have different coefficients. Moreover, because the periods allocated to a particular regime do not need to be consecutive, shifts may occur frequently. Both the timing of the shifts and the regime-specific slope coefficients are treated as non-random, meaning that—unlike in much of the regime-switching literature—no restrictive distributional assumptions are required. The new approach, henceforth referred to as "SHIFTIE" (SHIFTs in Interactive Effects models), is therefore very general in terms of the types of shifts it can accommodate.

The rest of the paper is organized as follows. In Section 5.2, we present the model and the proposed SHIFTIE approach used to estimate it. Sections 3 and 5.2 examine the asymptotic and finite-sample properties of the approach, respectively. Section 5 reports the results of an empirical application that showcases the importance of COVID-19 for models of economic growth. Section 6 concludes. All proofs are provided in the appendix.

2 Model and approach

Consider a scalar panel variable $y_{i,t}$, observed over $t = 1, \ldots, T$ time periods and $i = 1, \ldots, N$ cross-sectional units. The data-generating process (DGP) that we consider can be viewed as a prototypical interactive effects model, with the added feature that the slope coefficients are allowed to shift over time. It is given by

$$y_{i,t} = \boldsymbol{\beta}_{r_t}' \mathbf{x}_{i,t} + \boldsymbol{\gamma}_t' \mathbf{f}_i + \varepsilon_{i,t}, \tag{2.1}$$

$$\mathbf{x}_{i,t} = \boldsymbol{\Gamma}_t' \mathbf{f}_i + \mathbf{v}_{i,t}, \tag{2.2}$$

where $\mathbf{x}_{i,t} := [x_{1,i,t}, \ldots, x_{k,i,t}]'$ and $\boldsymbol{\beta}_{r_t} := [\beta_{1,r_t}, \ldots, \beta_{k,r_t}]'$ are $k \times 1$ vectors of known regressors and unknown slope coefficients, respectively. The slope coefficients are allowed to vary over time via the variable $r_t \in \mathcal{R} := \{1, \ldots, R\}$, which assigns each time period t into one of R regimes. Moreover, $\gamma_t := [\gamma_{1,t}, \ldots, \gamma_{m,t}]'$ and $\mathbf{\Gamma}_t := [\mathbf{\Gamma}_{1,t}, \ldots, \mathbf{\Gamma}_{m,t}]'$ are $m \times 1$ and $m \times k$ matrices, respectively, of unknown common factors with $\mathbf{f}_i := [f_{1,i}, \ldots, f_{m,i}]'$ being the associated $m \times 1$ vector of factor loadings, and $\varepsilon_{i,t}$ and $\mathbf{v}_{i,t}$ are idiosyncratic error terms.

The products $\gamma_t'\mathbf{f}_i$ and $\mathbf{\Gamma}_t'\mathbf{f}_i$ represent the interactive effects. As noted in Section 1, these effects appear in both (2.1) and (2.2), which makes $\mathbf{x}_{i,t}$ endogenous. Unlike in studies such as Pesaran (2006), this endogeneity arises from the loadings rather than the factors themselves (see Hansen and Liao, 2019, for a similar assumption).[3]

It is useful to write the above equations in stacked vector form as follows:

$$\mathbf{y}_t = \mathbf{X}_t\boldsymbol{\beta}_{r_t} + \mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t, \tag{2.3}$$

$$\mathbf{X}_t = \mathbf{F}\mathbf{\Gamma}_t + \mathbf{V}_t, \tag{2.4}$$

where $\mathbf{y}_t := [y_{1,t}, \ldots, y_{N,t}]'$ and $\boldsymbol{\varepsilon}_t := [\varepsilon_{1,t}, \ldots, \varepsilon_{N,t}]'$ are $N \times 1$, $\mathbf{X}_t := [\mathbf{x}_{1,t}, \ldots, \mathbf{x}_{N,t}]'$ and $\mathbf{V}_t := [\mathbf{v}_{1,t}, \ldots, \mathbf{v}_{N,t}]'$ are $N \times k$, and $\mathbf{F} := [\mathbf{f}_1, \ldots, \mathbf{f}_N]'$ is $N \times m$.

The model in (2.3) involves four types of parameters: the vector of regime-specific slope coefficients $\boldsymbol{\beta}_{r_t}$, the regime membership variable r_t, the vector of common factors γ_t, and the matrix of loadings $\mathbf{F}$. It is convenient to organize the first three into the $Rk \times 1$ vector $\boldsymbol{\beta} := [\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_R']'$, the $m \times 1$ vector $\mathbf{r} := [r_1, \ldots, r_T]'$, and the $T \times m$ matrix $\boldsymbol{\gamma} := [\gamma_1, \ldots, \gamma_T]'$. While $\boldsymbol{\gamma}$ and $\mathbf{F}$ are not of primary interest, it is necessary to control for them; otherwise, $\boldsymbol{\beta}$ and $\mathbf{r}$ would be inconsistently estimated due to the omitted variables problem. One approach is to follow the existing literature on grouped cross-sectional heterogeneity and estimate all four sets of parameters jointly (see, for example, Ando and Bai, 2016). However, joint estimation in this high-dimensional panel data setting is computationally burdensome. We therefore seek a stepwise procedure. An important

[3]Although the same set of loadings $\mathbf{f}_i$ appears in both (2.1) and (2.2), some of the elements of γ_t and $\mathbf{\Gamma}_t$ may be zero. As a result, $y_{i,t}$ and $\mathbf{x}_{i,t}$ are allowed to depend on different loadings.

observation in this regard is that while estimates of $\boldsymbol{\beta}$ and $\mathbf{r}$ necessarily depend on one another, the problems of estimating $(\boldsymbol{\beta}, \mathbf{r})$, on the one hand, and $(\gamma, \mathbf{F})$, on the other hand, are in principle separable.

As a source of inspiration, we consider the studies by Baltagi et al. (2016), Karavias et al. (2023), and Ditzen et al. (2025), which incorporate structural breaks and interactive effects through variations of the common correlated effects (CCE) approach of Pesaran (2006). CCE uses cross-sectional averages of the observables as proxies for the factors, which are then projected out in a regime-wise manner. While this de-factoring addresses the interactive effects, it introduces a dependency on the unknown break dates due to its regime-wise implementation. As a result, this approach does not achieve a clear separation between breaks and interactive effects, instead making all estimates interdependent—something we aim to avoid.

Our proposed SHIFTIE solution is inspired by Moon and Perron (2004). Although their context—unit root testing—differs from ours and they do not consider the CCE approach, they make a key observation: de-factoring with non-stationary factors is more complex than de-loading with non-random factor loadings. Since the goal is to eliminate interactive effects rather than estimate them, de-loading is sufficient. We face a similar situation in the sense that de-factoring is comparatively more complicated. In particular, unlike the factors, the loadings are time-invariant, which allows de-loading to be performed without any knowledge of the regime memberships.

To formalize our proposed solution, let $\mathbf{Z}_t := [\mathbf{y}_t, \mathbf{X}_t]$ be the $N \times (k+1)$ matrix of observables. By combining (2.3) and (2.4), the models for $\mathbf{y}_t$ and $\mathbf{X}_t$ can be rewritten as the following static factor model for $\mathbf{Z}_t$:

$$\mathbf{Z}_t = \mathbf{F}\mathbf{C}_t + \mathbf{U}_t, \tag{2.5}$$

where $\mathbf{C}_t := [\boldsymbol{\Gamma}_t\boldsymbol{\beta}_{r_t} + \boldsymbol{\gamma}_t, \boldsymbol{\Gamma}_t]$ is $m \times (k+1)$ and $\mathbf{U}_t := [\boldsymbol{\varepsilon}_t + \mathbf{V}_t\boldsymbol{\beta}_{r_t}, \mathbf{V}_t]$ is $N \times (k+1)$. In this notation, the proposed estimator $\widehat{\mathbf{F}}$ of (the space spanned by) $\mathbf{F}$ is given simply by

$$\widehat{\mathbf{F}} := \overline{\mathbf{Z}} = \mathbf{F}\overline{\mathbf{C}} + \overline{\mathbf{U}}, \tag{2.6}$$

where $\overline{\mathbf{A}} := T^{-1}\sum_{t=1}^{T}\mathbf{A}_t$ for any generic matrix $\mathbf{A}_t$. The justification for this estimator lies in the fact that under certain regulatory conditions laid out in Section 3, $\overline{\mathbf{U}}$ negligible. Hence, $\widehat{\mathbf{F}}$ is consistent for $\mathbf{F}\overline{\mathbf{C}}$, which under the conditions of Section 3 implies that pre-multiplication by the $N \times N$ projection error matrix $\mathbf{M}_{\widehat{F}} := \mathbf{I}_N - \widehat{\mathbf{F}}(\widehat{\mathbf{F}}'\widehat{\mathbf{F}})^{-1}\widehat{\mathbf{F}}'$ will eliminate the interactive effects asymptotically.

To estimate $(\boldsymbol{\beta}, \mathbf{r})$, we propose an algorithm inspired by the grouped fixed effects (GFE) approach of Bonhomme and Manresa (2015) to the estimation of latent cross-sectional group structures in fixed effects models. In particular, the idea is to estimate $\boldsymbol{\beta}$ and $\mathbf{r}$ jointly by minimizing the residual sum of squares of the de-loaded data;

$$(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{r}}) := \underset{(\boldsymbol{\beta},\mathbf{r})\in\mathcal{B}^R\times\mathcal{R}^T}{\arg\min} \sum_{t=1}^{T} \left\|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{r_t})\right\|^2, \tag{2.7}$$

where $\mathcal{B} \in \mathbb{R}^{k\times 1}$ is the parameter space of $\boldsymbol{\beta}_{r_t}$ and $\|\mathbf{A}\| := \sqrt{\operatorname{tr}(\mathbf{A}'\mathbf{A})}$ is the Frobenius norm of any matrix $\mathbf{A}$. The above joint estimation problem can be solved iteratively. Given $\mathbf{r}$, we can estimate $\boldsymbol{\beta}$, and given $\boldsymbol{\beta}$, we can reestimate $\mathbf{r}$. This process continues until convergence. The below estimation algorithm provides the details. Here and throughout this paper, $\mathbf{r}^{(s)} := [r_1^{(s)}, \ldots, r_T^{(s)}]'$ and $\boldsymbol{\beta}^{(s)} := [\boldsymbol{\beta}_1^{(s)\prime}, \ldots, \boldsymbol{\beta}_R^{(s)\prime}]'$ denote estimates of $\mathbf{r}$ and $\boldsymbol{\beta}$, respectively, at step s.

Estimation algorithm.

1. Set $s = 0$. Initiate $\mathbf{r}$ for a given R. Let $\mathbf{r}^{(0)}$ be this initial value.

2. Given $\mathbf{r}^{(s)}$, compute

$$\boldsymbol{\beta}^{(s)} := \arg\min_{\beta \in \mathcal{B}^R} \sum_{t=1}^{T} \left\| \mathbf{M}_{\widehat{F}} \left(\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_{r_t^{(s)}} \right) \right\|^2. \tag{2.8}$$

3. Given $\boldsymbol{\beta}^{(s)}$, update $\mathbf{r}$ as $\mathbf{r}^{(s+1)} = \left[r_1^{(s+1)}, \ldots, r_R^{(s+1)} \right]'$, where

$$r_t^{(s+1)} := \arg\min_{r_t \in \mathcal{R}} \left\| \mathbf{M}_{\widehat{F}} \left(\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_{r_t}^{(s)} \right) \right\|^2. \tag{2.9}$$

4. Set $s = s + 1$ and return to step 2. Repeat until numerical convergence. The solution $(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{r}})$ to (2.7) is given by the final step estimates.

According to our Monte Carlo simulations, convergence is very fast, partly because the objective function is non-increasing with each iteration. However, convergence to the global optimum is not guaranteed—a common limitation of this type of iterative algorithm. Therefore, multiple initializations of $\mathbf{r}$ should be tested, and the one yielding the lowest value of the objective function should be selected.

The GFE approach can be viewed as a variant of the classical k-means clustering algorithm, and serves as the workhorse of the literature on unknown groups (see, for example, Loyo and Boot, 2025, and Miao et al., 2020). However, it is not the only available method. One alternative is the expectation–maximization (EM) algorithm used by Ke et al. (2016) to maximize the full likelihood function. However, this approach is computationally costly and challenging to implement—likely one reason for its limited adoption in applied work. Additionally, it assumes that the slope coefficients are pre-ordered, which in practice requires sequencing the data. This is manageable with a single regressor but it is unclear how to proceed when multiple regressors are present. Another alternative to GFE is the penalized principal component (PPC) approach of Su and Ju (2018), which uses the principal components-based approach of Bai (2009) to deal with the interactive effects and the least absolute shrinkage and selection operator (LASSO) to deal with the groups. However, PPC involves solving a high-dimensional, non-convex optimization problem and requires grid search over multiple tuning parameters, including the number of factors m. Thus, from a computational standpoint, PPC is no more practical than EM.

3 Asymptotic analysis

To facilitate our asymptotic analysis, we need to introduce some notation. Specifically, if $\mathbf{A}$ is a matrix, $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ signify its smallest and largest eigenvalues, respectively, while $\text{rank}(\mathbf{A})$ signifies its rank. The symbols $\to_p$ and $\to_d$ signify convergence in probability and distribution, respectively. The indicator function for the event A is denoted $\mathbb{1}(A)$. It takes on the value 1 if A is true and 0 otherwise. We use w.p.1 (w.p.a.1) to denote with probability (approaching) 1. As usual, true parameter values are superscripted by a "0," such that, for example, r_t^0 is the true value of r_t.

Assumption 1.

(a) $\mathcal{B}$ is a compact subset of $\mathbb{R}^{k\times 1}$;

(b) $\boldsymbol{\beta}_{r_t}$, $\boldsymbol{\Gamma}_t$ and γ_t are non-random such that $\text{rank}(\overline{\mathbf{C}}) = m \leq k+1$. Also, $\overline{\mathbf{C}} \to \mathbf{C}$ as $T \to \infty$, where $\text{rank}(\mathbf{C}) = m \leq k+1$ and $\|\mathbf{C}\| < \infty$;

(c) $\mathbf{w}_{i,t} := [\varepsilon_{i,t}, \mathbf{v}_{i,t}]'$ as a covariance stationary process with absolutely summable autocovariances, $\mathbb{E}(\mathbf{w}_{i,t}) = \mathbf{0}_{(k+1)\times 1}$, $\sup_{i,t} \mathbb{E}(\|\mathbf{w}_{i,t}\|^4) < \infty$ and $\sup_t N^{-1} \sum_{i=1}^N \sum_{j=1}^N |\text{tr}\,[\mathbb{E}(\mathbf{w}_{i,t}\mathbf{w}_{j,t}')]| < \infty$;

(d) $\varepsilon_{i,t}$ and $\mathbf{v}_{j,s}$ are independent for all i, t, j and s;

(e) $\mathbf{f}_i$ is non-random such that $\sup_i \|\mathbf{f}_i\| < \infty$ and $N^{-1}\mathbf{F}'\mathbf{F} \to \mathbf{\Sigma}_F$ as $N \to \infty$, where $\mathbf{\Sigma}_F$ is positive definite;

(f) $(NT)^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{E}(\mathbf{U}_t'\mathbf{U}_s) \to \mathbf{\Omega}_u$ as N, $T \to \infty$, where $\mathbf{\Omega}_u$ is positive definite;

(g) $\max_{r' \in \mathcal{R}} \lambda_{\min}(\mathbf{S}_{r',r}) \geq \rho_{\min} > 0$ for all $r \in \mathcal{R}$ w.p.a.1, where $\mathbf{S}_{r',r} := (NT)^{-1} \sum_{t=1}^{T} \mathbb{1}(r_t' = r')\mathbb{1}(r_t = r)\mathbf{V}_t'\mathbf{M}_{F^0}\mathbf{V}_t$, and $\mathbf{F}^0$ is the limit of $\widehat{\mathbf{F}}$ in the sense of Lemma A.1 in the appendix;

(h) $T^{-1}T_r \to \pi_r > 0$ as $T \to \infty$ for all $r \in \mathcal{R}$, where $T_r := \sum_{t=1}^{T} \mathbb{1}(r_t = r)$;

(i) $\|\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0\| > 0$ for all $r' \neq r$.

Some remarks are in order. Assumption 1(a) requires that the parameter space for $\boldsymbol{\beta}_{r_t}$ is compact—a standard condition for extremum estimators. Unlike many other studies on interactive effects (see, for example, Pesaran, 2006, and Kaddoura and Westerlund, 2025), Assumptions 1(b) and 1(c) do not require $\mathbf{\Gamma}_t$, γ_t, or $\mathbf{f}_i$ to be randomly distributed. Instead, these matrices are treated as non-random, which is a more general consideration, since it is non-parametric. Random loadings can still be accommodated by interpreting $\mathbf{f}_i$ as a single realization from the underlying random loading process, and the same applies to random factors. Treating $\boldsymbol{\beta}_{r_t}$ as non-random in Assumption 1(b) is also a strength, especially when compared to the Markov switching literature, which—as discussed in Section 1—typically assumes that r_t evolves according to a Markov chain. Assumption 1(b) further requires that $\overline{\mathbf{C}}$ has full row rank. This condition ensures that the m columns of $\mathbf{F}$ can be recovered asymptotically from the $k+1$ columns of $\widehat{\mathbf{F}}$, and is analogous to the standard CCE rank condition on the cross-sectional averages of the loadings. It is a mild restriction, requiring only that the number of loadings is not underspecified. This contrasts with much of the interactive effects literature, which assumes that the number of loadings is known or can be estimated accurately (see Westerlund and Urbain, 2015, for a discussion).[4]

Assumption 1(c) allows for weak serial and cross-sectional dependence in $\varepsilon_{i,t}$ and $\mathbf{v}_{i,t}$, and is quite general. This is sufficient for some of our results but not all. Later on, at-most-weak dependence will need to be strengthened to independence—an expected requirement, as many studies on unknown groups assume independence in at least one of the two dimensions (see, for example, Miao et al., 2020). Assumption 1(d) rules out lagged dependent variables in $\mathbf{x}_{i,t}$. However, as noted above, $y_{i,t}$ is still allowed to exhibit serial correlation through both γ_t and $\varepsilon_{i,t}$. Assumption 1(d) also does not preclude endogeneity, as $\mathbf{x}_{i,t}$ may still be correlated with the regression errors in (2.1) via $\mathbf{f}_i$.

Assumptions 1(e) and 1(g) are standard non-collinearity conditions that ensure identification of $\boldsymbol{\beta}_r^0$. They require that the loadings are not collinear and that the regressors exhibit sufficient variation across both the cross-section and time, after projecting out all variation that can be explained by the loadings (see, for example, Bai, 2009). The latter condition is assumed to hold over the intersection of time periods assigned to all estimated and true regimes. Note also that for Assumption 1(g) to be satisfied, these intersections cannot be negligible relative to T. This requirement is formalized in Assumption 1(h), which is standard in the structural break literature (see, for example, Bonhomme and Manresa, 2015, and Loyo and Boot, 2025). However, while in that literature Assumption 1(h) is essential for accurate estimation, according to our Monte Carlo results this is not the case here. Assumption 1(f) requires that the average long-run covariance matrix of $\mathbf{U}_t$ is positive definite. This implies that $y_{i,t}$ and $\mathbf{x}_{i,t}$ cannot be over-differenced. Finally, Assumption 1(i)—commonly referred to as the regime "sepa-

[4]One way to relax Assumption 1(b) is to include time-invariant regressors that can be treated as observed loadings. These observed loadings may be appended to $\widehat{\mathbf{F}}$ and projected out, thereby removing the need for them to satisfy the rank condition in Assumption 1(b).

ration condition"—requires that $\boldsymbol{\beta}^0_{r'}$ and $\boldsymbol{\beta}^0_r$ are distinct.

With the above conditions in place, we can establish the consistency result in Theorem 1 below. The theorem is stated in terms of the following Hausdorff distance:

$$d_H(\mathbf{a}, \mathbf{b}) := \max\left\{\max_{r\in\mathcal{R}}\min_{r'\in\mathcal{R}}\|\mathbf{a}_r - \mathbf{b}_{r'}\|,\ \max_{r'\in\mathcal{R}}\min_{r\in\mathcal{R}}\|\mathbf{a}_r - \mathbf{b}_{r'}\|\right\}, \tag{3.1}$$

where $(\mathbf{a}, \mathbf{b}) \in \mathcal{B}^R \times \mathcal{B}^R$.[5] The reason for considering this distance is that our objective function is invariant to a relabeling of the regime index r and the regime-specific slopes $\boldsymbol{\beta}_r$ are therefore only identified up to a re-ordering of the regimes (see Bonhomme and Manresa, 2015, for a discussion). By showing that $d_H(\boldsymbol{\beta}^0, \widehat{\boldsymbol{\beta}}) = o_p(1)$ Theorem 1 establishes that the identified set is consistently estimated.

Theorem 1. *Suppose that Assumption 1 holds. Then, as $T \to \infty$ with $N < \infty$ or $N \to \infty$,*

$$d_H\left(\boldsymbol{\beta}^0, \widehat{\boldsymbol{\beta}}\right) = o_p(1).$$

Theorem 1 implies that the elements of $\widehat{\boldsymbol{\beta}}$ can be relabelled so that $\|\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}^0_r\| = o_p(1)$ for all $r \in \mathcal{R}$ (see the appendix for details). Hence, in what follows, we simply assume that the elements of $\widehat{\boldsymbol{\beta}}$ have been suitably relabelled, as in Bonhomme and Manresa (2015). Theorem 1 does not require that the regime membership r^0_t is consistently estimated for all time periods t, which is only possible if $N \to \infty$. This follows from the fact that $\widehat{r}_t$ is constructed using only cross-sectional variation. By not requiring consistent estimation of r^0_t it is possible to have N fixed. This allowance stands out when compared to existing results in the group literature, which typically require both N and T to grow large (see, for example, Bonhomme and Manresa, 2015; Loyo and Boot, 2025; and Miao et al., 2020).

[5]The exact definition of $d_H(\cdot,\cdot)$ depends on the problem at hand and hence varies from study to study. Our definition is the same as in Miao et al. (2020).

The $o_p(1)$ convergence rate in Theorem 1 is clearly not sharp. It can be sharpened at the expense of more restrictive conditions. This is done in Theorem 2. As we show in the appendix, the proof of Theorem 1 can be modified to establish consistency in the following mean square sense:

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\boldsymbol{\beta}^0_{r_t^0}-\widehat{\boldsymbol{\beta}}_{\widehat{r}_t}\right\|^2=o_p(1), \tag{3.2}$$

which is analogous to Theorem 1 in Bonhomme and Manresa (2015), and Loyo and Boot (2025). However, this requires that $\min_t \lambda_{\min}(N^{-1}\mathbf{V}_t'\mathbf{M}_{F^0}\mathbf{V}_t) > 0$ w.p.a.1 instead of Assumption 1(g), which in turn means that N cannot be fixed anymore.

We now turn to the question under what conditions the SHIFTIE estimator achieves consistent classification of individual time periods. Under consistent classification, $\widehat{\boldsymbol{\beta}}$ will be asymptotically equivalent to the infeasible estimator based on the true group memberships, henceforth denoted $\widehat{\boldsymbol{\beta}}^{\mathrm{inf}}$.

Assumption 2.

(a) $\mathbf{w}_{i,t}$ is independent over i and t;

(b) $\mathbb{P}(|w_{j,i,t}| > M) \leq \exp(1-(M/\nu)^{\kappa})$ for all i and t, where $w_{j,i,t}$ is the j-th row of $\mathbf{w}_{i,t} := [w_{1,i,t}, \ldots, w_{k+1,i,t}]'$ and M, ν, $\kappa > 0$;

(c) $\mathbb{E}(N^{-1}\mathbf{V}_t'\mathbf{V}_t) \to \boldsymbol{\Sigma}_{v,t}$ as $N \to \infty$, where $\boldsymbol{\Sigma}_{v,t}$ is positive definite with $\sup_t \|\boldsymbol{\Sigma}_{v,t}\| < \infty$.

Classification consistency typically requires restrictions on the dependence structure and tail behavior of the data (see Bonhomme and Manresa, 2015, for a discussion). This is where Assumption 2 comes into play. Part (a) strengthens the weak dependence condition of Assumption 1(c) to independence. This assumption is used in the proofs;

however, intuition and the Monte Carlo results reported in Section 5.2 suggest that independence may be stronger than necessary. In any case, $y_{i,t}$ and $\mathbf{x}_{i,t}$ are still allowed to be dependent through the interactive effects. Similarly, while Assumption 2(b) requires that the tail probability of $\mathbf{w}_{i,t}$ decays at an exponential rate, the observed data may still exhibit thick tails.

Theorem 2. *Suppose that Assumptions 1 and 2 hold. Then, for all $\delta > 0$, as $N, T \to \infty$,*

(a) $\mathbb{P}\left(\sup_{t\in\{1,\ldots,T\}} \left|\widehat{r}_t - r_t^0\right| > 0\right) = o(1) + o(TN^{-\delta})$,

(b) $\max_{r\in\mathcal{R}} \left\|\widehat{\boldsymbol{\beta}}_r - \widehat{\boldsymbol{\beta}}_r^{\text{inf}}\right\| = o_p(N^{-\delta})$.

Unlike Theorem 1, Theorem 2 requires that both N and T grow without bound, which, as already noted, is a standard condition in the literature. Moreover, for $\widehat{r}_t$ to be consistent for r_t^0, the relative growth rates of N and T must satisfy $TN^{-\delta} = O(1)$. This condition is not particularly restrictive, as the only requirement on δ (at this point) is that it be positive.

Assumption 3.

(a) $T_r^{-1}\sum_{t=1}^{T}\mathbb{1}(r_t = r)\boldsymbol{\Sigma}_{v,t} \to \boldsymbol{\Theta}_r$ as $T \to \infty$ for all $r \in \mathcal{R}$, where $\boldsymbol{\Sigma}_{v,r}$ is positive definite;

(b) $T_r^{-1}\sum_{t=1}^{T}\mathbb{1}(r_t = r)\sigma_{\varepsilon,t}^2\boldsymbol{\Sigma}_{v,t} \to \boldsymbol{\Phi}_r$ as $T \to \infty$ for all $r \in \mathcal{R}$, where $\sigma_{\varepsilon,t}^2 = \mathbb{E}(\varepsilon_{i,t}^2)$;

(c) $T^{-1}N \to \eta < \infty$ as $N, T \to \infty$.

Assumption 3 ensures that the infeasible estimator $\widehat{\boldsymbol{\beta}}_r^{\text{inf}}$ is asymptotically well defined. This is important because, by Theorem 2, $\widehat{\boldsymbol{\beta}}_r$ is asymptotically equivalent to $\widehat{\boldsymbol{\beta}}_r^{\text{inf}}$. As a result, the asymptotic distribution of $\sqrt{NT_r}(\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r^0)$ stated in Theorem 3 below is derived from that of $\sqrt{NT_r}(\widehat{\boldsymbol{\beta}}_r^{\text{inf}} - \boldsymbol{\beta}_r^0)$. Assumptions 3(a) and 3(b) correspond to the standard identification and covariance matrix conditions (see, for example, Bonhomme and Manresa, 2015; and Loyo and Boot, 2025). The expansion rate condition in part (c) is needed to address the incidental parameters problem arising from the interactive effects, and is standard in studies that allow for such effects (see, for example, Bai, 2009, and Westerlund and Urbain, 2015, for discussions).

Theorem 3. *Suppose that Assumptions 1–3 hold, and that $\delta > 1$. Then, uniformly in $r \in \mathcal{R}$, as $N, T \to \infty$,*

$$\sqrt{NT_r}\left(\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r^0\right) \to_d \mathcal{N}\left(-\sqrt{\eta}\pi_r^{-1/2}\boldsymbol{\Theta}_r^{-1}\mathbf{b}_r, \boldsymbol{\Theta}_r^{-1}\boldsymbol{\Phi}_r\boldsymbol{\Theta}_r^{-1}\right),$$

where $\mathbf{b}_r$ is a $k \times 1$ bias vector such that $\mathbf{b}_r = \mathbf{b}_{2,r} + \mathbf{b}_{3,r} + \mathbf{b}_{4,r} - \mathbf{b}_{1,r}$ with

$$\begin{aligned}
\mathbf{b}_{1,r} &:= \lim_{T\to\infty} \frac{1}{T_r}\sum_{t=1}^{T} \mathbb{1}(r_t = r)(\boldsymbol{\Sigma}_{v,t}[\boldsymbol{\beta}_r, \mathbf{I}_k] - \boldsymbol{\Gamma}_t'\mathbf{B}_m'\boldsymbol{\Omega}_u)\mathbf{P}_{-m}\boldsymbol{\Sigma}_u\mathbf{B}_m\gamma_t, \\
\mathbf{b}_{2,r} &:= \lim_{T\to\infty} \frac{1}{T_r}\sum_{t=1}^{T} \mathbb{1}(r_t = r)\sigma_{\varepsilon,t}^2(\boldsymbol{\Sigma}_{v,t}[\boldsymbol{\beta}_r, \mathbf{I}_k] - \boldsymbol{\Gamma}_t'\mathbf{B}_m'\boldsymbol{\Omega}_u)\mathbf{P}_{-m}[1, \mathbf{0}_{1\times k}]', \\
\mathbf{b}_{3,r} &:= \lim_{T\to\infty} \frac{1}{T_r}\sum_{t=1}^{T} \mathbb{1}(r_t = r)(\boldsymbol{\Sigma}_{v,t}[\boldsymbol{\beta}_r, \mathbf{I}_k] - \boldsymbol{\Gamma}_t'\mathbf{B}_m'\boldsymbol{\Omega}_u)\mathbf{B}_m\gamma_t, \\
\mathbf{b}_{4,r} &:= \lim_{T\to\infty} \frac{1}{T_r}\sum_{t=1}^{T} \mathbb{1}(r_t = r)\sigma_{\varepsilon,t}^2\boldsymbol{\Gamma}_t'\mathbf{B}_m'[1, \mathbf{0}_{1\times k}]',
\end{aligned}$$

where $\mathbf{P}_{-m} := \mathbf{B}_{-m}(\mathbf{B}_{-m}'\boldsymbol{\Omega}_u\mathbf{B}_{-m})^{-1}\mathbf{B}_{-m}'$ and $\mathbf{B}_{-m}$ is as in the proof of Lemma A.1 in the appendix.

Theorem 3 implies that $\widehat{\boldsymbol{\beta}}_r$ is consistent and that the rate of convergence is given by $(NT_r)^{-1/2}$. The bias term $\sqrt{\eta}\pi_r^{-1/2}\boldsymbol{\Theta}_r^{-1}\mathbf{b}_r$ does not affect consistency; however, it does lead to a miscentering of the asymptotic distribution of $\sqrt{NT_r}(\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r^0)$, which invalidates inference. Because $\pi_r^{-1/2} = \lim_{T\to\infty}\sqrt{T/T_r} \geq 1$, the only exception is when

$\eta = 0$, in which case the distribution simplifies to $\mathcal{N}(\mathbf{0}_{k\times 1}, \mathbf{\Theta}_r^{-1}\mathbf{\Phi}_r\mathbf{\Theta}_r^{-1})$.[6] In practice, this means that we require $T \gg N$, which need not be the case in a given sample. Another possibility is to estimate the bias directly and subtract it from $\widehat{\boldsymbol{\beta}}_r$, which would yield an asymptotically unbiased estimator, provided that the bias estimate is consistent. Unfortunately, consistent estimation of $\mathbf{b}_r$ is only possible if $m = k + 1$, which, again, may not hold in practice.

These considerations motivate the use of the split panel jackknife (SPJ) proposed by Dhaene and Jochmans (2015), which requires neither $T \gg N$ nor explicit bias estimation. The specific SPJ estimator we consider is given by

$$\widehat{\boldsymbol{\beta}}_r^{\text{spj}} := 2\widehat{\boldsymbol{\beta}}_r - \overline{\boldsymbol{\beta}}_r, \tag{3.3}$$

where $\overline{\boldsymbol{\beta}}_r := (\widehat{\boldsymbol{\beta}}_{1,r} + \widehat{\boldsymbol{\beta}}_{2,r})/2$ with $\widehat{\boldsymbol{\beta}}_{\ell,r}$ for $\ell \in \{1,2\}$ given by

$$\widehat{\boldsymbol{\beta}}_{\ell,r} := \left(\sum_{t\in\mathcal{T}_{\ell,r}} \mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t \right)^{-1} \sum_{t\in\mathcal{T}_{\ell,r}} \mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{y}_t, \tag{3.4}$$

and where $\mathcal{T}_{1,r} := \{1, \ldots, T_r/2\}$ and $\mathcal{T}_{2,r} := \{T_r/2 + 1, \ldots, T_r\}$.

Assumption 4. $\mathbf{\Sigma}_{v,t} = \mathbf{\Sigma}_{v,r}$ and $\sigma^2_{\varepsilon,t} = \sigma^2_{\varepsilon,r}$ for all t such that $r_t = r \in \mathcal{R}$.

Assumption 4 requires that $\mathbf{\Sigma}_{v,t}$ and $\sigma^2_{\varepsilon,t}$ are constant within regimes. This is needed or else split panel jackknife will not be able to eliminate the bias of $\widehat{\boldsymbol{\beta}}_r$.

Theorem 4. *Suppose that Assumptions 1–4 hold, and that $\delta > 1$. Then, as N, $T \to \infty$,*

$$\sqrt{NT_r}\left(\widehat{\boldsymbol{\beta}}_r^{\text{spj}} - \boldsymbol{\beta}_r^0\right) \to_d \mathcal{N}\left(\mathbf{0}_{k\times 1}, \sigma^2_{\varepsilon,r}\mathbf{\Sigma}_{v,r}^{-1}\right).$$

[6] Note that unless $\eta = 0$, the presence of $\pi_r^{-1/2}$ will inflate the bias for relatively small regimes.

Theorem 4 verifies that the SPJ correction is asymptotically successful. Inference based on this result requires consistent estimators $\widehat{\sigma}^2_{\varepsilon,r}$ and $\widehat{\mathbf{\Sigma}}_{v,r}$ of $\sigma^2_{\varepsilon,r}$ and $\mathbf{\Sigma}_{v,r}$, respectively. Natural suggestions towards this end are $\widehat{\sigma}^2_{\varepsilon,r} = (NT_r)^{-1}\sum_{t=1}^{T}\mathbb{1}(\widehat{r}_t = r)\widehat{\boldsymbol{\varepsilon}}_t'\widehat{\boldsymbol{\varepsilon}}_t$ and $\widehat{\mathbf{\Sigma}}_{v,r} = (NT_r)^{-1}\sum_{t=1}^{T}\mathbb{1}(\widehat{r}_t = r)\widehat{\mathbf{V}}_t'\widehat{\mathbf{V}}_t$, where $\widehat{\boldsymbol{\varepsilon}}_t := \mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t\widehat{\boldsymbol{\beta}}_{\widehat{r}_t})$ and $\widehat{\mathbf{V}}_t := \mathbf{M}_{\widehat{F}}\mathbf{X}_t$.

All the results reported so far assume that the true number of regimes, R^0, is known, which may not be the case in practice. If R^0 is unknown, we follow the bulk of the existing group literature and minimize an information criterion (see, for example, Bonhomme and Manresa, 2015, Miao et al., 2020, and Su and Ju, 2018). Our proposed estimator $\widehat{R}$ of R^0 is given by

$$\widehat{R} := \underset{1\leq R\leq R_{\max}}{\arg\min}\ \mathrm{IC}\left(R\right), \tag{3.5}$$

where $R_{\max} \geq R^0$ is a pre-specified maximum number of regimes, and

$$\mathrm{IC}(R) := \ln\left(\frac{1}{NT}\sum_{t=1}^{T}\widehat{\boldsymbol{\varepsilon}}_t\left(R\right)'\widehat{\boldsymbol{\varepsilon}}_t\left(R\right)\right) + R\cdot\varphi_{NT} \tag{3.6}$$

Here we have made the dependence of $\widehat{\boldsymbol{\varepsilon}}_t$ on R explicit by writing $\widehat{\boldsymbol{\varepsilon}}_t = \widehat{\boldsymbol{\varepsilon}}_t(R)$. Consistency requires that the penalty term φ_{NT} satisfies $\varphi_{NT} \to 0$ and $\varphi_{NT}\max\{N, T\} \to \infty$ as N, $T \to \infty$.

4 Monte Carlo study

In this section, we investigate the finite-sample properties of our proposec SHIFTIE approach. The DGP used for this purpose is similar to those considered by Kaddoura and Westerlund (2023), and Kaddoura (2025), and is given by a restricted version of

(2.1) and (2.2) with $k = 3$ and $r = 2$. The loadings are generated as

$$\boldsymbol{\Gamma}_t = \varphi_1 \boldsymbol{\Gamma}_{t-1} + \boldsymbol{\Pi}_t, \tag{4.1}$$

$$\boldsymbol{\gamma}_t = \varphi_2 \boldsymbol{\gamma}_{t-1} + \boldsymbol{\theta}_t, \tag{4.2}$$

with initial conditions set to zero, parameters $\varphi_1 = 0.5$ and $\varphi_2 = 0.1$, and innovations drawn independently from $\mathcal{N}(0, 1)$. We use the same DGP as for $\boldsymbol{\gamma}_t$ to generate $\mathbf{f}_i$. The idiosyncratic error $\varepsilon_{i,t}$ is generated in the following way:

$$\varepsilon_{i,t} = \vartheta \varepsilon_{i,t-1} + \zeta_{i,t} + \sum_{\ell=1}^{L} \vartheta \left(\zeta_{i-j,t} + \zeta_{i+j,t} \right), \tag{4.3}$$

where $\varepsilon_{1,0} = \cdots = \varepsilon_{N,0} = 0$, $\vartheta_2 = 0.3$, $L = 10$, and $\zeta_{i,t} \sim \mathcal{N}(0, \sigma_i^2)$ with $\sigma_i \sim \mathcal{U}(0.5, 1)$. This means that $\varepsilon_{i,t}$ is both serially and cross-section correlated with the $2L$ nearest cross-sectional neighbours.[7] The DGP of $\mathbf{v}_{i,t}$ is the same, except that the innovations are drawn from $\mathcal{N}(0, 1)$. Hence, while $\varepsilon_{i,t}$ is heteroskedastic, $\mathbf{v}_{i,t}$ is not.

The primary focus of this simulation study is the slope coefficients. We consider two distinct DGPs in which the slopes display (DGP1) a temporary structural break and (DGP2) Markov switching. Both DGPs are inspired by our empirical illustration in Section 5. In DGP1, $\boldsymbol{\beta}_{r_t}$ is generated as follows:

$$\boldsymbol{\beta}_{r_t} = \begin{cases} \mathbf{0}_{p\times 1} & \text{for} \quad 1 \le t < \lfloor 0.5T \rfloor \\ \mathbf{1}_{p\times 1} & \text{for} \quad \lfloor 0.5T \rfloor \le t < \lfloor 0.6T \rfloor \\ \mathbf{0}_{p\times 1} & \text{for} \quad \lfloor 0.6T \rfloor \le t \le T \end{cases}, \tag{4.4}$$

where $\lfloor \cdot \rfloor$ is the floor function. Thus, $\boldsymbol{\beta}_{r_t}$ experiences a temporary structural break that

[7]The cross-sectional sum in $\mathbf{v}_{i,t}$ is truncated at beginning and end when not enough cross-sections are available.

subsequently reverts back to the original level. The number of regimes, R^0, is therefore given by two. In DGP2, $\boldsymbol{\beta}_{r_t}$ evolves according to a two-regime Markow switching process with the following transition matrix:

$$\mathbf{P} := \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}, \tag{4.5}$$

where $p_{ab} := \mathbb{P}\{r_t = b \mid r_{t-1} = a\}$ is the probability of moving from regime a at time $t-1$ to regime b at time t. For our simulation, we set $p_{11} = p_{22} = 0.8$ and $p_{12} = p_{21} = 0.2$.

The performance of SHIFTIE is evaluated based on 1000 replications of the above DGPs for all combinations of $N, T \in \{40, 60, 80, 100\}$. Hence, not only is the DGP challenging but most sample sizes are also very small. Note in particular how in DGP1 the second regime where $\lfloor 0.5T \rfloor \leq t < \lfloor 0.6T \rfloor$ only contains 4 time periods when $T = 40$. Three metrics are considered; the mean squared error (MSE) and bias of $\widehat{\boldsymbol{\beta}}_r$ and $\widehat{\boldsymbol{\beta}}_r^{\text{spj}}$, and the misclassification frequency of $\widehat{r}_t$. The bias and MSE are reported separately for each regime and all together. All results are based on treating R^0 as known.

INSERT TABLES 1 AND 2 ABOUT HERE

Tables 1 and 2 report the bias and MSE results for DGP1 and DGP2, respectively. As expected given Theorem 3, the baseline SHIFTIE estimator is biased, especially in the smallest samples. We also see that in DGP1 most of the inaccuracy originates from the second regime, which—as pointed out earlier—is relatively small. Things improve as N and T increase; however, some distortions remain, as to be expected given that all samples considered are quite small. Note in particular how the unbiasedness requirement that $T \gg N$ is not met. The SPJ takes care of this problem. In fact, the accuracy of the SPJ SHIFTIE version is very high, even in the smallest samples and regardless of the DGP considered. This corroborates Theorem 4.

INSERT TABLE 3 ABOUT HERE

In Table 3, we report the misclassification frequency for both DGPs. As expected given Theorem 2(a) this frequency goes down as N increases. It goes down also when T increases but not always. Note in particular how in DGP2 the misclassification is increasing in T when N is relatively small, which is again consistent with Theorem 2(a). Intuitively, unless N is large enough to ensure accurate estimation, more time periods to classify just means more errors.

5 Empirical illustration

5.1 Motivation

There is a substantial empirical literature concerned with economic growth and the business cycle in the US (see Owyang et al., 2005, and Hamilton and Owyang, 2012, and the references provided therein). The coronavirus pandemic (COVID–19) had a substantial effect on the US economy. In fact, as Federal Reserve Bank of Chicago (2022) point out, COVID–19 was highly disruptive even from an historical perspective. Real GDP contracted by 3.5% in 2020, which is the largest contraction since 1946 and the first contraction since the 2008 financial crisis. The US Census Bureau's Household Pulse Survey published weekly statistics of the effects of the pandemic on Americans' lives. For week 12 (July 16–21) of 2020, 51.1% of respondents reported a loss of employment income since March 13. During roughly the same period, the number of persons with jobs decreased by 14.7 million. Recovery began relatively quickly, though, with the recession lasting only two months—March and April—according to the National Bureau of Economic Research (NBER).[8] Of course, this does not mean that the effect of COVID–19 was gone after two months. In fact, unemployment did not return to its pre-pandemic level until 2022, and it took even longer for real GDP to return to its pre-pandemic trend (2023:Q4). Parts of the US economy have not completely recovered even today. The duration of the COVID–19 effect on economic growth is therefore unclear, although it seem to have been temporary. This is unlike the global financial crisis of 2007–2008, which shifted the level of real GDP down permanently (Federal Reserve, 2024).

Another unsettled issue is whether the recovery took place abruptly or gradually. A key factor here is the many interventions by the US government and the Federal Reserve aimed at keeping financial markets functioning and to provide stimulus. In particular, there was the Coronavirus Aid, Relief, and Economic Security Act—also known as the "CARES Act"—of March 2020, and the American Rescue Plan Act of March 2021—also called the "COVID–19 Stimulus Package" or "American Rescue Plan," which are the largest economic stimulus packages in US history. There were other stimulus packages as well, such as the Families First Coronavirus Response Act of March 2020, the Paycheck Protection Program and Healthcare Enhancement Act of April 2020, and the Consolidated Appropriations Act of December 2020, but these were not quite as large. There was also a massive quantitative easing programme launched by the Federal Reserve in March of 2020 and a lowering of the target for the federal funds rate to (almost) 0. While many interventions were introduced in March of 2020, they were in effect for many months, and there were subsequent interventions.

As the discussion in the last two paragraphs makes clear, there is great uncertainty

[8] One reason for this speedy recovery is the American Rescue Plan Act of 2021, also called the "COVID–19 Stimulus Package" or "American Rescue Plan", which was a USD 1.9 trillion economic stimulus package. Another reason is the Federal Reserve's cut of the federal funds rate and massive quantitative easing programs aimed at keeping financial markets functioning and to provide stimulus.

over both the duration and the size of the effect of COVID–19 on growth. Surprisingly, the part of the empirical literature that is concerned with this issues is very scarce, and the few studies that exists are either policy reports or research not focusing on the US or growth (see, for example, Congress, 2025, Sunge et al., 2024, and Ditzen et al., 2025). In fact, most growth studies are pre-COVID–19. There therefore much to do.

The present paper is the first to consider the effect of COVID–19 on US economic growth. This is our first contribution. Our second contribution lies in our choice of econometric method. Starting with the seminal work of Hamilton (1989), there is a long tradition in the existing business cycle literature to apply Markov switching approaches to aggregate data. As Owyang et al. (2005) point out, however, such data are not necessarily reflective of all regions of the country. Disaggregate panel data are more informative in this regard. Motivated by this observation, Owyang et al. employ state-level data for the 1979:01–2002:06 period. Their proposed approach has two steps. In the first step, Hamilton's (1989) time series Markov switching approach is used to identify state-specific regime shifts in the intercept of the dependent variable, which is given by the monthly state coincident index produced by the Federal Reserve Bank of Philadelphia. In the second step, the sample is split by estimated regime and regime-specific, pooled panel regression models are estimated where the dependent variable is regressed onto several regressors thought to explain growth.

While reasonable and appealing in its simplicity, Owyang et al.'s (2005) two-step approach has (at least) five drawbacks. One drawback is that the first-step estimation of the regimes is done state-by-state using only the time series variation of the data. This wasteful in the sense that when the regimes of one state are estimated, the information contained in all other states is ignored. The states are treated as though they are independent, which they are not. Another drawback is that the regimes are estimated based on the dependent variable only, which is again wasteful as it ignores the regressors. Stated differently, if the second-step regression model with regime specific slopes is the truth, the first-step intercept-only model is misspecified. A third drawback is that the error coming from the estimation of the regimes is unlikely to be negligible in the second step, as it is again based on a relatively imprecise time series-only approach. A fourth drawback is that the regimes are assumed to be Markov with only two states and constant transition probabilities, which need not be a mandate of the data. A fifth and final drawback is that the second-step regression model is fitted with only an overall intercept, and there is no attempt to account for unobserved heterogeneity.

The SHIFTIE methodology developed in Section 2 overcomes these drawbacks. The regimes and slopes are estimated jointly as opposed to sequentially in two steps, which means that no information is disregarded and that all estimation errors are accounted for. The estimated model is very general in that it allows for unobserved heterogeneity in the form of interactive fixed effects, and it places no assumptions on the nature of the shifts. SHIFTIE does assume that the shifts are homogenous; however, this is a reasonable assumption because all states were affected by COVID–19.

5.2 Data

The dependent variable is based on the same coincident index used not only by Owyang et al. (2005), but by a number of other studies (see, for example, Agudze et al., 2022, and Crone and Clayton-Matthews, 2005).[9] We use this index because it is reported monthly, as opposed to GDP, which is only available at a quarterly frequency. The regressors are extracted from the FRED–SD data set of Bokun et al. (2023), which contains 28 variables per state at either monthly or quarterly frequency. All quarterly variables were discarded. While the coincident indices are available from 1979:01, most variables in the

[9] More information and data are available online at `https://philadelphiafed.org/research-and-data/regional-economy/indexes/coincident`.

FRED–SD data are not available until later. Most variables are available from 1990:01. We therefore take this as the starting date of our sample. We take the latest end date available to us, which is 2025:03. We then removed all variables for which there were missing observations within the sample rage. This brought the number of variables down to 12, most of which are labor market related. Many of the variables are non-stationary, which, as pointed out in Section 3, is not permitted under our assumptions. We therefore transform them by taking logs when possible, and then first differences if needed to achieve stationarity. This process is copied from McCracken and Ng (2016). McCracken and Ng are not considering the same data set; however, most variables in the FRED–SD data set are also in the aggregate FRED–MD data set that they study. The coincident index is used in place of GDP, and is transformed accordingly by taking logs and first difference. Because of the differencing, the sample used in the analysis starts in 1990:02, for a total of $T = 422$ months and $N = 50$ states. This means that T is large relative to N, and that it is substantially larger than the values considered in the Monte Carlo study of Section . We therefore expect SHIFTIE to work well here.

The included regressors, which are inspired by studies such as Agudze et al. (2022), Owyang et al. (2005), and Hamilton and Owyang (2012), are the growth rate of the number of employees in the construction (CON), finance (FIN), government (GOV), information (INF), manufacturing (MAN), and private services (PRS) sectors, and the first difference of the participation rate (PAR).[10] In terms of the notation of Section , these seven variables are the ones that go into $\mathbf{x}_{i,t}$. The dependent variable, $y_{i,t}$, is the growth rate of the coincident index. The estimated loadings in $\widehat{\mathbf{F}}$ are given by the time series averages of all eight variables. A vector of ones is also included, which is tantamount to treating the presence of time fixed effects as known. We can therefore allow for as many

[10] We experimented with different model specifications, all of which yielded qualitatively similar results.

as nine unknown factors.

5.3 Results

We begin by considering the results on the estimated regimes. The number of shifts, R^0, is estimated using $\widehat{R}$. The maximum number of regimes is set to $R_{\max} = 5$ and the penalty term is specified as $\varphi_{NT} = (NT)^{-1}(T+N) \cdot \ln[(N+T)^{-1}NT]$, similarly to Bonhomme and Manresa (2015). We estimate $\widehat{R} = 2$ regimes. As alluded to earlier, many studies that assume Markov switching also assume two regimes. We require neither of these assumptions. However, our results actually support the use of two regimes.

INSERT FIGURES 1 AND 2 ABOUT HERE

The estimated regime memberships are illustrated in Figure 1. Regime 2 only contains four months, 2020:04–2020:06 and 2020:08, which are represented by the vertical red lines. All other months belong to regime 1. The red lines are plotted together with the average coincident index across states. We see that in 2020:04 the average index experiences a substantial drop but that it immediately bounces back up to a level that is even higher than before the drop. The drop in April coincides with NBERs recessions, which dates the peak and trough of the COVID–19 recession to 2020:02 and 2020:04, respectively. Regime 1 therefore covers the lowest point of the recession and the subsequent recovery. As already pointed out, many fiscal and monetary interventions occurred in March 2020. Regime 1 is located directly after these interventions. While there were other factors that also affected growth, the interventions described earlier were the largest events at the time and they are likely to have affected growth. Note also that while there are other major historical events in the sample, including the 2007–2008 global financial crisis, these do not show up in the results. COVID–19 was therefore unique.

Interestingly, if we increase the number of regimes from two to three, many of these other events appear, as is evident form Figure 2. The "COVID–19 regime", regime 2, is still there and seems very robust to changes in the number of regimes. The new regime, regime 3, contain a number of events that were not as significant as COVID–19 but that did nevertheless affect growth. One example occurs at the beginning of the sample, where we estimate parts of 1990 to regime 3. This coincides with the recession of that time. The global financial crisis is also a part of regime 3, as is the 2001 recession, and Black Monday of 2011. Moreover, while the COVID–19 regime lasts four months, the transition back to the "normal" regime, regime 1, does not occur directly but via regime 3. Regime 3 can therefore be meaningfully interpreted as a crisis regime that excludes the initial COVID–19 effect (regime 2). However, because $\widehat{R} = 2$, while economically interpretable, regime 3 is not statistically important. We therefore disregard it in what remains of this section.

INSERT TABLE 4 ABOUT HERE

The estimation results based on the best-performing SPJ-version of SHIFTIE are reported in Table 4. The first thing to note is that while the sign of the estimated coefficients is the same in the two regimes, the size and significance of those estimates differ markedly between regimes. The only exception is FIN for which there is a change of sign but then this regressor is always insignificant. The differences in the results for the other sectors/regressors depend on how exposed they were to the effects of the pandemic. Consider CON. This regressor is highly significant in regime 1 but not in regime 2. Construction is an important sector that under the pandemic was classified as "critical", which means that it was prioritized for continued operation (CISA, 2020). This explains not only the importance of CON for growth in normal times (regime 1), but also its disconnect during the pandemic (regime 2), as it was shielded from exposure. Parts of the government was also deemed critical, and we see that the effect of GOV is qualitatively similar to that of CON. Manufacturing was not protected in the same way. Both output and hours worked fell dramatically, especially in the motor vehicle production, which basically came to a complete stop. However, the sector quickly rebounced, as did growth; in fact, the increase in output and hours worked in the third quarter of 2020 were the largest ever recorded (US Bureau of Labor Statistics, 2022). This explains why the estimated effect of MAN is relatively large in regime 2. The same is true for PRS. The estimated effect of PAR is small but significant. The fact that it is negative is counterintuitive and requires an explanation. Except for a few episodes like the global financial crisis and COVID–19, US economic activity has been steadily increasing during the sample period. Yet, since the mid-1990's, the labor force participation rate for the total population ages 16 and over has actually been declining. The reason for this is a rise in the percentage of the population ages 65 and over, which is relatively less likely to be working (see US Census Bureau, 2021). This explains the negative association between growth and PAR.

The above results show how the pandemic stands apart from other crises. They corroborate the following conclusion by Federal Reserve Bank of Chicago (2022): "the Covid-19 pandemic recession and the subsequent expansion—hereafter referred to as the 'pandemic cycle' or 'pandemic era'—have less in common with those of other business cycles."

6 Conclusion

This study proposes a new approach to the estimation of panel data regression models with interactive effects and possibly time-varying slope coefficients. A major strength of the new methodology is that the type of time-variation that can be accommodated is very broad and includes most common specifications in the literature, such as structural breaks and Markov switching, as special cases. This generality is important not only from a theoretical point of view, but also in practice, as existing approaches require empirical researchers to take a stance as to the shift-generating process, which is almost always unknown. We therefore expects the new approach to become a valuable addition to the already existing menu of tools for handling shifts in model coefficients.

References

Abadir, K. M., and J. R. Magnus (2005). *Matrix Algebra, Econometric Exercises 1*. New York: Cambridge University Press.

Agudze, K. M., M. Billio, R. Casarin and F. Ravazzolo (2022). Markov switching panel with endogenous synchronization effects. *Journal of Econometrics* **230**, 281–298.

Ando, T. and J. Bai (2016). Panel data models with grouped factor structure under unknown group membership. *Journal of Applied Econometrics* **31**, 163–191.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica* **77**, 1229–1279.

Baltagi, B. H., Q. Feng and C. Kao (2016). Estimation of heterogeneous panels with structural breaks. *Journal of Econometrics* **191**, 176–195.

Bokun, K. O., L. E. Jackson, K. L. Kliesen and M. T. Owyang (2023). FRED–SD: A real-time database for state-level data with forecasting applications. *International Journal of Forecasting* **39**, 279–297.

Boldea, O., B. Drepper and Z. Gan (2020). Change point estimation in panel data with time-varying individual effects. *Journal of Applied Econometrics* **35**, 712–727.

Bonhomme, S., and E. Manresa (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, **83**, 1147–1184.

Cheng T., J. Gao and Y. Yan (2010). Regime switching panel data models with interactive fixed effects. *Economics Letters* **177**, 47–51.

CISA (2020). Identifying Critical Infrastructure During COVID–19. Available online at: `https://www.cisa.gov/topics/risk-management/coronavirus/identifying-c-`

ritical-infrastructure-during-covid-19.

Congress (2025). U.S. Economic Recovery in the Wake of COVID19: Successes and Challenges. Available online at: https://www.congress.gov/crs-product/R47115.

Dhaene, G., and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies* **82**, 991–1030.

Ditzen, J., Y. Karavias and J. Westerlund (2025). Multiple structural breaks in interactive effects panel data models. Forthcoming in *Journal of Applied Econometrics*.

Fang, Y., K. A. Loparo and X. Feng (1994). Inequalities for the trace of matrix product. *IEEE Transactions on Automatic Control* **39**, 2489–2490.

Federal Reserve (2024). Why is the U.S. GDP recovering faster than other advanced economies? Available online at: https://www.federalreserve.gov/econres/notes/feds-notes/why-is-the-u-s-gdp-recovering-faster-than-other-advanced-economies-20240517.html.

Hamilton, J. D., and M. T. Owyang (2012). The propagation of regional recessions. *Review of Economics and Statistics* **94**, 935–947.

Hamilton, J. D. (2016). Chapter 3 – Macroeconomic regimes and regime shifts. In Taylor, J. B., and H. Uhlig (Eds.), *Handbook of Macroeconomics* **2**, 163–201. Elsevier.

Hansen, C., and Y. Liao (2019). The factor-lasso and K-step bootstrap approach for inference in high-dimensional economic applications. *Econometric Theory* **35**, 465–509.

Karabiyik, H., S. Reese and J. Westerlund (2017). On the role of the rank condition in CCE estimation of factor-augmented panel regressions. *Journal of Econometrics* **197**, 60–64.

Kaddoura, Y., and J. Westerlund (2023). Estimation of panel data models with random interactive effects and multiple structural breaks when T is Fixed. *Journal of Business & Economic Statistics* **41**, 778–790.

Kaddoura, Y., and J. Westerlund (2025). CCE under non-random heterogeneity. Forthcoming in *Econometrics Journal*.

Karavias, Y., P. K. Narayan and J. Westerlund (2023). Structural breaks in interactive effects panels and the stock market reaction to COVID–19. *Journal of Business & Economic Statistics* **41**, 653–666.

Ke, Y., J. Li and W. Zhang (2016). Structure identification in panel data analysis. *Annals of Statistics* **44**, 1193–1233.

Loyo, J. A., and T. Boot (2024). Grouped heterogeneity in linear panel data models with heterogeneous error variances. *Journal of Business & Economic Statistics* **43**, 68–80.

Miao, K., K. Li and L. Su (2020). Panel threshold models with interactive fixed effects. *Journal of Econometrics* **219**, 137–170.

Moon, H. R., and B. Perron (2004). Testing for a unit root in panels with dynamic factors. *Journal of Econometrics* **122**, 81–126.

Owyang, M. T., J. Piger and H. J. Wall (2005). Business cycle phases in US states. *Review of Economics and Statistics* **87**, 604–616.

Perron, P. (2006). Dealing with structural breaks. In Patterson, K., and T. C. Mills (Eds.), *Palgrave Handbook of Econometrics 1: Econometric Theory*. Basingstoke: Palgrave Macmillan.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* **74**, 967–1012.

Piger, J. (2009). Econometrics: Models of regime changes. In Meyers, R. (Ed.), *Complex Systems in Finance and Econometrics*. New York: Springer.

Qian, J., and L. Su (2016). Shrinkage estimation of common breaks in panel data models via adaptive group fused Lasso. *Journal of Econometrics* **191**, 86–109.

Su, L., and G. Ju (2018). Identifying latent grouped patterns in panel data models with interactive fixed effects. *Journal of Econometrics* **206**, 554–573.

Sunge, R., C. Mudzingiri and N. Mkhize (2024). The COVID-19 pandemic and economic recovery: The mediating role of governance, a global perspective. *Heliyon* **10**, e39869.

US Bureau of Labor Statistics (2022). U.S. manufacturing output, hours worked, and productivity recover from COVID–19. Available online at: `https://www.bls.gov/pub/ted/2022/u-s-manufacturing-output-hours-worked-and-productivity-recover-from-covid-19.htm`.

US Census Bureau (2021). Why Did Labor Force Participation Rate Decline When the Economy Was Good? Available online at: `https://www.census.gov/library/stories/2021/06/why-did-labor-force-participation-rate-decline-when-economy-was-good.html`.

Westerlund, J. (2018). CCE in panels with general unknown factors. *Econometrics Journal* **21**, 264–276.

Westerlund, J., and J.-P. Urbain (2015). Cross-sectional averages versus principal components. *Journal of Econometrics* **185**, 372–377.

Zhu, H., V. Sarafidis and M. J. Silvapulle (2020). A new structural break test for panels with common factors. *Econometrics Journal* **23**, 137–155.

Table 1: Bias and MSE in DGP1.

N	T	SHIFTIE Total	SHIFTIE Regime 1	SHIFTIE Regime 2	SPJ Total	SPJ Regime 1	SPJ Regime 2
				Bias			
40	40	-0.092551	0.013958	-0.199059	-0.014811	0.006829	-0.036451
40	60	-0.053652	0.004663	-0.111968	-0.012752	0.002677	-0.028180
40	80	-0.029844	0.007988	-0.067675	-0.009111	0.002500	-0.020722
40	100	-0.025014	0.002352	-0.052380	-0.010456	0.000949	-0.021861
60	40	-0.065158	0.009285	-0.139602	-0.019352	0.003645	-0.042350
60	60	-0.027844	0.003836	-0.059524	-0.012269	0.001340	-0.025878
60	80	-0.017835	0.005487	-0.041157	-0.011625	0.003911	-0.027162
60	100	-0.013561	0.002859	-0.029982	-0.008278	0.001888	-0.018444
80	40	-0.035505	0.005175	-0.076185	-0.014790	0.006889	-0.036469
80	60	-0.018815	0.003751	-0.041381	-0.010748	0.002569	-0.024064
80	80	-0.012985	0.001984	-0.027953	-0.010637	0.001730	-0.023003
80	100	-0.008291	0.001010	-0.017592	-0.006427	0.001790	-0.014645
100	40	-0.027847	0.007087	-0.062780	-0.017358	0.005357	-0.040074
100	60	-0.015441	0.002822	-0.033705	-0.012965	0.002373	-0.028303
100	80	-0.008888	0.007316	-0.025092	-0.007271	0.003352	-0.017894
100	100	-0.006378	0.000405	-0.013162	-0.006380	0.000659	-0.013419
				MSE			
40	40	0.123408	0.025348	0.221468	0.036066	0.006652	0.065481
40	60	0.062230	0.014835	0.109625	0.024118	0.004425	0.043811
40	80	0.042042	0.014438	0.069647	0.018846	0.003380	0.034312
40	100	0.025286	0.008470	0.042102	0.014054	0.002572	0.025537
60	40	0.081578	0.018056	0.145100	0.029529	0.005664	0.053395
60	60	0.035504	0.009050	0.061959	0.018898	0.003550	0.034246
60	80	0.019370	0.006347	0.032394	0.014136	0.002805	0.025466
60	100	0.015833	0.005477	0.026190	0.010965	0.002091	0.019840
80	40	0.037548	0.005688	0.069407	0.021742	0.004634	0.038851
80	60	0.021953	0.007307	0.036599	0.015515	0.002934	0.028096
80	80	0.012613	0.003355	0.021871	0.011923	0.002239	0.021608
80	100	0.008986	0.001733	0.016240	0.009377	0.001730	0.017025
100	40	0.030367	0.008440	0.052293	0.020193	0.003883	0.036504
100	60	0.015583	0.005112	0.026054	0.014251	0.002610	0.025891
100	80	0.016801	0.007994	0.025609	0.010308	0.002057	0.018559
100	100	0.007125	0.001559	0.012692	0.008027	0.001593	0.014461

Notes: The table reports the bias and MSE of the estimated slopes, $\widehat{\boldsymbol{\beta}}_r$. The results are reported separately for each regime and all together. "SHIFTIE" and "SPJ" refer to the basic SHIFTIE estimator and its split panel jackknife version, respectively. The DGP is described in the text.

Table 2: Bias and MSE in DGP2.

		SHIFTIE			SPJ		
N	T	Total	Regime 1	Regime 2	Total	Regime 1	Regime 2
				Bias			
40	40	-0.187871	0.024302	-0.400043	-0.011139	0.006071	-0.028348
40	60	-0.148834	0.025061	-0.322729	-0.011213	0.002076	-0.024502
40	80	-0.111652	0.019381	-0.242685	-0.007219	0.002426	-0.016863
40	100	-0.104958	0.019939	-0.229855	-0.009008	0.000881	-0.018896
60	40	-0.152110	0.016738	-0.320957	-0.015626	0.002961	-0.034214
60	60	-0.104771	0.013413	-0.222955	-0.010137	0.000953	-0.021227
60	80	-0.088480	0.007632	-0.184593	-0.008864	0.002054	-0.019783
60	100	-0.066627	0.008063	-0.141317	-0.007161	0.001740	-0.016063
80	40	-0.106731	0.012961	-0.226424	-0.010712	0.006322	-0.027746
80	60	-0.073646	0.011019	-0.158311	-0.009096	0.003526	-0.021718
80	80	-0.056092	0.008733	-0.120917	-0.008750	0.001366	-0.018867
80	100	-0.040993	0.007286	-0.089271	-0.004670	0.001702	-0.011042
100	40	-0.086420	0.013852	-0.186691	-0.012966	0.004869	-0.030801
100	60	-0.053053	0.009932	-0.116038	-0.009943	0.002119	-0.022005
100	80	-0.036325	0.007584	-0.080234	-0.006046	0.003057	-0.015148
100	100	-0.028008	0.001619	-0.057634	-0.005218	0.000535	-0.010971
				MSE			
40	40	0.188017	0.042474	0.333561	0.034362	0.006585	0.062138
40	60	0.141960	0.031473	0.252446	0.028081	0.005569	0.050592
40	80	0.096567	0.020778	0.172356	0.020705	0.004555	0.036856
40	100	0.079148	0.018417	0.139878	0.019168	0.003819	0.034518
60	40	0.145974	0.040419	0.251528	0.023678	0.004409	0.042948
60	60	0.096169	0.024787	0.167551	0.018535	0.003556	0.033514
60	80	0.063095	0.018975	0.107216	0.014993	0.002927	0.027059
60	100	0.046753	0.014080	0.079427	0.013655	0.002583	0.024728
80	40	0.105628	0.031660	0.179597	0.018409	0.003344	0.033474
80	60	0.068641	0.020423	0.116858	0.013718	0.002763	0.024672
80	80	0.045808	0.015408	0.076208	0.011665	0.002225	0.021105
80	100	0.031326	0.011560	0.051091	0.010039	0.002029	0.018050
100	40	0.095274	0.032040	0.158508	0.013771	0.002558	0.024983
100	60	0.053045	0.015895	0.090195	0.010806	0.002084	0.019527
100	80	0.033558	0.013477	0.053638	0.009147	0.001711	0.016583
100	100	0.019950	0.006429	0.033471	0.007862	0.001564	0.014160

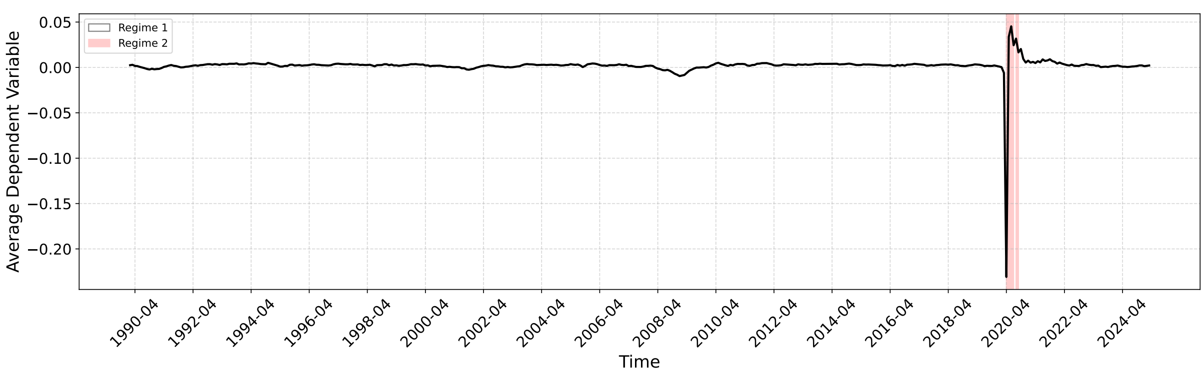
Notes: The table reports the bias and MSE of the estimated slopes, $\widehat{\boldsymbol{\beta}}_r$. The results are reported separately for each regime and all together. "SHIFTIE" and "SPJ" refer to the basic SHIFTIE estimator and its split panel jackknife version, respectively. The DGP is described in the text.

Table 3: Misclassification frequency.

$N \setminus T$	40	60	80	100
		DGP1		
40	2.010	1.457	1.340	0.972
60	1.202	0.745	0.293	0.126
80	0.376	0.293	0.172	0.125
100	0.286	0.126	0.363	0.033
		DGP2		
40	8.040	9.923	9.876	11.639
60	6.061	6.404	6.535	6.518
80	4.049	4.084	4.139	3.933
100	3.184	2.813	2.666	2.086

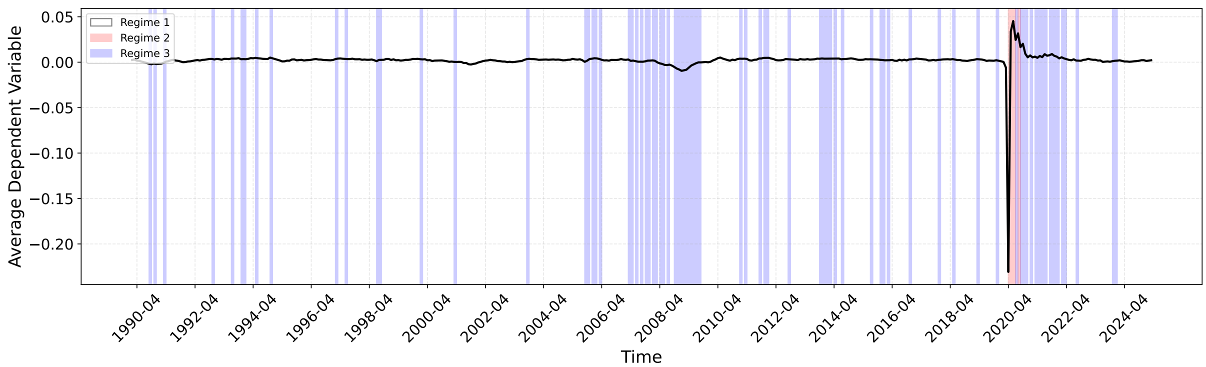
Notes: The table reports the misclassification frequency in % of the estimated regime memberships, $\widehat{r}_t$. DGP1 and DGP2 are described in the text.

Figure 1: Estimated regime memberships.



Notes: The black dashed line represents the cross-sectional mean of the dependent variable, the coincident index. The red vertical lines are months that SHIFTIE estimate to belong to regime 2, which are 2020:04–2020:06 and 2020:08. All other months are estimated to belong to the regime 1.

Figure 2: Estimated regime memberships with three regimes.



Notes: The black dashed line represents the cross-sectional mean of the dependent variable, the coincident index. The red (blue) vertical lines are months that SHIFTIE estimate to belong to regimes 2 and 3, respectively. All other months are estimated to belong to the regime 1.

Table 4: The estimated regime-specific slope coefficients.

Regressor	Estimate	SE	t-statistic	p-value
		Regime 1		
CON	0.022399	0.001811	12.3667	0.0000
FIN	-0.003316	0.003924	-0.8449	0.3982
GOV	0.033948	0.002443	13.8986	0.0000
INF	-0.003277	0.001296	-2.5283	0.0115
MAN	0.030352	0.002033	14.9282	0.0000
PRS	0.270826	0.007392	36.6392	0.0000
PAR	-0.005731	0.000135	-42.4967	0.0000
		Regime 2		
CON	0.008094	0.077188	0.1049	0.9165
FIN	0.879779	0.669142	1.3148	0.1886
GOV	0.202993	0.271041	0.7489	0.4539
INF	-0.124576	0.177773	-0.7008	0.4835
MAN	0.473399	0.127820	3.7036	0.0002
PRS	1.498865	0.288253	5.1998	0.0000
PAR	-0.011569	0.004072	-2.8413	0.0045

Notes: "CON", "FIN", "GOV'", "INF", "MAN" and "PRS" refer to the growth rate of the number of employees in the construction, finance, government, information, manufacturing, and private services sectors, respectively, while "PAR" refers to the first difference of the participation rate. Regime 2 contains the months 2020:04–2020:06 and 2020:08. All other months belong to regime 1. The estimated slopes are based on the SPJ version of SHIFTIE. "SE" refers to the estimated standard errors. The reported t-statistic and the associated p-value test the null hypothesis that the relevant slope coefficient is zero.

Appendix

Our asymptotic results rely on $\widehat{\mathbf{F}}$ being consistent in some sense. The limit of this object depends on whether $m = k+1$ or $m < k+1$ (similarly to Karabiyik et al., 2017, but now applied to loadings instead of factors). Suppose first that $m < k+1$. This means that some of the columns of $\widehat{\mathbf{F}}$ are degenerate. Note in particular that if $\operatorname{rank}\overline{\mathbf{C}} = m$, which holds under out Assumption 1, we may without loss of generality partition $\overline{\mathbf{C}}$ as $\overline{\mathbf{C}} =: [\overline{\mathbf{C}}_m, \overline{\mathbf{C}}_{-m}]$, where $\overline{\mathbf{C}}_m$ is an $m \times m$ full rank matrix and $\overline{\mathbf{C}}_{-m}$ is $m \times (k+1-m)$. By similarly partitioning $\overline{\mathbf{U}} =: [\overline{\mathbf{U}}_m, \overline{\mathbf{U}}_{-m}]$, we have

$$\widehat{\mathbf{F}} = [\mathbf{F}\overline{\mathbf{C}}_m, \mathbf{F}\overline{\mathbf{C}}_{-m}] + [\overline{\mathbf{U}}_m, \overline{\mathbf{U}}_{-m}]. \tag{A.1}$$

Define

$$\overline{\mathbf{B}} := \begin{bmatrix} \overline{\mathbf{C}}_m^{-1} & -\overline{\mathbf{C}}_m^{-1}\overline{\mathbf{C}}_{-m} \\ \mathbf{0}_{(k+1-m)\times m} & \mathbf{I}_{k+1-m} \end{bmatrix} = [\overline{\mathbf{B}}_m, \overline{\mathbf{B}}_{-m}], \tag{A.2}$$

with obvious definitions of $\overline{\mathbf{B}}_m$ and $\overline{\mathbf{B}}_{-m}$. Note that while $\overline{\mathbf{B}}_m$ is $(k+1) \times m$, $\overline{\mathbf{B}}_{-m}$ is $(k+1) \times (k+1-m)$, which means that $\overline{\mathbf{B}}$ is $(k+1) \times (k+1)$. The matrix $\overline{\mathbf{B}}$ is also full rank, because $\operatorname{rank}\overline{\mathbf{B}} = \operatorname{rank}\overline{\mathbf{C}}_m^{-1} + \operatorname{rank}\mathbf{I}_{k+1-m} = m + (k+1-m) = k+1$ (see Abadir and Magnus, 2005, Exercise 5.43), and can therefore be inverted. Moreover, $\overline{\mathbf{C}}\overline{\mathbf{B}} = [\mathbf{I}_m, \mathbf{0}_{m\times(k+1-m)}]$, which we can use to construct the following:

$$\widehat{\mathbf{F}}\overline{\mathbf{B}} = \mathbf{F}\overline{\mathbf{C}}\overline{\mathbf{B}} + \overline{\mathbf{U}}\overline{\mathbf{B}} = [\mathbf{F}, \mathbf{0}_{N\times(k+1-m)}] + [\overline{\mathbf{U}}\overline{\mathbf{B}}_m, \overline{\mathbf{U}}\overline{\mathbf{B}}_{-m}]. \tag{A.3}$$

By the definition of $\mathbf{U}_t$, $\mathbf{U}_t := [\boldsymbol{\varepsilon}_t + \mathbf{V}_t\boldsymbol{\beta}_{r_t}, \mathbf{V}_t] = \mathbf{W}_t\mathbf{Q}_{r_t}$, where $\mathbf{W}_t := [\boldsymbol{\varepsilon}_t, \mathbf{V}_t]$ is

$N \times (k+1)$ and

$$\mathbf{Q}_{r_t} := \begin{bmatrix} 1 & \mathbf{0}_{1\times k} \\ \boldsymbol{\beta}_{r_t} & \mathbf{I}_k \end{bmatrix} \tag{A.4}$$

is $(k+1) \times (k+1)$ and full rank (see Abadir and Magnus, 2005, Exercise 5.48). Hence, denoting by $\mathbf{u}'_{i,t}$ ($\mathbf{w}'_{i,t}$) the i-th row of $\mathbf{U}_t$ ($\mathbf{W}_t$), we have $\mathbf{u}_{i,t} = [\varepsilon_{i,t} + \boldsymbol{\beta}'_{r_t}\mathbf{v}_{i,t}, \mathbf{v}'_{i,t}]' = \mathbf{Q}'_{r_t}\mathbf{w}_{i,t}$. We also use the following: $\text{tr}\,(\mathbf{A}'\mathbf{B}) \leq \sqrt{\text{tr}\,(\mathbf{A}'\mathbf{A})\text{tr}\,(\mathbf{B}'\mathbf{B})} = \|\mathbf{A}\|\|\mathbf{B}\|$ if $\mathbf{A}$ and $\mathbf{B}$ are of the same dimension (see Abadir and Magnus, 2005, Exercise 12.5). Applying these results to $\mathbb{E}(\|\sqrt{T}N^{-1/2}\overline{\mathbf{U}}\|^2)$, we can show that

$$\begin{aligned}
&\mathbb{E}(\|\sqrt{T}N^{-1/2}\overline{\mathbf{U}}\|^2) \\
&= TN^{-1}\mathbb{E}[\text{tr}\,(\overline{\mathbf{U}}'\overline{\mathbf{U}})] = \frac{T}{N}\sum_{i=1}^{N}\mathbb{E}[\text{tr}\,(\overline{\mathbf{u}}_i\overline{\mathbf{u}}'_i)] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}[\text{tr}\,(\mathbf{u}_{i,t}\mathbf{u}'_{i,s})] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}[\text{tr}\,(\mathbf{Q}'_{r_t}\mathbf{w}_{i,t}\mathbf{w}'_{i,s}\mathbf{Q}_{r_s})] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\text{tr}\,[\mathbf{Q}_{r_s}\mathbf{Q}'_{r_t}\mathbb{E}(\mathbf{w}_{i,t}\mathbf{w}'_{i,s})] \leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\|\mathbf{Q}_{r_s}\mathbf{Q}'_{r_t}\|\|\mathbb{E}(\mathbf{w}_{i,t}\mathbf{w}'_{i,s})\| \\
&\leq \sup_{t,s}\|\mathbf{Q}_{r_s}\mathbf{Q}'_{r_t}\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\|\mathbb{E}(\mathbf{w}_{i,t}\mathbf{w}'_{i,s})\| = O(1),
\end{aligned} \tag{A.5}$$

where the last inequality is due to the fact that the autocovariances of $\mathbf{w}_{i,t}$ are absolute summable by Assumption 1, and $\sup_{t,s}\|\mathbf{Q}_{r_s}\mathbf{Q}'_{r_t}\| \leq \sup_{t,s}\|\mathbf{Q}_{r_s}\|\|\mathbf{Q}_{r_t}\| \leq (\sup_t\|\mathbf{Q}_{r_t}\|)^2 = O(1)$ since $\|\boldsymbol{\beta}_r\| < \infty$ by the compactness of $\mathcal{B}$. Hence, since the variance is $O(1)$,

$$\|\sqrt{T}N^{-1/2}\overline{\mathbf{U}}\| = O_p(1), \tag{A.6}$$

which holds regardless of whether N is fixed or tending to infinity. Note in particular how $\|\overline{\mathbf{U}}\| = O_p(T^{-1/2})$ for a fixed N, which means that the last $k+1-m$ columns of $\widehat{\mathbf{F}}\overline{\mathbf{B}}$ are degenerate. In order to address this issue, we introduce $(k+1)\times(k+1)$ normalization matrix $\mathbf{D}_T := \mathrm{diag}(\mathbf{I}_m, \sqrt{T}\mathbf{I}_{k+1-m})$. Post-multiplication by this matrix the above equation becomes

$$
\begin{aligned}
\widehat{\mathbf{F}}\overline{\mathbf{B}}\mathbf{D}_T &= [\mathbf{F}, \mathbf{0}_{N\times(k+1-m)}] + [\overline{\mathbf{U}}\overline{\mathbf{B}}_m, \sqrt{T}\overline{\mathbf{U}}\overline{\mathbf{B}}_{-m}] \\
&= [\mathbf{F}, \sqrt{T}\overline{\mathbf{U}}\overline{\mathbf{B}}_{-m}] + [\overline{\mathbf{U}}\overline{\mathbf{B}}_m, \mathbf{0}_{N\times(k+1-m)}].
\end{aligned} \tag{A.7}
$$

Hence, defining $\widehat{\mathbf{F}}^0 := \widehat{\mathbf{F}}\overline{\mathbf{B}}\mathbf{D}_T$ and $\mathbf{F}^0 := [\mathbf{F}, \sqrt{T}\overline{\mathbf{U}}\overline{\mathbf{B}}_{-m}]$ and making use of the fact that $\|\overline{\mathbf{U}}\| = O_p(T^{-1/2})$ and $\|\overline{\mathbf{C}}\| < \infty$ by Assumption 1, we can show that

$$
\|\widehat{\mathbf{F}}^0 - \mathbf{F}^0\| = \|[\overline{\mathbf{U}}\overline{\mathbf{B}}_m, \mathbf{0}_{N\times(k+1-m)}]\| \le \|\overline{\mathbf{U}}\|\|\overline{\mathbf{B}}_m\| = O_p(T^{-1/2}), \tag{A.8}
$$

The result in (A.8) assumes that $m < k+1$. The corresponding result for the case when $m = k+1$ follows by the same arguments after redefining $\overline{\mathbf{C}} := \overline{\mathbf{C}}_m$, $\overline{\mathbf{B}} := \overline{\mathbf{C}}^{-1}$ and $\mathbf{D}_T := \mathbf{I}_m$, such that $\widehat{\mathbf{F}}^0 = \widehat{\mathbf{F}}\overline{\mathbf{C}}^{-1}$ and $\mathbf{F}^0 = \mathbf{F}$. Hence, with this modification to the notation (A.8) holds regardless of whether $m = k+1$ or $m < k+1$.

The above results presume that N is fixed. If N is large, then we can make use of $\|N^{-1/2}\overline{\mathbf{U}}\|^2 = O_p(T^{-1})$ to show that

$$
\|N^{-1/2}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\|^2 \le \|N^{-1/2}\overline{\mathbf{U}}\|^2\|\overline{\mathbf{B}}_m\|^2 = O_p(T^{-1}). \tag{A.9}
$$

Note that the scaling by N here is inconsequential if N fixed. This last result therefore holds regardless of whether $N < \infty$ or $N \to \infty$. Lemma A.1 summarizes this.

Lemma A.1. *Suppose that Assumption 1 holds. Then, as $T \to \infty$ with $N < \infty$ or $N \to \infty$,*

$$
\|N^{-1/2}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\| = O_p(T^{-1/2}). \tag{A.10}
$$

Before we state our next lemma, analogously to Bonhomme and Manresa (2015, Proof of Theorem 1), we introduce the following functions:

$$Q(\boldsymbol{\beta}, \mathbf{r}) := \frac{1}{NT} \sum_{t=1}^{T} \|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_{r_t})\|^2, \tag{A.11}$$

$$Q^0(\boldsymbol{\beta}, \mathbf{r}) := \frac{1}{NT} \sum_{t=1}^{T} \|\mathbf{M}_{F^0} \mathbf{V}_t (\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t})\|^2 + \frac{1}{NT} \sum_{t=1}^{T} \|\mathbf{M}_{F^0} \boldsymbol{\varepsilon}_t\|^2. \tag{A.12}$$

Note that the scaling by N and T here is again inconsequential. The proposed estimator $(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{r}})$ of $(\boldsymbol{\beta}, \mathbf{r})$ minimizes $Q(\cdot, \cdot)$. Lemma A.2 makes clear that $Q^0(\cdot, \cdot)$ can be regarded as the limiting version of $Q(\cdot, \cdot)$. In order to avoid cluttering the notation, the dependence of these functions on N and T is omitted.

Lemma A.2. *Suppose that Assumption 1 holds. Then, as $T \to \infty$ with $N < \infty$ or $N \to \infty$,*

$$\sup_{(\boldsymbol{\beta}, \mathbf{r}) \in \mathcal{B}^R \times \mathcal{R}^T} |Q(\boldsymbol{\beta}, \mathbf{r}) - Q^0(\boldsymbol{\beta}, \mathbf{r})| = O_p(T^{-1/2}). \tag{A.13}$$

Proof: The following basic result will be used repeatedly in this proof: $\|\mathbf{A} + \mathbf{B}\|^2 = \|\mathbf{A}\|^2 + \|\mathbf{B}\|^2 + 2\mathrm{tr}\,(\mathbf{A}'\mathbf{B})$. We begin by applying it to $\|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_{r_t})\|^2 = \|\mathbf{M}_{\widehat{F}} \mathbf{X}_t (\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t}) + \mathbf{M}_{\widehat{F}}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)\|^2$, giving

$$\begin{aligned} \|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_{r_t})\|^2 = {} & \|\mathbf{M}_{\widehat{F}} \mathbf{X}_t (\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t})\|^2 + \|\mathbf{M}_{\widehat{F}}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)\|^2 \\ & + 2\mathrm{tr}\,[(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)\mathbf{M}_{\widehat{F}} \mathbf{X}_t (\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t})]. \end{aligned} \tag{A.14}$$

Consider the first term on the right-hand side. From $\overline{\mathbf{C}}\overline{\mathbf{B}}_m = \mathbf{I}_m$, we obtain

$$\mathbf{X}_t = \widehat{\mathbf{F}}\overline{\mathbf{B}}_m \boldsymbol{\Gamma}_t - (\widehat{\mathbf{F}} - \mathbf{F}\overline{\mathbf{C}})\overline{\mathbf{B}}_m \boldsymbol{\Gamma}_t + \mathbf{V}_t = \widehat{\mathbf{F}}\overline{\mathbf{B}}_m \boldsymbol{\Gamma}_t - \overline{\mathbf{U}}\overline{\mathbf{B}}_m \boldsymbol{\Gamma}_t + \mathbf{V}_t. \tag{A.15}$$

Hence, since $\widehat{\mathbf{F}}'\mathbf{M}_{\widehat{F}} = \mathbf{0}_{(k+1)\times N}$, we obtain $\mathbf{X}_t'\mathbf{M}_{\widehat{F}} = (\mathbf{V}_t - \overline{\mathbf{U}}\overline{\mathbf{B}}_m \boldsymbol{\Gamma}_t)'\mathbf{M}_{\widehat{F}}$. By using this and another application of the above result for $\|\mathbf{A}+\mathbf{B}\|^2$,

$$\begin{aligned}
\|\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}^0_{r^0_t}-\boldsymbol{\beta}_{r_t})\|^2 &= \|\mathbf{M}_{\widehat{F}}(\mathbf{V}_t-\overline{\mathbf{UB}}_m\boldsymbol{\Gamma}_t)(\boldsymbol{\beta}^0_{r^0_t}-\boldsymbol{\beta}_{r_t})\|^2 \\
&= \|\mathbf{M}_{\widehat{F}}\mathbf{V}_t(\boldsymbol{\beta}^0_{r^0_t}-\boldsymbol{\beta}_{r_t})-\mathbf{M}_{\widehat{F}}\overline{\mathbf{UB}}_m\boldsymbol{\Gamma}_t(\boldsymbol{\beta}^0_{r^0_t}-\boldsymbol{\beta}_{r_t})\|^2 \\
&= \|\mathbf{M}_{\widehat{F}}\mathbf{V}_t(\boldsymbol{\beta}^0_{r^0_t}-\boldsymbol{\beta}_{r_t})\|^2+\|\mathbf{M}_{\widehat{F}}\overline{\mathbf{UB}}_m\boldsymbol{\Gamma}_t(\boldsymbol{\beta}^0_{r^0_t}-\boldsymbol{\beta}_{r_t})\|^2 \\
&\quad -2\mathrm{tr}\,[(\boldsymbol{\beta}^0_{r^0_t}-\boldsymbol{\beta}_{r_t})'\mathbf{V}_t'\mathbf{M}_{\widehat{F}}\overline{\mathbf{UB}}_m\boldsymbol{\Gamma}_t(\boldsymbol{\beta}^0_{r^0_t}-\boldsymbol{\beta}_{r_t})]. \qquad \text{(A.16)}
\end{aligned}$$

The same result can be applied again to show that

$$\begin{aligned}
\|\mathbf{M}_{\widehat{F}}\mathbf{V}_t(\boldsymbol{\beta}^0_{r^0_t}-\boldsymbol{\beta}_{r_t})\|^2 &= \|\mathbf{M}_{F^0}\mathbf{V}_t(\boldsymbol{\beta}^0_{r^0_t}-\boldsymbol{\beta}_{r_t})+(\mathbf{M}_{\widehat{F}}-\mathbf{M}_{F^0})\mathbf{V}_t(\boldsymbol{\beta}^0_{r^0_t}-\boldsymbol{\beta}_{r_t})\|^2 \\
&= \|\mathbf{M}_{F^0}\mathbf{V}_t(\boldsymbol{\beta}^0_{r^0_t}-\boldsymbol{\beta}_{r_t})\|^2+\|(\mathbf{M}_{\widehat{F}}-\mathbf{M}_{F^0})\mathbf{V}_t(\boldsymbol{\beta}^0_{r^0_t}-\boldsymbol{\beta}_{r_t})\|^2 \\
&\quad +2\mathrm{tr}\,[(\boldsymbol{\beta}^0_{r^0_t}-\boldsymbol{\beta}_{r_t})'\mathbf{V}_t'\mathbf{M}_{F^0}(\mathbf{M}_{\widehat{F}}-\mathbf{M}_{F^0})\mathbf{V}_t(\boldsymbol{\beta}^0_{r^0_t}-\boldsymbol{\beta}_{r_t})] \qquad \text{(A.17)}
\end{aligned}$$

The second term on the right-hand side of (A.14) can be written similarly using $\mathbf{F}'\mathbf{M}_{F^0}=\mathbf{0}_{m\times N}$ as

$$\begin{aligned}
\|\mathbf{M}_{\widehat{F}}(\mathbf{F}\boldsymbol{\gamma}_t+\boldsymbol{\varepsilon}_t)\|^2 &= \|\mathbf{M}_{F^0}\boldsymbol{\varepsilon}_t+(\mathbf{M}_{\widehat{F}}-\mathbf{M}_{F^0})(\mathbf{F}\boldsymbol{\gamma}_t+\boldsymbol{\varepsilon}_t)\|^2 \\
&= \|\mathbf{M}_{F^0}\boldsymbol{\varepsilon}_t\|^2+\|(\mathbf{M}_{\widehat{F}}-\mathbf{M}_{F^0})(\mathbf{F}\boldsymbol{\gamma}_t+\boldsymbol{\varepsilon}_t)\|^2 \\
&\quad +2\mathrm{tr}\,[\boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}(\mathbf{M}_{\widehat{F}}-\mathbf{M}_{F^0})(\mathbf{F}\boldsymbol{\gamma}_t+\boldsymbol{\varepsilon}_t)]. \qquad \text{(A.18)}
\end{aligned}$$

Inserting these results into $Q(\boldsymbol{\beta}, \mathbf{r})$ yields

$$\begin{aligned}
Q(\boldsymbol{\beta}, \mathbf{r}) &= \frac{1}{NT}\sum_{t=1}^{T} \|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{r_t})\|^2 \\
&= Q^0(\boldsymbol{\beta}, \mathbf{r}) + \frac{1}{NT}\sum_{t=1}^{T} \|(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\mathbf{V}_t(\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t})\|^2 \\
&+ \frac{2}{NT}\sum_{t=1}^{T} \mathrm{tr}\,[(\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t})'\mathbf{V}_t'\mathbf{M}_{F^0}(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\mathbf{V}_t(\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t})] \\
&+ \frac{1}{NT}\sum_{t=1}^{T} \|\mathbf{M}_{\widehat{F}}\overline{\mathbf{U}\mathbf{B}}_m\boldsymbol{\Gamma}_t(\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t})\|^2 \\
&- \frac{2}{NT}\sum_{t=1}^{T} \mathrm{tr}\,[(\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t})'\mathbf{V}_t'\mathbf{M}_{\widehat{F}}\overline{\mathbf{U}\mathbf{B}}_m\boldsymbol{\Gamma}_t(\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t})] \\
&+ \frac{1}{NT}\sum_{t=1}^{T} \|(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)\|^2 \\
&+ \frac{2}{NT}\sum_{t=1}^{T} \mathrm{tr}\,[\boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)] \\
&+ \frac{2}{NT}\sum_{t=1}^{T} \mathrm{tr}\,[(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t})] \\
&= Q^0(\boldsymbol{\beta}, \mathbf{r}) + \sum_{j=1}^{7} I_j,
\end{aligned} \tag{A.19}$$

with implicit definitions of $I_1, \ldots, I_7$. We now consider each of these seven terms.

Consider I_1. By the Cauchy–Schwarz inequality,

$$\begin{aligned}
I_1 &= \frac{1}{NT}\sum_{t=1}^{T} \|(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\mathbf{V}_t(\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t})\|^2 \\
&\leq \left(\frac{1}{T}\sum_{t=1}^{T} \|N^{-1/2}(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\mathbf{V}_t\|^4\right)^{1/2} \left(\frac{1}{T}\sum_{t=1}^{T} \|\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t}\|^4\right)^{1/4},
\end{aligned} \tag{A.20}$$

where $\|\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t}\| = O(1)$ since $\|\boldsymbol{\beta}_{r_t}\| < \infty$ by the compactness of $\mathcal{B}$. Consider $\mathbf{M}_{F^0} -$

$\mathbf{M}_{\widehat{F}}$. Since $\overline{\mathbf{B}}$ and $\mathbf{D}_T$ are invertible, we have $\mathbf{P}_{\widehat{F}} = \mathbf{P}_{\widehat{F}\overline{B}D_T} = \mathbf{P}_{\widehat{F}^0}$, which in turn implies

$$\begin{aligned}
\mathbf{M}_{F^0} - \mathbf{M}_{\widehat{F}} &= \mathbf{P}_{\widehat{F}^0} - \mathbf{P}_{F^0} \\
&= (\widehat{\mathbf{F}}^0 - \mathbf{F}^0)(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)' + (\widehat{\mathbf{F}}^0 - \mathbf{F}^0)(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\mathbf{F}^{0\prime} \\
&+ \mathbf{F}^0(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)' + \mathbf{F}^0[(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1} - (\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}]\mathbf{F}^{0\prime}.
\end{aligned} \tag{A.21}$$

By inserting this into $\|N^{-1/2}(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\mathbf{V}_t\|^2$, we obtain

$$\begin{aligned}
&\|N^{-1/2}(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\mathbf{V}_t\|^2 \\
&= N^{-1}|\text{tr}\,[\mathbf{V}_t'(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})'(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\mathbf{V}_t]| \\
&\leq N^{-1}|\text{tr}\,[\mathbf{V}_t'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'\mathbf{V}_t]| \\
&+ 2N^{-1}|\text{tr}\,[\mathbf{V}_t'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\mathbf{F}^{0\prime}\mathbf{V}_t]| \\
&+ 2N^{-1}|\text{tr}\,[\mathbf{V}_t'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'\mathbf{F}^0(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'\mathbf{V}_t]| \\
&+ 2N^{-1}|\text{tr}\,[\mathbf{V}_t'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'\mathbf{F}^0[(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1} - (\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}]\mathbf{F}^{0\prime}\mathbf{V}_t]| \\
&+ 2N^{-1}|\text{tr}\,[\mathbf{V}_t'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'\mathbf{F}^0(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'\mathbf{V}_t]| \\
&+ 2N^{-1}|\text{tr}\,[\mathbf{V}_t'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'\mathbf{V}_t]| \\
&+ 2N^{-1}|\text{tr}\,[\mathbf{V}_t'\mathbf{F}^0(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'\mathbf{F}^0(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'\mathbf{V}_t]| \\
&+ N^{-1}|\text{tr}\,[\mathbf{V}_t'\mathbf{F}^0(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\mathbf{F}^{0\prime}\mathbf{V}_t]| \\
&+ N^{-1}|\text{tr}\,[\mathbf{V}_t(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'\mathbf{V}_t]| \\
&+ N^{-1}|\text{tr}\,[\mathbf{F}^0[(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1} - (\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}]\mathbf{F}^{0\prime}\mathbf{F}^0[(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1} - (\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}]\mathbf{F}^{0\prime}]|.
\end{aligned} \tag{A.22}$$

We now make use of the fact that $\text{tr}\,(\mathbf{A}'\mathbf{B})^2 \leq \text{tr}\,(\mathbf{A}'\mathbf{A})\text{tr}\,(\mathbf{B}'\mathbf{B}) = \|\mathbf{A}\|^2\|\mathbf{B}\|^2$ if $\mathbf{A}$ and $\mathbf{B}$ are of the same dimension (see Abadir and Magnus, 2005, Exercise 12.5). This implies that $|\mathrm{tr}\,(\mathbf{A}'\mathbf{B})| \leq \|\mathbf{A}\|\|\mathbf{B}\|$, which we can use to show the following:

$$
\begin{aligned}
&\|N^{-1/2}(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\mathbf{V}_t\|^2 \\
&\leq \|N^{-1}\mathbf{V}_t'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\|^2\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|^2 N^{-1}\|\widehat{\mathbf{F}}^0 - \mathbf{F}^0\|^2 \\
&+ 2\|N^{-1}\mathbf{V}_t'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\|\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|^2 N^{-1}\|\widehat{\mathbf{F}}^0 - \mathbf{F}^0\|^2\|N^{-1}\mathbf{F}^{0\prime}\mathbf{V}_t\| \\
&+ 2\|N^{-1}\mathbf{V}_t'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\|^2\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|^2 N^{-1/2}\|\widehat{\mathbf{F}}^0 - \mathbf{F}^0\| N^{-1/2}\|\mathbf{F}^0\| \\
&+ 2\|N^{-1}\mathbf{V}_t'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\|\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\| N^{-1/2}\|\widehat{\mathbf{F}}^0 - \mathbf{F}^0\| N^{-1/2}\|\mathbf{F}^0\| \\
&\times \|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1} - (N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\|\|N^{-1}\mathbf{F}^{0\prime}\mathbf{V}_t\| \\
&+ 2\|N^{-1}\mathbf{V}_t'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\|^2\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|^2 N^{-1/2}\|\widehat{\mathbf{F}}^0 - \mathbf{F}^0\| N^{-1/2}\|\mathbf{F}^0\| \\
&+ 2\|N^{-1}\mathbf{V}_t'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\|^2\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|^2 N^{-1}\|\mathbf{F}^0\|^2 \\
&+ 2\|N^{-1}\mathbf{V}_t'\mathbf{F}^0\|\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|^2 N^{-1/2}\|\widehat{\mathbf{F}}^0 - \mathbf{F}^0\| N^{-1/2}\|\mathbf{F}^0\|\|N^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'\mathbf{V}_t\| \\
&+ \|N^{-1}\mathbf{V}_t'\mathbf{F}^0\|^2\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|^2 N^{-1}\|\widehat{\mathbf{F}}^0 - \mathbf{F}^0\|^2 \\
&+ \|N^{-1}\mathbf{V}_t(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\|^2\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|^2 N^{-1}\|\mathbf{F}^0\|^2 \\
&+ N^{-2}\|\mathbf{F}^0\|^4\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1} - (N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\|^2. \qquad \text{(A.23)}
\end{aligned}
$$

Consider $N^{-1}\mathbf{V}_t'\mathbf{F}^0 = [N^{-1}\mathbf{V}_t'\mathbf{F}, N^{-1}\sqrt{T}\mathbf{V}_t'\overline{\mathbf{U}\mathbf{B}}_{-m}]$. By the at-most-weak dependence of $\mathbf{v}_{i,t}$ over i (Assumption 1) and $\sup_{i,j}|\mathbf{f}_i'\mathbf{f}_j| \leq \sup_{i,j}\|\mathbf{f}_i\|\|\mathbf{f}_j\| \leq (\sup_i\|\mathbf{f}_i\|)^2 = O(1)$

(Assumption 1) similarly to the proof of Lemma A.1,

$$\begin{aligned}
\mathbb{E}(\|N^{-1}\mathbf{V}_t'\mathbf{F}\|^2) &= N^{-2}\mathbb{E}[\text{tr}\,(\mathbf{V}_t'\mathbf{F}\mathbf{F}'\mathbf{V}_t)] = N^{-2}\mathbb{E}[\text{tr}\,(\mathbf{V}_t'\mathbf{F}\mathbf{F}'\mathbf{V}_t)] \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}[\text{tr}\,(\mathbf{v}_{i,t}\mathbf{f}_i'\mathbf{f}_j\mathbf{v}_{j,t}')] = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbf{f}_i'\mathbf{f}_j\text{tr}\,[\mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}_{j,t}')] \\
&\leq \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}|\mathbf{f}_i'\mathbf{f}_j||\text{tr}\,[\mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}_{j,t}')]| \\
&\leq N^{-1}\sup_{i,j}|\mathbf{f}_i'\mathbf{f}_j|\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|\text{tr}\,[\mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}_{j,t}')]| \\
&\leq N^{-1}(\sup_{i}\|\mathbf{f}_i\|)^2\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|\text{tr}\,[\mathbb{E}(\mathbf{v}_{i,t}\mathbf{v}_{j,t}')]| = O(N^{-1}), \qquad \text{(A.24)}
\end{aligned}$$

which in turn implies $\|N^{-1}\mathbf{V}_t'\mathbf{F}\| = O_p(N^{-1/2})$. Next up is

$$N^{-1}\sqrt{T}\mathbf{V}_t'\overline{\mathbf{U}} = \frac{1}{N\sqrt{T}}\sum_{s=1}^{T}\mathbf{V}_t'\mathbf{U}_s = \frac{1}{N\sqrt{T}}\sum_{s=1}^{T}[\mathbf{V}_t'\boldsymbol{\varepsilon}_s + \mathbf{V}_t'\mathbf{V}_s\boldsymbol{\beta}_{r_s}, \mathbf{V}_t'\mathbf{V}_s]. \qquad \text{(A.25)}$$

Since $\boldsymbol{\varepsilon}_t$ and $\mathbf{V}_t$ are independent of one another, and at most weakly dependent over time and cross-sectional units by Assumption 1, we have $\|(NT)^{-1/2}\sum_{s=1}^{T}\mathbf{V}_t'\boldsymbol{\varepsilon}_s\| = O_p(1)$. The term $N^{-1}T^{-1/2}\sum_{s=1}^{T}\mathbf{V}_t'\boldsymbol{\varepsilon}_s$ in $N^{-1}\sqrt{T}\mathbf{V}_t'\overline{\mathbf{U}}$ is therefore negligible. However, this is generally not true for the other terms. In order to appreciate this, note how

$$\frac{1}{N\sqrt{T}}\sum_{s=1}^{T}\mathbf{V}_t'\mathbf{V}_s = \frac{1}{N\sqrt{T}}\sum_{i=1}^{N}\mathbf{v}_{i,t}\mathbf{v}_{i,t}' + \frac{1}{N\sqrt{T}}\sum_{s\neq t}\sum_{i=1}^{N}\mathbf{v}_{i,t}\mathbf{v}_{i,s}'. \qquad \text{(A.26)}$$

While the first term on the right-hand side is clearly $O_p(T^{-1/2})$, and hence negligible, the second term is not. By imposing additional conditions on the cross-sectional and serial dependencies of $\mathbf{v}_{i,t}$ also the second term can be made negligible. One possibility is to require that $\mathbf{v}_{i,t}$ is serially uncorrelated, in which case the second term will be $O_p(N^{-1/2})$. In the present lemma, however, we only require that $\mathbf{v}_{i,t}$ is at most weakly correlated, which means that

$$\left\|\frac{1}{N\sqrt{T}}\sum_{s=1}^{T}\mathbf{V}_t'\mathbf{V}_s\right\| = \left\|\frac{1}{N\sqrt{T}}\sum_{s\neq t}\sum_{i=1}^{N}\mathbf{v}_{i,t}\mathbf{v}_{i,s}'\right\| + O_p(T^{-1/2}) = O_p(1). \tag{A.27}$$

This implies $\|N^{-1}\sqrt{T}\mathbf{V}_t'\overline{\mathbf{U}}\| = O_p(1)$, and so

$$\|N^{-1}\mathbf{V}_t'\mathbf{F}^0\| = O_p(1). \tag{A.28}$$

We want to point out that this order need not be exact. As already pointed out, and as we exploit further in Proof of Lemma A.3, one exception occurs when $\mathbf{v}_{i,t}$ is serially uncorrelated, in which case $\|N^{-1}\mathbf{V}_t'\mathbf{F}^0\| = O_p(N^{-1/2}) + O_p(T^{-1/2})$. Another possibility is that $m = k+1$, in which case $\|N^{-1}\mathbf{V}_t'\mathbf{F}^0\| = \|N^{-1}\mathbf{V}_t'\mathbf{F}\| = O_p(N^{-1/2})$. In this proof, however, we treat $\|N^{-1}\mathbf{V}_t'\mathbf{F}^0\|$ as $O_p(1)$. From $\|N^{-1}\sqrt{T}\mathbf{V}_t'\overline{\mathbf{U}}\| = O_p(1)$ and $\widehat{\mathbf{F}}^0 - \mathbf{F}^0 = [\overline{\mathbf{U}}\overline{\mathbf{B}}_m, \mathbf{0}_{N\times(k+1-m)}]$ (Lemma A.1), we obtain

$$\|N^{-1}\mathbf{V}_t'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\| \leq \|N^{-1}\mathbf{V}_t'\overline{\mathbf{U}}\|\|\overline{\mathbf{B}}_m\| = O_p(T^{-1/2}). \tag{A.29}$$

Another term in (A.23) is $N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0$. The steps used earlier to evaluate $E(\|N^{-1}\mathbf{V}_t'\mathbf{F}\|^2)$ can be applied also to $E(\|\sqrt{T}N^{-1}\mathbf{F}'\overline{\mathbf{U}}\|^2)$. This yields $\|\sqrt{T}N^{-1}\mathbf{F}'\overline{\mathbf{U}}\| = O_p(N^{-1/2})$.

In order to characterize the limit of $TN^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}}$ it is useful to introduce the indicator function $\mathbb{1}(A)$, which takes on the value 1 if the event A is true and 0 other wise. In this notation, $\sum_{t=1}^{T}\boldsymbol{\beta}_{r_t} = \sum_{t=1}^{T}\sum_{r=1}^{R}\mathbb{1}(r_t = r)\boldsymbol{\beta}_r$ and hence $\sum_{t=1}^{T}\mathbf{Q}_{r_t}'\mathbf{w}_{i,t} = \sum_{t=1}^{T}\sum_{r=1}^{R}\mathbb{1}(r_t =$

$r)\mathbf{Q}_r'\mathbf{w}_{i,t}$. This implies

$$
\begin{aligned}
TN^{-1}\overline{\mathbf{U}}'\overline{\mathbf{U}} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{u}_{i,t}\mathbf{u}_{i,s}' = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbf{Q}_{r_t}'\mathbf{w}_{i,t}\mathbf{w}_{i,s}'\mathbf{Q}_{r_s} \\
&= \sum_{r=1}^{R}\sum_{r'=1}^{R}\mathbf{Q}_r'\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{1}(r_t=r)\mathbb{1}(r_s=r')\mathbf{w}_{i,t}\mathbf{w}_{i,s}'\right)\mathbf{Q}_{r'} \\
&\to_p \sum_{r=1}^{R}\sum_{r'=1}^{R}\mathbf{Q}_r'\left(\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{1}(r_t=r)\mathbb{1}(r_s=r')E(\mathbf{w}_{i,t}\mathbf{w}_{i,s}')\right)\mathbf{Q}_{r'} \\
&=: \mathbf{\Omega}_u
\end{aligned} \tag{A.30}
$$

as N, $T\to\infty$, where $\mathbf{\Omega}_u$ is positive definite by Assumption 1. Hence, since $N^{-1}\mathbf{F}'\mathbf{F} = N^{-1}\sum_{i=1}^{N}\mathbf{f}_i\mathbf{f}_i' \to \mathbf{\Sigma}_F$ as $N\to\infty$, where $\mathbf{\Sigma}_F$ is positive definite too (Assumption 1),

$$
\begin{aligned}
N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0 &= \begin{bmatrix} N^{-1}\mathbf{F}'\mathbf{F} & \sqrt{T}N^{-1}\mathbf{F}'\overline{\mathbf{U}}\overline{\mathbf{B}}_{-m} \\ \sqrt{T}N^{-1}\overline{\mathbf{B}}_{-m}'\overline{\mathbf{U}}'\mathbf{F} & TN^{-1}\overline{\mathbf{B}}_{-m}'\overline{\mathbf{U}}'\overline{\mathbf{U}}\overline{\mathbf{B}}_{-m} \end{bmatrix} \\
&= \begin{bmatrix} N^{-1}\mathbf{F}'\mathbf{F} & \mathbf{0}_{m\times(k+1-m)} \\ \mathbf{0}_{(k+1-m)\times m} & TN^{-1}\overline{\mathbf{B}}_{-m}'\overline{\mathbf{U}}'\overline{\mathbf{U}}\overline{\mathbf{B}}_{-m} \end{bmatrix} + O_p(N^{-1/2}) \\
&\to_p \begin{bmatrix} \mathbf{\Sigma}_F & \mathbf{0}_{m\times(k+1-m)} \\ \mathbf{0}_{(k+1-m)\times m} & \mathbf{\Omega}_u \end{bmatrix}
\end{aligned} \tag{A.31}
$$

as N, $T\to\infty$. The limiting block-diagonal matrix here is positive definite because the blocks are positive definite (see Abadir and Magnus, 2005, Exercise 8.44). This implies that $N^{-1}\|\mathbf{F}^0\|^2 = O_p(1)$ and $\|(N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\| = O_p(1)$. We also need $\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1} - (N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\|$. From

$$
\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0 = \mathbf{F}^{0\prime}\mathbf{F}^0 + (\widehat{\mathbf{F}}^0-\mathbf{F}^0)'\mathbf{F}^0 + \mathbf{F}^{0\prime}(\widehat{\mathbf{F}}^0-\mathbf{F}^0) + (\widehat{\mathbf{F}}^0-\mathbf{F}^0)'(\widehat{\mathbf{F}}^0-\mathbf{F}^0), \tag{A.32}
$$

and using Lemma A.1, we obtain

$$N^{-1}\|\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0 - \mathbf{F}^{0\prime}\mathbf{F}^0\| \leq 2N^{-1/2}\|\widehat{\mathbf{F}}^0 - \mathbf{F}^0\|N^{-1/2}\|\mathbf{F}^0\| + N^{-1}\|\widehat{\mathbf{F}}^0 - \mathbf{F}^0\|^2 = O_p(T^{-1/2}), \tag{A.33}$$

which in view of

$$(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1} - (\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1} = (\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\mathbf{F}^{0\prime}\mathbf{F}^0 - \widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)(\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1} \tag{A.34}$$

leads to the following:

$$\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1} - (N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\| \leq \|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|N^{-1}\|\mathbf{F}^{0\prime}\mathbf{F}^0 - \widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0\|\|(N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\| = O_p(T^{-1/2}). \tag{A.35}$$

We now have all the pieces we need in order to evaluate (A.23). The dominating terms are given by $\|N^{-1}\mathbf{V}_t(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\|^2\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|^2N^{-1}\|\mathbf{F}^0\|^2$ and $N^{-2}\|\mathbf{F}^0\|^4\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1} - (N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\|^2$, which are both $O_p(T^{-1})$. It follows that

$$\|N^{-1/2}(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\mathbf{V}_t\|^2 = O_p(T^{-1}), \tag{A.36}$$

which in turn implies that I_1 is of the same order;

$$I_1 \leq \left(\frac{1}{T}\sum_{t=1}^{T}\|N^{-1/2}(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\mathbf{V}_t\|^4\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t}\|^4\right)^{1/4} = O_p(T^{-1}). \tag{A.37}$$

For I_2 we again use (A.21) followed by $\mathbf{M}_{F^0}\mathbf{F}^0 = \mathbf{0}_{N\times(k+1-m)}$, giving

$$\begin{aligned}
&\|N^{-1}\mathbf{V}_t'\mathbf{M}_{F^0}(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\mathbf{V}_t\| \\
&= \|N^{-1}\mathbf{V}_t'\mathbf{M}_{F^0}[(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)' + (\widehat{\mathbf{F}}^0 - \mathbf{F}^0)(\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\mathbf{F}^{0\prime}]\mathbf{V}_t\| \\
&\leq \|N^{-1}\mathbf{V}_t'\mathbf{M}_{F^0}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\|\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|\|N^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'\mathbf{V}_t\| \\
&+ \|N^{-1}\mathbf{V}_t'\mathbf{M}_{F^0}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\|\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|\|N^{-1}\mathbf{F}^{0\prime}\mathbf{V}_t\|, \qquad (A.38)
\end{aligned}$$

where all terms are known, except for

$$\begin{aligned}
&\|N^{-1}\mathbf{V}_t'\mathbf{M}_{F^0}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\| \\
&\leq \|N^{-1}\mathbf{V}_t'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\| + \|N^{-1}\mathbf{V}_t'\mathbf{F}^0\|\|(N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\|N^{-1/2}\|\mathbf{F}^0\|N^{-1/2}\|\widehat{\mathbf{F}}^0 - \mathbf{F}^0\| \\
&= O_p(T^{-1/2}). \qquad (A.39)
\end{aligned}$$

Hence, since $\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|$, $\|N^{-1}\mathbf{F}^{0\prime}\mathbf{V}_t\|$ are both $O_p(1)$, while $\|N^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'\mathbf{V}_t\| = O_p(T^{-1/2})$, we have

$$\|N^{-1}\mathbf{V}_t'\mathbf{M}_{F^0}(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\mathbf{V}_t\| = O_p(T^{-1/2}). \qquad (A.40)$$

By using this last result, $\|\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t}\| = O(1)$ and $|\mathrm{tr}\,(\mathbf{A}'\mathbf{B})| \leq \|\mathbf{A}\|\|\mathbf{B}\|$, we arrive at the following bound for $|I_2|$:

$$
\begin{aligned}
|I_2| &= \left| \frac{2}{NT} \sum_{t=1}^{T} \mathrm{tr}\, [(\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t})' \mathbf{V}_t' \mathbf{M}_{F^0} (\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0}) \mathbf{V}_t (\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t})] \right| \\
&\leq \frac{2}{NT} \sum_{t=1}^{T} |\mathrm{tr}\, [(\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t})(\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t})' \mathbf{V}_t' \mathbf{M}_{F^0} (\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0}) \mathbf{V}_t]| \\
&\leq \frac{2}{NT} \sum_{t=1}^{T} \|\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t}\|^2 \|\mathbf{V}_t' \mathbf{M}_{F^0} (\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0}) \mathbf{V}_t\| \\
&\leq 2 \left(\frac{1}{T} \sum_{t=1}^{T} \|\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t}\|^4 \right)^{1/2} \left(\frac{1}{T} \sum_{t=1}^{T} \|N^{-1} \mathbf{V}_t' \mathbf{M}_{F^0} (\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0}) \mathbf{V}_t\|^2 \right)^{1/2} \\
&= O_p(T^{-1/2}).
\end{aligned} \tag{A.41}
$$

For I_3, we use $(a+b)^2 \leq 2(a^2+b^2)$, which together with $\|N^{-1/2}\overline{\mathbf{U}}\| = O_p(T^{-1/2})$ imply

$$
\begin{aligned}
\|N^{-1/2}\mathbf{M}_{F^0}\overline{\mathbf{U}}\|^2 &\leq 2\|N^{-1/2}\overline{\mathbf{U}}\|^2 + 2\|N^{-1/2}\mathbf{P}_{F^0}\overline{\mathbf{U}}\|^2 \\
&= 2\|N^{-1/2}\overline{\mathbf{U}}\|^2 + 2\mathrm{tr}\, [N^{-1}\overline{\mathbf{U}}'\mathbf{F}^0(N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}N^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}] \\
&\leq 2\|N^{-1/2}\overline{\mathbf{U}}\|^2 + 2\|N^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\|^2\|(N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\| \\
&\leq 2\|N^{-1/2}\overline{\mathbf{U}}\|^2 + 2\|N^{-1/2}\overline{\mathbf{U}}\|^2\|N^{-1/2}\mathbf{F}^0\|^2\|(N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\| \\
&= O_p(T^{-1}).
\end{aligned} \tag{A.42}
$$

The term $\|N^{-1/2}(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\overline{\mathbf{U}}\|^2$ is of the same order, which is easily appreciated using the above given expansion of $\|N^{-1/2}(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\mathbf{V}_t\|^2$. We can therefore show that

$$
\|N^{-1/2}\mathbf{M}_{\widehat{F}}\overline{\mathbf{U}}\|^2 \leq 2\|N^{-1/2}\mathbf{M}_{F^0}\overline{\mathbf{U}}\|^2 + 2\|N^{-1/2}(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\overline{\mathbf{U}}\|^2 = O_p(T^{-1}), \tag{A.43}
$$

so that

$$
\begin{aligned}
I_3 &= \frac{1}{NT}\sum_{t=1}^{T} \|\mathbf{M}_{\widehat{F}}\overline{\mathbf{U}}\overline{\mathbf{B}}_m \boldsymbol{\Gamma}_t(\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t})\|^2 \\
&\le \|N^{-1/2}\mathbf{M}_{\widehat{F}}\overline{\mathbf{U}}\|^2 \|\overline{\mathbf{B}}_m\|^2 \left(\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{\Gamma}_t\|^4\right)^{1/2} \left(\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t}\|^4\right)^{1/2} \\
&= O_p(T^{-1}). \qquad \text{(A.44)}
\end{aligned}
$$

The evaluations of I_4, I_5 and I_6 is similar to those of I_1, I_2 and I_3. $|I_4|$ and $|I_6|$ is of the same order as $|I_2|$, by exactly the same arguments, and I_5 is of the same order as I_1. Only I_7 remains. We start by writing

$$
\begin{aligned}
I_7 &= \frac{2}{NT}\sum_{t=1}^{T} \mathrm{tr}\left[(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t})\right] \\
&= \frac{2}{NT}\sum_{t=1}^{T} \boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}\mathbf{V}_t(\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t}) \\
&- \frac{2}{NT}\sum_{t=1}^{T} (\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\mathbf{V}_t(\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t}) \\
&- \frac{2}{NT}\sum_{t=1}^{T} \boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}\overline{\mathbf{U}}\overline{\mathbf{B}}_m\boldsymbol{\Gamma}_t(\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t}) \\
&+ \frac{2}{NT}\sum_{t=1}^{T} (\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\overline{\mathbf{U}}\overline{\mathbf{B}}_m\boldsymbol{\Gamma}_t(\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t}). \qquad \text{(A.45)}
\end{aligned}
$$

The first term on the right dominates (see Proof of Lemma A.3). When evaluating this term we follow Bonhomme and Manresa (2015, Proof of Theorem 1), and split it in two, one part that involves $\boldsymbol{\beta}^0_{r^0_t}$ and one that involves $\boldsymbol{\beta}_{r_t}$, and we evaluate the sum regime-wise. Note first how

$$
\begin{aligned}
\left|\frac{1}{NT}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}\mathbf{V}_t\boldsymbol{\beta}_{r_t}\right| &= \left|\frac{1}{NT}\sum_{r=1}^{R}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}\mathbf{V}_t 1(r_t=r)\boldsymbol{\beta}_r\right| \\
&\leq \left(\sum_{r=1}^{R}\left\|\frac{1}{NT}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}\mathbf{V}_t 1(r_t=r)\right\|^2\right)^{1/2}\left(\sum_{r=1}^{R}\|\boldsymbol{\beta}_r\|^2\right)^{1/2} \\
&\leq \left(\sum_{r=1}^{R}\left\|\frac{1}{NT}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}\mathbf{V}_t\right\|^2\right)^{1/2}\left(\sum_{r=1}^{R}\|\boldsymbol{\beta}_r\|^2\right)^{1/2} \\
&\leq \sqrt{R}\left\|\frac{1}{NT}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}\mathbf{V}_t\right\|\left(\sum_{r=1}^{R}\|\boldsymbol{\beta}_r\|^2\right)^{1/2}, \qquad \text{(A.46)}
\end{aligned}
$$

where $\sum_{r=1}^{R}\|\boldsymbol{\beta}_r\|^2 = O(1)$ by the compactness of $\mathcal{B}$. We also have

$$
\left\|\frac{1}{NT}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}\mathbf{V}_t\right\| \leq \left\|\frac{1}{NT}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t'\mathbf{V}_t\right\| + \left\|\frac{1}{NT}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t'\mathbf{F}^0(\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\mathbf{F}^{0\prime}\mathbf{V}_t\right\|, \qquad \text{(A.47)}
$$

where the first term on the right is clearly $O_p((NT)^{-1/2})$ since $\boldsymbol{\varepsilon}_t$ and $\mathbf{V}_t$ are mean zero, independent of one another, and at most weakly dependent over both time and the cross-section (Assumption 1). For the second term, we use the fact that $\|\text{vec}\,\mathbf{A}\|^2 = (\text{vec}\,\mathbf{A})'\text{vec}\,\mathbf{A} = \text{tr}\,(\mathbf{A}'\mathbf{A}) = \|\mathbf{A}\|^2$ by Exercise 10.20 in Abadir and Magnus (2005). By using this, $\text{vec}\,(\mathbf{ABC}) = (\mathbf{C}'\otimes\mathbf{A})\text{vec}\,\mathbf{B}$ (see Abadir and Magnus, Exercise 10.18) and

$\|\mathbf{F}^0(\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\mathbf{F}^{0\prime}\|^2 = \operatorname{tr}\mathbf{I}_m = m$, we obtain

$$\begin{aligned}
\left\|\frac{1}{NT}\sum_{t=1}^{T}\mathbf{V}_t'\mathbf{F}^0(\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\mathbf{F}^{0\prime}\boldsymbol{\varepsilon}_t\right\| &= \left\|\operatorname{vec}\left(\frac{1}{NT}\sum_{t=1}^{T}\mathbf{V}_t'\mathbf{F}^0(\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\mathbf{F}^{0\prime}\boldsymbol{\varepsilon}_t\right)\right\| \\
&= \left\|\frac{1}{NT}\sum_{t=1}^{T}(\boldsymbol{\varepsilon}_t\otimes\mathbf{V}_t)'\operatorname{vec}[\mathbf{F}^0(\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\mathbf{F}^{0\prime}]\right\| \\
&\leq \left\|\frac{1}{NT}\sum_{t=1}^{T}(\boldsymbol{\varepsilon}_t\otimes\mathbf{V}_t)\right\|\|\operatorname{vec}[\mathbf{F}^0(\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\mathbf{F}^{0\prime}]\| \\
&\leq \left\|\frac{1}{NT}\sum_{t=1}^{T}(\boldsymbol{\varepsilon}_t\otimes\mathbf{V}_t)\right\|\|\mathbf{F}^0(\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\mathbf{F}^{0\prime}\| \\
&= \sqrt{m}\left\|\frac{1}{NT}\sum_{t=1}^{T}(\boldsymbol{\varepsilon}_t\otimes\mathbf{V}_t)\right\|, \qquad \text{(A.48)}
\end{aligned}$$

where by Assumption 1 and a number of basic rules for Kronecker products,

$$\begin{aligned}
\mathbb{E}\left(\left\|\frac{1}{NT}\sum_{t=1}^{T}(\boldsymbol{\varepsilon}_t\otimes\mathbf{V}_t)\right\|^2\right) &= \mathbb{E}\left(\operatorname{tr}\left[\frac{1}{NT}\sum_{t=1}^{T}(\boldsymbol{\varepsilon}_t\otimes\mathbf{V}_t)\left(\frac{1}{NT}\sum_{t=1}^{T}(\boldsymbol{\varepsilon}_t\otimes\mathbf{V}_t)\right)'\right]\right) \\
&= \mathbb{E}\left(\operatorname{tr}\left[\left(\frac{1}{NT}\sum_{t=1}^{T}(\boldsymbol{\varepsilon}_t\otimes\mathbf{V}_t)'\right)\frac{1}{NT}\sum_{t=1}^{T}(\boldsymbol{\varepsilon}_t\otimes\mathbf{V}_t)\right]\right) \\
&= \operatorname{tr}\left(\frac{1}{(NT)^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{E}(\boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_s\otimes\mathbf{V}_t'\mathbf{V}_s)\right) \\
&= T^{-1}\operatorname{tr}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}[N^{-1}\mathbb{E}(\boldsymbol{\varepsilon}_t'\boldsymbol{\varepsilon}_s)\otimes N^{-1}E(\mathbf{V}_t'\mathbf{V}_s)]\right) \\
&= O(T^{-1}) \qquad \text{(A.49)}
\end{aligned}$$

Hence, since the variance is $O(T^{-1})$, we have $\|(NT)^{-1}\sum_{t=1}^{T}(\boldsymbol{\varepsilon}_t\otimes\mathbf{V}_t)\| = O_p(T^{-1/2})$, which in turn implies

$$\left\|\frac{1}{NT}\sum_{t=1}^{T}\mathbf{V}_t'\mathbf{F}^0(\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\mathbf{F}^{0\prime}\boldsymbol{\varepsilon}_t\right\| = O_p(T^{-1/2}) \qquad \text{(A.50)}$$

By adding the results,

$$\left\| \frac{1}{NT} \sum_{t=1}^{T} \boldsymbol{\varepsilon}_t' \mathbf{M}_{F^0} \mathbf{V}_t \right\| \leq \left\| \frac{1}{NT} \sum_{t=1}^{T} \boldsymbol{\varepsilon}_t' \mathbf{V}_t \right\| + \left\| \frac{1}{NT} \sum_{t=1}^{T} \boldsymbol{\varepsilon}_t' \mathbf{F}^0 (\mathbf{F}^{0\prime} \mathbf{F}^0)^{-1} \mathbf{F}^{0\prime} \mathbf{V}_t \right\|$$
$$= O_p((NT)^{-1/2}) + O_p(T^{-1/2}) = O_p(T^{-1/2}), \tag{A.51}$$

and so

$$\left| \frac{1}{NT} \sum_{t=1}^{T} \boldsymbol{\varepsilon}_t' \mathbf{M}_{F^0} \mathbf{V}_t \boldsymbol{\beta}_{r_t} \right| \leq \sqrt{R} \left\| \frac{1}{NT} \sum_{t=1}^{T} \boldsymbol{\varepsilon}_t' \mathbf{M}_{F^0} \mathbf{V}_t \right\| \left(\sum_{r=1}^{R} \| \boldsymbol{\beta}_r \|^2 \right)^{1/2} = O_p(T^{-1/2}). \tag{A.52}$$

Clearly, this last result holds also if $\boldsymbol{\beta}_{r_t} = \boldsymbol{\beta}^0_{r_t^0}$. It follows that

$$\left| \frac{1}{NT} \sum_{t=1}^{T} \boldsymbol{\varepsilon}_t' \mathbf{M}_{F^0} \mathbf{V}_t (\boldsymbol{\beta}^0_{r_t^0} - \boldsymbol{\beta}_{r_t}) \right| \leq \left| \frac{1}{NT} \sum_{t=1}^{T} \boldsymbol{\varepsilon}_t' \mathbf{M}_{F^0} \mathbf{V}_t \boldsymbol{\beta}^0_{r_t^0} \right| + \left| \frac{1}{NT} \sum_{t=1}^{T} \boldsymbol{\varepsilon}_t' \mathbf{M}_{F^0} \mathbf{V}_t \boldsymbol{\beta}_{r_t} \right|$$
$$= O_p(T^{-1/2}), \tag{A.53}$$

which in turn implies that $|I_7|$ is of the same order.

By adding the results, (A.19) becomes

$$Q(\boldsymbol{\beta}, \mathbf{r}) = Q^0(\boldsymbol{\beta}, \mathbf{r}) + O_p(T^{-1/2}), \tag{A.54}$$

where the order of the reminder is uniform in $(\boldsymbol{\beta}, \mathbf{r}) \in \mathcal{B}^R \times \mathcal{R}^T$.

The above results require that $N \to \infty$. This is not necessary. In fact, having N fixed only affects the results insofar that $N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0$ is no longer asymptotically block diagonal. However, it is still $O_p(1)$, as is its inverse, which is what matters in the end. ■

Lemma A.3. *Suppose that Assumption 1 holds. Then, as $T \to \infty$ with $N < \infty$ or $N \to \infty$,*

$$\min_{r^0 \in \mathcal{R}} \max_{\widehat{r} \in \mathcal{R}} \|\widehat{\boldsymbol{\beta}}_{\widehat{r}} - \boldsymbol{\beta}^0_{r^0}\|^2 = O_p(T^{-1/2}).$$

Proof: We will prove that $\max_{r \in \mathcal{R}} \min_{\widehat{r} \in \mathcal{R}} \|\boldsymbol{\beta}^0_r - \widehat{\boldsymbol{\beta}}_{\widehat{r}}\|^2 = O_p(T^{-1/2})$ by considering the following inequality:

$$Q(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{r}}) - Q(\boldsymbol{\beta}^0, \mathbf{r}^0) \leq 0, \tag{A.55}$$

which holds because, as already pointed out, $(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{r}})$ minimises $Q(\boldsymbol{\beta}, \mathbf{r})$. We employ contradiction to show that, unless $\max_{r \in \mathcal{R}} \min_{\widehat{r} \in \mathcal{R}} \|\boldsymbol{\beta}^0_r - \widehat{\boldsymbol{\beta}}_{\widehat{r}}\|^2 = O_p(T^{-1/2})$, the aforementioned inequality will not hold.

We begin by noting that from the definition of $Q^0(\boldsymbol{\beta}, \mathbf{r})$, we have

$$Q^0(\boldsymbol{\beta}^0, \mathbf{r}^0) = \frac{1}{NT} \sum_{t=1}^{T} \|\mathbf{M}_{F^0} \boldsymbol{\varepsilon}_t\|^2, \tag{A.56}$$

and hence

$$\begin{aligned}
&Q^0(\boldsymbol{\beta}, \mathbf{r}) - Q^0(\boldsymbol{\beta}^0, \mathbf{r}^0) \\
&= \frac{1}{NT} \sum_{t=1}^{T} \|\mathbf{M}_{F^0} \mathbf{V}_t (\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t})\|^2 \\
&= \frac{1}{NT} \sum_{t=1}^{T} \operatorname{tr}\left[(\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t})' \mathbf{V}_t' \mathbf{M}_{F^0} \mathbf{V}_t (\boldsymbol{\beta}^0_{r^0_t} - \boldsymbol{\beta}_{r_t})\right] \\
&= \sum_{r^0=1}^{R} \sum_{r=1}^{R} \operatorname{tr}\left[(\boldsymbol{\beta}^0_{r^0} - \boldsymbol{\beta}_r)' \frac{1}{NT} \sum_{t=1}^{T} \mathbb{1}(r_t = r') \mathbb{1}(r^0_t = r) \mathbf{V}_t' \mathbf{M}_{F^0} \mathbf{V}_t (\boldsymbol{\beta}^0_r - \boldsymbol{\beta}_{r'})\right] \\
&= \sum_{r=1}^{R} \sum_{r'=1}^{R} \operatorname{tr}\left[(\boldsymbol{\beta}^0_r - \boldsymbol{\beta}_{r'})' \mathbf{S}_{r',r} (\boldsymbol{\beta}^0_r - \boldsymbol{\beta}_{r'})\right],
\end{aligned} \tag{A.57}$$

where $\mathbf{S}_{r',r} := (NT)^{-1} \sum_{t=1}^{T} \mathbb{1}(r_t' = r') \mathbb{1}(r_t = r) \mathbf{V}_t' \mathbf{M}_{F^0} \mathbf{V}_t$. Suppose that $\mathbf{A}$ and $\mathbf{B}$ are positive semi-definite matrices. Suppose also that $\mathbf{A}$ is symmetric. Then, $\lambda_{\min}(\mathbf{A})\text{tr}\,\mathbf{B} \leq \text{tr}\,(\mathbf{AB}) \leq \lambda_{\max}(\mathbf{A})\text{tr}\,\mathbf{B}$, where $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ are the smallest and largest eigenvalues of $\mathbf{A}$, respectively (see Fang et al., 1994, Theorem 1). Applying this to $\text{tr}\,[(\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'})'\mathbf{S}_{r',r}(\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'})]$ yields

$$\begin{aligned}\text{tr}\,[(\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'})(\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'})'\mathbf{S}_{r',r}] &\geq \lambda_{\min}(\mathbf{S}_{r',r})\text{tr}\,[(\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'})(\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'})'] \\ &= \lambda_{\min}(\mathbf{S}_{r',r})\|\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'}\|^2, \end{aligned} \tag{A.58}$$

so that w.p.a.1,

$$\begin{aligned} Q^0(\boldsymbol{\beta}, \mathbf{r}) - Q^0(\boldsymbol{\beta}^0, \mathbf{r}^0) &\geq \sum_{r=1}^{R}\sum_{r'=1}^{R} \lambda_{\min}(\mathbf{S}_{r',r})\|\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'}\|^2 \\ &\geq \max_{r\in\mathcal{R}} \sum_{r'=1}^{R} \lambda_{\min}(\mathbf{S}_{r',r})\|\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'}\|^2 \\ &\geq \max_{r\in\mathcal{R}} \left(\min_{r'\in\mathcal{R}} \|\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'}\|^2\right) \sum_{r'=1}^{R} \lambda_{\min}(\mathbf{S}_{r',r}) \\ &\geq \max_{r\in\mathcal{R}} \left(\min_{r'\in\mathcal{R}} \|\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'}\|^2\right) \max_{r'\in\mathcal{R}} \lambda_{\min}(\mathbf{S}_{r',r}) \\ &\geq \max_{r\in\mathcal{R}} \left(\min_{r'\in\mathcal{R}} \|\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'}\|^2\right) \rho_{\min}, \end{aligned} \tag{A.59}$$

where $\rho_{\min} > 0$ by Assumption 1. Hence, invoking Lemma A.2, we have

$$\begin{aligned} Q(\boldsymbol{\beta}, \mathbf{r}) - Q(\boldsymbol{\beta}^0, \mathbf{r}^0) &= Q^0(\boldsymbol{\beta}, \mathbf{r}) - Q^0(\boldsymbol{\beta}^0, \mathbf{r}^0) + O_p(T^{-1/2}) \\ &\geq \rho_{\min} \max_{r\in\mathcal{R}} \min_{r'\in\mathcal{R}} \|\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'}\|^2 + O_p(T^{-1/2}). \end{aligned} \tag{A.60}$$

This holds for all $(\boldsymbol{\beta}, \mathbf{r}) \in \mathcal{B}^R \times \mathcal{R}^T$, including $(\boldsymbol{\beta}, \mathbf{r}) = (\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{r}})$. We can therefore show that

$$Q(\widehat{\boldsymbol{\beta}},\widehat{\mathbf{r}}) - Q(\boldsymbol{\beta}^0,\mathbf{r}^0) \geq \rho_{\min} \max_{r\in\mathcal{R}} \min_{\widehat{r}\in\mathcal{R}} \|\boldsymbol{\beta}_r^0 - \widehat{\boldsymbol{\beta}}_{\widehat{r}}\|^2 + O_p(T^{-1/2}) \geq 0. \tag{A.61}$$

Because $\rho_{\min} > 0$ w.p.a.1, the limit of $Q(\widehat{\boldsymbol{\beta}},\widehat{\mathbf{r}}) - Q(\boldsymbol{\beta}^0,\mathbf{r}^0)$ will be positive if first term on the right-hand side dominates, which will be the case if $\max_{r\in\mathcal{R}} \min_{\widehat{r}\in\mathcal{R}} \|\boldsymbol{\beta}_{r^0}^0 - \widehat{\boldsymbol{\beta}}_{\widehat{r}}\|^2 > O_p(T^{-1/2})$. But then this would contradict (A.55). The only way in which (A.55) can be satisfied asymptotically is therefore if $\max_{r\in\mathcal{R}} \min_{\widehat{r}\in\mathcal{R}} \|\boldsymbol{\beta}_r^0 - \widehat{\boldsymbol{\beta}}_{\widehat{r}}\|^2 \leq O_p(T^{-1/2})$. This means that

$$Q(\widehat{\boldsymbol{\beta}},\widehat{\mathbf{r}}) - Q(\boldsymbol{\beta}^0,\mathbf{r}^0) \to_p 0, \tag{A.62}$$

so that asymptotically (A.55) is satisfied with strict equality. This completes the proof.■

Lemma A.3 establishes consistency in a minimax sense. Consistency in mean square can be established in the same way, although under sightly different conditions. In particular, suppose that instead of $\rho_{\min} > 0$ w.p.a.1 we have $\min_t \lambda_{\min}(N^{-1}\mathbf{V}_t'\mathbf{M}_{F^0}\mathbf{V}_t) > 0$ w.p.a.1. Then, by using exactly the same steps as in Proof of Lemma A.3 above,

$$\begin{aligned} Q^0(\boldsymbol{\beta},\mathbf{r}) - Q^0(\boldsymbol{\beta}^0,\mathbf{r}^0) &= \frac{1}{T}\sum_{t=1}^{T} \operatorname{tr}\left[(\boldsymbol{\beta}_{r_t^0}^0 - \boldsymbol{\beta}_{r_t})' N^{-1}\mathbf{V}_t'\mathbf{M}_{F^0}\mathbf{V}_t(\boldsymbol{\beta}_{r_t^0}^0 - \boldsymbol{\beta}_{r_t})\right] \\ &\geq \frac{1}{T}\sum_{t=1}^{T} \lambda_{\min}(N^{-1}\mathbf{V}_t'\mathbf{M}_{F^0}\mathbf{V}_t)\|\boldsymbol{\beta}_{r_t^0}^0 - \boldsymbol{\beta}_{r_t}\|^2 \\ &\geq \min_t \lambda_{\min}(N^{-1}\mathbf{V}_t'\mathbf{M}_{F^0}\mathbf{V}_t)\frac{1}{T}\sum_{t=1}^{T} \|\boldsymbol{\beta}_{r_t^0}^0 - \boldsymbol{\beta}_{r_t}\|^2 \geq 0, \end{aligned} \tag{A.63}$$

so that by Lemma A.2 evaluated at $(\boldsymbol{\beta}, \mathbf{r}) = (\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{r}})$,

$$
\begin{aligned}
&Q(\widehat{\boldsymbol{\beta}}, \widehat{\mathbf{r}}) - Q(\boldsymbol{\beta}^0, \mathbf{r}^0) \\
&\geq \min_t \lambda_{\min}(N^{-1}\mathbf{V}_t'\mathbf{M}_{F^0}\mathbf{V}_t)\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{\beta}^0_{r_t^0} - \widehat{\boldsymbol{\beta}}_{\widehat{r}_t}\|^2 + O_p(T^{-1/2}) \geq 0.
\end{aligned}
\tag{A.64}
$$

This implies that

$$
\frac{1}{T}\sum_{t=1}^{T}\|\boldsymbol{\beta}^0_{r_t^0} - \widehat{\boldsymbol{\beta}}_{\widehat{r}_t}\|^2 \leq O_p(T^{-1/2}) \tag{A.65}
$$

for otherwise (A.55) cannot be satisfied asymptotically.

Proof of Theorem 1.

The proof is analogous to that of Lemma B.3 in Bonhomme and Manresa (2015) (see also Proof of Lemma A.2 of Miao et al., 2020). We need to show that the following two results hold:

$$
\max_{r\in\mathcal{R}} \min_{r'\in\mathcal{R}} \|\boldsymbol{\beta}^0_r - \widehat{\boldsymbol{\beta}}_{r'}\| = o_p(1), \tag{A.66}
$$

$$
\max_{r'\in\mathcal{R}} \min_{r\in\mathcal{R}} \|\boldsymbol{\beta}^0_r - \widehat{\boldsymbol{\beta}}_{r'}\| = o_p(1). \tag{A.67}
$$

The first result is implied by Lemma A.3. We therefore proceed to establish the second. Define the function $\sigma : \mathcal{R} \mapsto \mathcal{R}$ such that

$$
\sigma(r) := \arg\min_{r'\in\mathcal{R}} \|\boldsymbol{\beta}^0_r - \widehat{\boldsymbol{\beta}}_{r'}\|. \tag{A.68}
$$

We start by showing that $\sigma(\cdot)$ one-to-one w.p.a.1. By applying first $\|a - b\| \geq \|a\| - \|b\|$ (the reverse triangle inequality) and then $\|a + b\| \leq \|a\| + \|b\|$ (the regular version of the same inequality),

$$
\begin{aligned}
\|\widehat{\boldsymbol{\beta}}_{\sigma(r)} - \widehat{\boldsymbol{\beta}}_{\sigma(r')}\| &= \|\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'}^0 - (\boldsymbol{\beta}_r^0 - \widehat{\boldsymbol{\beta}}_{\sigma(r)} + \widehat{\boldsymbol{\beta}}_{\sigma(r')} - \boldsymbol{\beta}_{r'}^0)\| \\
&\geq \|\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'}^0\| - \|\boldsymbol{\beta}_r^0 - \widehat{\boldsymbol{\beta}}_{\sigma(r)} + \widehat{\boldsymbol{\beta}}_{\sigma(r')} - \boldsymbol{\beta}_{r'}^0\| \\
&\geq \|\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'}^0\| - \|\boldsymbol{\beta}_r^0 - \widehat{\boldsymbol{\beta}}_{\sigma(r)}\| - \|\widehat{\boldsymbol{\beta}}_{\sigma(r')} - \boldsymbol{\beta}_{r'}^0\|. \qquad \text{(A.69)}
\end{aligned}
$$

For the first term on the right-hand side, we use the fact that $\|\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_{r'}^0\| > 0$ for $r \neq r'$ by Assumption 2. The second and third terms are $o_p(1)$ because by Lemma A.3 $\|\boldsymbol{\beta}_r^0 - \widehat{\boldsymbol{\beta}}_{r'}\| = o_p(1)$ for all $(r', r) \in \mathcal{R}^2$. Hence, for $r \neq r'$,

$$
\|\widehat{\boldsymbol{\beta}}_{\sigma(r)} - \widehat{\boldsymbol{\beta}}_{\sigma(r')}\| \to_p c > 0 \qquad \text{(A.70)}
$$

as $T \to \infty$ with $N < \infty$ or $N \to \infty$. This last result implies that $\sigma(r) \neq \sigma(r')$ w.p.a.1, which in turn means that $\sigma(\cdot)$ is one-to-one and hence a proper permutation function. Moreover, the inverse of $\sigma(\cdot)$ exists and is given by $\sigma^{-1}(\cdot)$. It follows that for all $r' \in \mathcal{R}$,

$$
\min_{r \in \mathcal{R}} \|\boldsymbol{\beta}_r^0 - \widehat{\boldsymbol{\beta}}_{r'}\| \leq \|\boldsymbol{\beta}_{\sigma^{-1}(r')}^0 - \widehat{\boldsymbol{\beta}}_{r'}\| \leq \min_{r'' \in \mathcal{R}} \|\boldsymbol{\beta}_{\sigma^{-1}(r')}^0 - \widehat{\boldsymbol{\beta}}_{r''}\| = o_p(1), \qquad \text{(A.71)}
$$

where the last equality is due to Lemma A.3. Because this holds for all r' it holds also for the maximum. This establishes (A.67) and therefore the proof is complete. ■

The proof of Theorem 1 shows that there is a permutation $\sigma(r)$ of the regime index r such that $\|\boldsymbol{\beta}_r^0 - \widehat{\boldsymbol{\beta}}_{\sigma(r)}\| = o_p(1)$. By relabelling the elements of $\widehat{\boldsymbol{\beta}}$, we can take $\sigma(r) = r$. In what follows, we adopt this convention (similarly to Bonhomme and Manresa, 2015). We therefore write $\|\boldsymbol{\beta}_r^0 - \widehat{\boldsymbol{\beta}}_r\| = o_p(1)$ for $\|\boldsymbol{\beta}_r^0 - \widehat{\boldsymbol{\beta}}_{\sigma(r)}\| = o_p(1)$. This justifies considering the following set:

$$\mathcal{N} := \{\boldsymbol{\beta} \in \mathcal{B}^R : \|\boldsymbol{\beta}_r^0 - \boldsymbol{\beta}_r\| < \eta \ \text{ for all } \ r \in \mathcal{R}\}. \tag{A.72}$$

We now continue onto the proof of Theorem 2, which, following the lead of Bonhomme and Manresa (2015), proceeds as follows: Define

$$Q_t(\boldsymbol{\beta}_r) := N^{-1}\|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_r)\|^2. \tag{A.73}$$

For a given $\boldsymbol{\beta}$, our estimator $\widehat{r}_t = \widehat{r}_t(\boldsymbol{\beta}_r)$ of r_t^0 is the minimizer of this function. In Lemma A.4 below we show that the probability of misclassification is negligible when $\boldsymbol{\beta} \in \mathcal{N}$. This is then used to show that $\widehat{\boldsymbol{\beta}}$ is asymptotically equivalent to the oracle estimator that treats the regimes as known (Lemma A.5) and that the probability of misclassification is negligible also for $\widehat{r}_t(\widehat{\boldsymbol{\beta}}_r)$.

Lemma A.4. *Suppose that Assumptions 1 and 2 hold. Then, as N, $T \to \infty$,*

$$\sup_{\boldsymbol{\beta} \in \mathcal{N}} \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}(\widehat{r}_t(\boldsymbol{\beta}_r) \neq r_t^0) = o_p(N^{-\delta}), \tag{A.74}$$

where $\delta > 0$.

Proof: By the definition of $\widehat{r}_t(\cdot)$, time period t is assigned to regime r if the value of the objective function is lowest for that group; that is,

$$\mathbb{1}(\widehat{r}_t(\boldsymbol{\beta}_r) = r) \leq \mathbb{1}(Q_t(\boldsymbol{\beta}_r) \leq Q_t(\boldsymbol{\beta}_{r_t^0})). \tag{A.75}$$

Hence, if we define

$$Z_t(\boldsymbol{\beta}_r) := \mathbb{1}(r_t^0 \neq r)\mathbb{1}(Q_t(\boldsymbol{\beta}_r) \leq Q_t(\boldsymbol{\beta}_{r_t^0})) \tag{A.76}$$

for the wrongly assigned time periods, we have

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}(\widehat{r}_t(\boldsymbol{\beta}_r) \neq r_t^0) &= \sum_{r=1}^{R}\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}(r_t^0 \neq r)\mathbb{1}(\widehat{r}_t(\boldsymbol{\beta}_r) = r) \\ &\leq \sum_{r=1}^{R}\frac{1}{T}\sum_{t=1}^{T}Z_t(\boldsymbol{\beta}_r). \end{aligned} \tag{A.77}$$

Consider $Z_t(\boldsymbol{\beta}_r)$. From

$$\begin{aligned}\|\mathbf{M}_{\widehat{F}}\mathbf{X}_t\boldsymbol{\beta}_r\|^2 - \|\mathbf{M}_{\widehat{F}}\mathbf{X}_t\boldsymbol{\beta}_{r_t^0}\|^2 &= \mathrm{tr}\,(\boldsymbol{\beta}_r'\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t\boldsymbol{\beta}_r) - \mathrm{tr}\,(\boldsymbol{\beta}_{r_t^0}'\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t\boldsymbol{\beta}_{r_t^0}) \\ &= (\boldsymbol{\beta}_r - \boldsymbol{\beta}_{r_t^0})'\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_r + \boldsymbol{\beta}_{r_t^0}), \qquad \text{(A.78)} \\ -\mathrm{tr}\,(\mathbf{y}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t\boldsymbol{\beta}_r) + \mathrm{tr}\,(\mathbf{y}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t\boldsymbol{\beta}_{r_t^0}) &= -\mathbf{y}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_r - \boldsymbol{\beta}_{r_t^0}), \qquad \text{(A.79)}\end{aligned}$$

we obtain

$$NQ_t(\boldsymbol{\beta}_r) = \|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_r)\|^2 = \|\mathbf{M}_{\widehat{F}}\mathbf{y}_t\|^2 + \|\mathbf{M}_{\widehat{F}}\mathbf{X}_t\boldsymbol{\beta}_r\|^2 - 2\mathrm{tr}\,(\mathbf{y}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t\boldsymbol{\beta}_r), \tag{A.80}$$

which in turn implies

$$
\begin{aligned}
&Q_t(\boldsymbol{\beta}_r) - Q_t(\boldsymbol{\beta}_{r_t^0}) \\
&= N^{-1}[\|\mathbf{M}_{\widehat{F}}\mathbf{X}_t\boldsymbol{\beta}_r\|^2 - 2\mathrm{tr}\,(\mathbf{y}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t\boldsymbol{\beta}_r) - \|\mathbf{M}_{\widehat{F}}\mathbf{X}_t\boldsymbol{\beta}_{r_t^0}\|^2 + 2\mathrm{tr}\,(\mathbf{y}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t\boldsymbol{\beta}_{r_t^0})] \\
&= N^{-1}[(\boldsymbol{\beta}_r - \boldsymbol{\beta}_{r_t^0})'\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_r + \boldsymbol{\beta}_{r_t^0}) - 2\mathbf{y}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_r - \boldsymbol{\beta}_{r_t^0})] \\
&= N^{-1}[\mathbf{X}_t(\boldsymbol{\beta}_r + \boldsymbol{\beta}_{r_t^0}) - 2\mathbf{y}_t]'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_r - \boldsymbol{\beta}_{r_t^0}) \\
&= N^{-1}[2\mathbf{y}_t - \mathbf{X}_t(\boldsymbol{\beta}_{r_t^0} + \boldsymbol{\beta}_r)]'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r_t^0} - \boldsymbol{\beta}_r) \\
&= N^{-1}[2(\mathbf{X}_t\boldsymbol{\beta}_{r_t^0}^0 + \mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t) - \mathbf{X}_t(\boldsymbol{\beta}_{r_t^0} + \boldsymbol{\beta}_r)]'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r_t^0} - \boldsymbol{\beta}_r) =: \Delta_t(\boldsymbol{\beta}_{r_t^0}, \boldsymbol{\beta}_r). \quad \text{(A.81)}
\end{aligned}
$$

which in turn implies that $Z_t(\boldsymbol{\beta}_r)$ can be written as

$$
Z_t(\boldsymbol{\beta}_r) = \mathbb{1}(r_t^0 \neq r)\mathbb{1}(\Delta_t(\boldsymbol{\beta}_{r_t^0}, \boldsymbol{\beta}_r) \leq 0). \tag{A.82}
$$

We now want to bound $\mathbb{P}(Z_t(\boldsymbol{\beta}_r) = 1)$. In so doing, it is convenient to define

$$
I_{0,t}(r, r') := (\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)'N^{-1}\mathbf{V}_t'\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0), \tag{A.83}
$$

$$
I_{2,t}(r, r') := \Delta_t(\boldsymbol{\beta}_{r'}, \boldsymbol{\beta}_r) - \Delta_t(\boldsymbol{\beta}_{r'}^0, \boldsymbol{\beta}_r^0), \tag{A.84}
$$

$$
I_{1,t}(r, r') := \Delta_t(\boldsymbol{\beta}_{r'}^0, \boldsymbol{\beta}_r^0) - I_{0,t}(r, r'). \tag{A.85}
$$

This means that $\Delta_t(\boldsymbol{\beta}_{r'}, \boldsymbol{\beta}_r)$ can be expanded in the following fashion:

$$
\begin{aligned}
\Delta_t(\boldsymbol{\beta}_{r'}, \boldsymbol{\beta}_r) &= I_{0,t}(r, r') + [\Delta_t(\boldsymbol{\beta}_{r'}^0, \boldsymbol{\beta}_r^0) - I_{0,t}(r, r')] + [\Delta_t(\boldsymbol{\beta}_{r'}, \boldsymbol{\beta}_r) - \Delta_t(\boldsymbol{\beta}_{r'}^0, \boldsymbol{\beta}_r^0)] \\
&= I_{0,t}(r, r') + I_{1,t}(r, r') + I_{2,t}(r, r'). \qquad \text{(A.86)}
\end{aligned}
$$

By using this and the same max bound as in Bonhomme and Manresa (2015, Proof of

Lemma B.4), we obtain

$$
\begin{aligned}
Z_t(\boldsymbol{\beta}_r) &\leq \max_{r' \in \mathcal{R} \setminus \{r\}} \mathbb{1}(\Delta_t(\boldsymbol{\beta}_{r'}, \boldsymbol{\beta}_r) \leq 0) = \max_{r' \in \mathcal{R} \setminus \{r\}} \mathbb{1}(I_{1,t}(r, r') \leq -I_{0,t}(r, r') - I_{2,t}(r, r')) \\
&\leq \max_{r' \in \mathcal{R} \setminus \{r\}} \mathbb{1}(I_{1,t}(r, r') \leq -I_{0,t}(r, r') + |I_{2,t}(r, r')|), \qquad \text{(A.87)}
\end{aligned}
$$

so that if we bound the maximum over $r' \neq r$ by the corresponding sum,

$$
\mathbb{P}(Z_t(\boldsymbol{\beta}_r) = 1) \leq \sum_{r' \neq r} \mathbb{P}(I_{1,t}(r, r') \leq -I_{0,t}(r, r') + |I_{2,t}(r, r')|). \qquad \text{(A.88)}
$$

Consider $I_{2,t}(r, r')$. From the definition of $a_t(\boldsymbol{\beta}_{r'}, \boldsymbol{\beta}_r)$,

$$
\begin{aligned}
I_{2,t}(r, r') &= \Delta_t(\boldsymbol{\beta}_{r'}, \boldsymbol{\beta}_r) - \Delta_t(\boldsymbol{\beta}^0_{r'}, \boldsymbol{\beta}^0_r) \\
&= N^{-1}[2(\mathbf{X}_t\boldsymbol{\beta}^0_{r'} + \mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t) - \mathbf{X}_t(\boldsymbol{\beta}_{r'} + \boldsymbol{\beta}_r)]'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'} - \boldsymbol{\beta}_r) \\
&- N^{-1}[2(\mathbf{X}_t\boldsymbol{\beta}^0_{r'} + \mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t) - \mathbf{X}_t(\boldsymbol{\beta}^0_{r'} + \boldsymbol{\beta}^0_r)]'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}^0_{r'} - \boldsymbol{\beta}^0_r) \\
&= 2N^{-1}[(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'} - \boldsymbol{\beta}_r) - (\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}^0_{r'} - \boldsymbol{\beta}^0_r)] \\
&+ 2N^{-1}[\boldsymbol{\beta}^{0\prime}_{r'}\mathbf{X}'_t\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'} - \boldsymbol{\beta}_r) - \boldsymbol{\beta}^{0\prime}_{r'}\mathbf{X}'_t\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}^0_{r'} - \boldsymbol{\beta}^0_r)] \\
&+ N^{-1}[-(\boldsymbol{\beta}_{r'} + \boldsymbol{\beta}_r)'\mathbf{X}'_t\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'} - \boldsymbol{\beta}_r) \\
&+ (\boldsymbol{\beta}^0_{r'} + \boldsymbol{\beta}^0_r)'\mathbf{X}'_t\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}^0_{r'} - \boldsymbol{\beta}^0_r)], \qquad \text{(A.89)}
\end{aligned}
$$

where

$$
\begin{aligned}
&-(\boldsymbol{\beta}_{r'}+\boldsymbol{\beta}_r)'\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_r)+(\boldsymbol{\beta}_{r'}^0+\boldsymbol{\beta}_r^0)'\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}^0-\boldsymbol{\beta}_r^0)\\
&=-(\boldsymbol{\beta}_{r'}+\boldsymbol{\beta}_r)'\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_r)+(\boldsymbol{\beta}_{r'}^0+\boldsymbol{\beta}_r^0)'\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_r)\\
&-(\boldsymbol{\beta}_{r'}^0+\boldsymbol{\beta}_r^0)'\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_r)+(\boldsymbol{\beta}_{r'}^0+\boldsymbol{\beta}_r^0)'\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}^0-\boldsymbol{\beta}_r^0)\\
&=-[(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^0)+(\boldsymbol{\beta}_r-\boldsymbol{\beta}_r^0)]'\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_r)\\
&-(\boldsymbol{\beta}_{r'}^0+\boldsymbol{\beta}_r^0)'\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t[(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^0)-(\boldsymbol{\beta}_r-\boldsymbol{\beta}_r^0)] \qquad \text{(A.90)}
\end{aligned}
$$

implying

$$
\begin{aligned}
I_{2,t}(r,r') &= 2N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t+\boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}\mathbf{X}_t[(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^0)-(\boldsymbol{\beta}_r-\boldsymbol{\beta}_r^0)]\\
&+2N^{-1}\boldsymbol{\beta}_{r'}^{0\prime}\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t[(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^0)-(\boldsymbol{\beta}_r-\boldsymbol{\beta}_r^0)]\\
&-[(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^0)+(\boldsymbol{\beta}_r-\boldsymbol{\beta}_r^0)]'N^{-1}\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_r)\\
&-(\boldsymbol{\beta}_{r'}^0+\boldsymbol{\beta}_r^0)'N^{-1}\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t[(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^0)-(\boldsymbol{\beta}_r-\boldsymbol{\beta}_r^0)]. \qquad \text{(A.91)}
\end{aligned}
$$

We now consider each of the terms on the right-hand side of the above equation, starting with the first. By using the fact that $\|\boldsymbol{\beta}_r-\boldsymbol{\beta}_r^0\|<\eta$ for $\boldsymbol{\beta}\in\mathcal{N}$ and the same steps as in the proof of Lemma A.2, we can show that

$$
\begin{aligned}
&|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t+\boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}\mathbf{X}_t[(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^0)-(\boldsymbol{\beta}_r-\boldsymbol{\beta}_r^0)]|\\
&\leq\|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t+\boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}\mathbf{X}_t\|(\|\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^0\|+\|\boldsymbol{\beta}_r-\boldsymbol{\beta}_r^0\|)\\
&\leq 2\eta\|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t+\boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}\mathbf{X}_t\|\\
&=2\eta\|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t+\boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}(\mathbf{V}_t-\overline{\mathbf{U}}\overline{\mathbf{B}}_m\boldsymbol{\Gamma}_t)\|\\
&\leq 2\eta[\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}\mathbf{V}_t\|+\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}\overline{\mathbf{U}}\|\|\overline{\mathbf{B}}_m\|\|\boldsymbol{\Gamma}_t\|+\|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t+\boldsymbol{\varepsilon}_t)'(\mathbf{M}_{\widehat{F}}-\mathbf{M}_{F^0})\mathbf{V}_t\|\\
&+\|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t+\boldsymbol{\varepsilon}_t)'(\mathbf{M}_{\widehat{F}}-\mathbf{M}_{F^0})\overline{\mathbf{U}}\|\|\overline{\mathbf{B}}_m\|\|\boldsymbol{\Gamma}_t\|]. \qquad \text{(A.92)}
\end{aligned}
$$

As for the first term on the right-hand side, we have

$$\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}\mathbf{V}_t\| \le \|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t\| + \|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{F}^0\|\|(N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\|\|N^{-1}\mathbf{F}^{0\prime}\mathbf{V}_t\|. \quad \text{(A.93)}$$

From the proof of Lemma A.2, we know that $\|N^{-1}\mathbf{F}^{0\prime}\mathbf{V}_t\|$ and $\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{F}^0\|$ are $O_p(1)$ under the conditions of that lemma. This is not enough here, though. In particular, the condition that $\mathbf{v}_{i,t}$ is at most weakly serially correlated must be strengthened to independence (Assumption 2), so that

$$\left\|\frac{1}{N\sqrt{T}}\sum_{s\neq t}\sum_{i=1}^{N}\mathbf{v}_{i,t}\mathbf{v}_{i,s}'\right\| = O_p(N^{-1/2}), \quad \text{(A.94)}$$

in which case (A.27) reads

$$\left\|\frac{1}{N\sqrt{T}}\sum_{s=1}^{T}\mathbf{V}_t'\mathbf{V}_s\right\| = O_p(\mathcal{C}_{NT}^{-1}). \quad \text{(A.95)}$$

where $\mathcal{C}_{NT} := \max\{\sqrt{N},\sqrt{T}\}$. This implies that $\|N^{-1}\sqrt{T}\mathbf{V}_t'\overline{\mathbf{U}}\|$ and hence also $\|N^{-1}\mathbf{V}_t'\mathbf{F}^0\|$ are of the same order. By using this, $\|(N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\| = O_p(1)$ and $\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t\| = O_p(N^{-1/2})$, we can show that

$$\begin{aligned}\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}\mathbf{V}_t\| &\le \|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t\| + \|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{F}^0\|\|(N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\|\|N^{-1}\mathbf{F}^{0\prime}\mathbf{V}_t\| \\ &= \|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t\| + O_p(\mathcal{C}_{NT}^{-2}) \quad \text{(A.96)}\end{aligned}$$

The second term on the right-hand side of (A.92) involves $\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}\overline{\mathbf{U}}\|$. This term has exactly the same structure as $\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}\mathbf{V}_t\|$ but with $\overline{\mathbf{U}}$ in place of $\mathbf{V}_t$. It is therefore of smaller order in magnitude than $O_p(\mathcal{C}_{NT}^{-1})$. From Proof of Lemma A.2, we know that $\|N^{-1}\overline{\mathbf{U}}'\mathbf{F}\|$ and $\|N^{-1}\sqrt{T}\overline{\mathbf{U}}'\overline{\mathbf{U}}\|$ are $O_p((NT)^{-1/2})$ and $O_p(T^{-1/2})$, respectively. Hence, $\|N^{-1}\overline{\mathbf{U}}'\mathbf{F}^0\| = \|[N^{-1}\overline{\mathbf{U}}'\mathbf{F}, N^{-1}\sqrt{T}\overline{\mathbf{U}}'\overline{\mathbf{U}}\mathbf{B}_{-m}]\| = O_p(T^{-1/2})$. Also, $\|N^{-1}\boldsymbol{\varepsilon}_t'\overline{\mathbf{U}}\|$ is of the same order as $\|N^{-1}\mathbf{V}_t'\overline{\mathbf{U}}\|$ (Proof of Lemma A.2). We can therefore show that

$$\begin{aligned}\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{M}_{F^0}\overline{\mathbf{U}}\| &\leq \|N^{-1}\boldsymbol{\varepsilon}_t'\overline{\mathbf{U}}\| + \|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{F}^0\|\|(N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\|\|N^{-1}\mathbf{F}^{0\prime}\overline{\mathbf{U}}\| \\ &= O_p(T^{-1}) + O_p((NT)^{-1/2}) \leq O_p(\mathcal{C}_{NT}^{-2}). \end{aligned} \tag{A.97}$$

We now move on to the third term, $\|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'(\mathbf{M}_{\widehat{F}} - \mathbf{M}_{F^0})\mathbf{V}_t\|$. Here we use $\|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'\mathbf{F}^0\| = O_p(1)$ and

$$\|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\| \leq \|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'\overline{\mathbf{U}}\|\|\overline{\mathbf{B}}_m\| = O_p(\mathcal{C}_{NT}^{-2}). \tag{A.98}$$

Together with (A.21) and some of the orders derived as a part of the proof of Lemma A.2, these results imply

$$\begin{aligned}&\|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'(\mathbf{M}_{F^0} - \mathbf{M}_{\widehat{F}})\mathbf{V}_t\| \\ &\leq \|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\|\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|\|N^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'\mathbf{V}_t\| \\ &+ \|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)\|\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|\|N^{-1}\mathbf{F}^{0\prime}\mathbf{V}_t\| \\ &+ \|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'\mathbf{F}^0\|\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1}\|\|N^{-1}(\widehat{\mathbf{F}}^0 - \mathbf{F}^0)'\mathbf{V}_t\| \\ &+ \|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'\mathbf{F}^0\|\|(N^{-1}\widehat{\mathbf{F}}^{0\prime}\widehat{\mathbf{F}}^0)^{-1} - (N^{-1}\mathbf{F}^{0\prime}\mathbf{F}^0)^{-1}\|\|N^{-1}\mathbf{F}^{0\prime}\mathbf{V}_t\| \\ &= O_p(\mathcal{C}_{NT}^{-2}). \end{aligned} \tag{A.99}$$

Just as with the first and second terms on the right-hand side of (A.92), the fourth has the same structure as the third but with $\overline{\mathbf{U}}$ in place of $\mathbf{V}_t$. Adjusting the above arguments to this change leads to the following:

$$\|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'(\mathbf{M}_{F^0} - \mathbf{M}_{\widehat{F}})\overline{\mathbf{U}}\| = O_p(\mathcal{C}_{NT}^{-2}). \tag{A.100}$$

The above results imply that (A.92) reduces to

$$|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t+\boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}\mathbf{X}_t[(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^0)-(\boldsymbol{\beta}_r-\boldsymbol{\beta}_r^0)]|=2\eta(\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t\|+O_p(\mathcal{C}_{NT}^{-2})). \tag{A.101}$$

This completes our evaluation of the first term of $I_{2,t}(r,r')$. We now move on to the second. Because $\|\boldsymbol{\beta}_r-\boldsymbol{\beta}_r^0\|<\eta$, we have $\|\boldsymbol{\beta}_r\|\leq\|\boldsymbol{\beta}_r-\boldsymbol{\beta}_r^0\|+\|\boldsymbol{\beta}_r^0\|\leq 2\eta$, which holds for all $\boldsymbol{\beta}_r$, including $\boldsymbol{\beta}_r^0$. Hence,

$$\begin{aligned}
&|N^{-1}\boldsymbol{\beta}_{r'}^{0\prime}\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t[(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^0)-(\boldsymbol{\beta}_r-\boldsymbol{\beta}_r^0)]|\\
&\leq\|\boldsymbol{\beta}_{r'}^0\|\|N^{-1}\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t\|(\|\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^0\|+\|\boldsymbol{\beta}_r-\boldsymbol{\beta}_r^0\|)\\
&\leq 4\eta^2\|N^{-1}\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t\|\\
&=4\eta^2(\|N^{-1}\mathbf{V}_t'\mathbf{V}_t\|+O_p(\mathcal{C}_{NT}^{-2})).
\end{aligned} \tag{A.102}$$

For the last equality, we employ the same steps used earlier to evaluate $\|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t+\boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}\mathbf{X}_t\|$ to show that

$$\begin{aligned}
N^{-1}\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t&=N^{-1}(\mathbf{V}_t-\overline{\mathbf{U}}\overline{\mathbf{B}}_m\boldsymbol{\Gamma}_t)'\mathbf{M}_{\widehat{F}}(\mathbf{V}_t-\overline{\mathbf{U}}\overline{\mathbf{B}}_m\boldsymbol{\Gamma}_t)\\
&=N^{-1}(\mathbf{V}_t-\overline{\mathbf{U}}\overline{\mathbf{B}}_m\boldsymbol{\Gamma}_t)'\mathbf{M}_{F^0}(\mathbf{V}_t-\overline{\mathbf{U}}\overline{\mathbf{B}}_m\boldsymbol{\Gamma}_t)\\
&+N^{-1}(\mathbf{V}_t-\overline{\mathbf{U}}\overline{\mathbf{B}}_m\boldsymbol{\Gamma}_t)'(\mathbf{M}_{\widehat{F}}-\mathbf{M}_{F^0})(\mathbf{V}_t-\overline{\mathbf{U}}\overline{\mathbf{B}}_m\boldsymbol{\Gamma}_t)\\
&=N^{-1}\mathbf{V}_t'\mathbf{V}_t+O_p(\mathcal{C}_{NT}^{-2}),
\end{aligned} \tag{A.103}$$

where, as usual, $\mathbf{A}=\mathbf{B}+o_p(1)$ means $\|\mathbf{A}-\mathbf{B}\|=o_p(1)$.

The above arguments can be applied also to the third term of $I_{2,t}(r,r')$, from which we obtain

$$
\begin{aligned}
&|[(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^{0})+(\boldsymbol{\beta}_{r}-\boldsymbol{\beta}_{r}^{0})]'N^{-1}\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r})| \\
&\leq 8\eta^2(\|N^{-1}\mathbf{V}_t'\mathbf{V}_t\|+O_p(\mathcal{C}_{NT}^{-2})). \qquad \text{(A.104)}
\end{aligned}
$$

The fourth term has the same bound.

The reminders of the last three terms of $I_{2,t}(r,r')$ are dominated by the first term. It follows that if η is small such that $\eta>\eta^2$, then

$$
\begin{aligned}
|I_{2,t}(r,r')| &\leq 2|N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t+\boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}\mathbf{X}_t[(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^{0})-(\boldsymbol{\beta}_{r}-\boldsymbol{\beta}_{r}^{0})]| \\
&\quad +2|N^{-1}\boldsymbol{\beta}_{r'}^{0\prime}\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t[(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^{0})-(\boldsymbol{\beta}_{r}-\boldsymbol{\beta}_{r}^{0})]| \\
&\quad +|[(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^{0})+(\boldsymbol{\beta}_{r}-\boldsymbol{\beta}_{r}^{0})]'N^{-1}\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r})| \\
&\quad +|(\boldsymbol{\beta}_{r'}^{0}+\boldsymbol{\beta}_{r}^{0})'N^{-1}\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t[(\boldsymbol{\beta}_{r'}-\boldsymbol{\beta}_{r'}^{0})-(\boldsymbol{\beta}_{r}-\boldsymbol{\beta}_{r}^{0})]| \\
&\leq 4\eta(\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t\|+O_p(\mathcal{C}_{NT}^{-2}))+24\eta^2(\|N^{-1}\mathbf{V}_t'\mathbf{V}_t\|+O_p(\mathcal{C}_{NT}^{-2})) \\
&\leq \eta(4\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t\|+24\|N^{-1}\mathbf{V}_t'\mathbf{V}_t\|)+O_p(\eta\mathcal{C}_{NT}^{-2}). \qquad \text{(A.105)}
\end{aligned}
$$

We now move on to $I_{1,t}(r,r')$. From the definitions of $\Delta_t(\boldsymbol{\beta}_{r'}^0, \boldsymbol{\beta}_r^0)$ and $I_{0,t}(r,r')$,

$$\begin{aligned}
I_{1,t}(r,r') &= \Delta_t(\boldsymbol{\beta}_{r'}^0, \boldsymbol{\beta}_r^0) - I_{0,t}(r,r') \\
&= N^{-1}[2(\mathbf{X}_t\boldsymbol{\beta}_{r'}^0 + \mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t) - \mathbf{X}_t(\boldsymbol{\beta}_{r'}^0 + \boldsymbol{\beta}_r^0)]'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0) \\
&- (\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)'N^{-1}\mathbf{V}_t'\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0) \\
&= 2N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0) + 2\boldsymbol{\beta}_{r'}^{0\prime}N^{-1}\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0) \\
&- (\boldsymbol{\beta}_{r'}^0 + \boldsymbol{\beta}_r^0)'N^{-1}\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0) - (\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)'N^{-1}\mathbf{V}_t'\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0) \\
&= 2N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0) + (\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)'N^{-1}\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0) \\
&- (\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)'N^{-1}\mathbf{V}_t'\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0) \\
&= 2N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0) \\
&+ (\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)'N^{-1}(\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t - \mathbf{V}_t'\mathbf{V}_t)(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0). \qquad \text{(A.106)}
\end{aligned}$$

We have already shown that $\|N^{-1}(\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t - \mathbf{V}_t'\mathbf{V}_t)\| = O_p(\mathcal{C}_{NT}^{-2})$, which in view of $\|\boldsymbol{\beta}_r^0\| \leq 2\eta$ gives

$$\begin{aligned}
&|(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)'N^{-1}(\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t - \mathbf{V}_t\mathbf{V}_t)(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)| \\
&\leq 16\eta^2\|N^{-1}(\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t - \mathbf{V}_t'\mathbf{V}_t)\| \leq O_p(\eta\mathcal{C}_{NT}^{-2}). \qquad \text{(A.107)}
\end{aligned}$$

Hence, since $N^{-1}(\mathbf{F}\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t)'\mathbf{M}_{\widehat{F}}\mathbf{X}_t = N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t + O_p(\mathcal{C}_{NT}^{-2})$,

$$I_{1,t}(r,r') = 2N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0) + O_p(\eta\mathcal{C}_{NT}^{-2}). \qquad \text{(A.108)}$$

For $I_{0,t}(r,r')$, we write

$$I_{0,t}(r,r') = \mathrm{tr}\,[(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)'N^{-1}\mathbf{V}_t'\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)] = \|N^{-1/2}\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)\|^2. \qquad \text{(A.109)}$$

We now make use of the above results to evaluate $\mathbb{P}(I_{1,t}(r,r') \leq -I_{0,t}(r,r') + |I_{2,t}(r,r')|)$. In so doing, it is convenient to also make use of the fact that $\mathbb{P}(Y + o_p(1) \leq a) = \mathbb{P}(Y \leq a - o_p(1)) \leq \mathbb{P}(Y \leq a + |o_p(1)|)$, where Y is a random variable and a is a constant. This implies

$$
\begin{aligned}
&\mathbb{P}(I_{1,t}(r,r') \leq -I_{0,t}(r,r') + |I_{2,t}(r,r')|) \\
&= \mathbb{P}(2N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0) + O_p(\eta\mathcal{C}_{NT}^{-2}) \leq -\|N^{-1/2}\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)\|^2 \\
&+ 4\eta\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t\| + 24\eta^2\|N^{-1}\mathbf{V}_t'\mathbf{V}_t\| + |O_p(\eta\mathcal{C}_{NT}^{-2})|) \\
&\leq \mathbb{P}(2N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0) \leq -\|N^{-1/2}\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)\|^2 \\
&+ \eta(4\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t\| + 24\|N^{-1}\mathbf{V}_t'\mathbf{V}_t\|) + 2|O_p(\eta\mathcal{C}_{NT}^{-2})|) \\
&\leq \mathbb{P}(2N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0) \leq -\|N^{-1/2}\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)\|^2 \\
&+ \eta(4\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t\| + 24\|N^{-1}\mathbf{V}_t'\mathbf{V}_t\|) + 2|O_p(\eta\mathcal{C}_{NT}^{-2})|). \qquad \text{(A.110)}
\end{aligned}
$$

From Loyo and Boot (2024, Proof of Lemma C.3): If Y and X are random variables and a and b constants, then

$$
\begin{aligned}
\mathbb{P}(X < Y + a) &= \mathbb{P}(X < Y + a, Y > b) + \mathbb{P}(X < Y + a, Y < b) \\
&\leq \mathbb{P}(X < a + b) + \mathbb{P}(Y > b). \qquad \text{(A.111)}
\end{aligned}
$$

This result can be used repeatedly to show the following:

$$
\begin{aligned}
&\mathbb{P}(I_{1,t}(r,r') \leq -I_{0,t}(r,r') + |I_{2,t}(r,r')|) \\
&\leq \mathbb{P}\left(2N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0) \leq -\frac{c_t(r,r')}{2} + \eta(4M + 24M + 2M)\right) \\
&+ \mathbb{P}\left(\|N^{-1/2}\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)\|^2 \leq \frac{c_t(r,r')}{2}\right) + \mathbb{P}(\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t\| \geq M) \\
&+ \mathbb{P}(\|N^{-1}\mathbf{V}_t'\mathbf{V}_t\| \geq M) + \mathbb{P}(|O(\mathcal{C}_{NT}^{-2})| \geq M) \qquad \text{(A.112)}
\end{aligned}
$$

for some $M > 0$ and where $c_t(r,r') := (\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)'\boldsymbol{\Sigma}_{v,t}(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)$. It is important to note that from $\mathrm{tr}\,(\mathbf{AB}) \geq \lambda_{\min}(\mathbf{A})\mathrm{tr}\,\mathbf{B}$ and Assumption 2, we have

$$\begin{aligned} c_t(r,r') &= \mathrm{tr}\,[(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)'\boldsymbol{\Sigma}_{v,t}(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)] = \mathrm{tr}\,[\boldsymbol{\Sigma}_{v,t}(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)'] \\ &\geq \lambda_{\min}(\boldsymbol{\Sigma}_{v,t})\mathrm{tr}\,[(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)'] = \lambda_{\min}(\boldsymbol{\Sigma}_{v,t})\|\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0\|^2 > 0 \quad \text{(A.113)} \end{aligned}$$

whenever $r' \neq r$. We are now going to bound each of the terms appearing on the right-hand side of (A.112). In so doing, we will employ Lemma B.5 of Bonhomme and Manresa (2015), which states that if z_t is a strongly mixing process with zero mean and tail probability $\mathbb{P}(|z_t| > M) \leq \exp(1 - (M/\nu)^\kappa)$, where M, ν, $\kappa > 0$. Then, for all $\delta > 0$,

$$\mathbb{P}\left(\left|\frac{1}{T}\sum_{t=1}^{T} z_t\right| \geq M\right) = o(T^{-\delta}). \tag{A.114}$$

We will apply this lemma to the cross-sectional dimension. This is possible because under Assumption 2, $\varepsilon_{i,t}$ and $\mathbf{v}_{i,t}$ are mean zero, independent and hence strongly mixing, and with the required tails. We begin by applying Bonhomme and Manresa's lemma to the third term on the right-hand side of (A.112), which is possible since $\varepsilon_{i,t}\mathbf{v}_{i,t}$ is mean zero and independent over i. It follows that

$$\mathbb{P}(\|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t\| \geq M) = o(N^{-\delta}), \tag{A.115}$$

For the fourth term, $P(\|N^{-1}\mathbf{V}_t'\mathbf{V}_t\| \geq M)$, we use

$$\|N^{-1}\mathbf{V}_t'\mathbf{V}_t\| \leq \|N^{-1}\mathbf{V}_t'\mathbf{V}_t - \mathbb{E}(N^{-1}\mathbf{V}_t'\mathbf{V}_t)\| + \|\mathbb{E}(N^{-1}\mathbf{V}_t'\mathbf{V}_t)\|, \tag{A.116}$$

where $\|\mathbb{E}(N^{-1}\mathbf{V}_t'\mathbf{V}_t)\| \to \|\boldsymbol{\Sigma}_{v,t}\| \leq c$ as $N \to \infty$ for some $c < \infty$ by Assumption 2. By using this last result to handle $\|\mathbb{E}(N^{-1}\mathbf{V}_t'\mathbf{V}_t)\|$, and applying Bonhomme and Manresa's

Lemma B.5 to $\|N^{-1}\mathbf{V}_t'\mathbf{V}_t - \mathbb{E}(N^{-1}\mathbf{V}_t'\mathbf{V}_t)\|$, we obtain the following large-N bound for $P(\|N^{-1}\mathbf{V}_t'\mathbf{V}_t\| \geq M)$:

$$\begin{aligned}\mathbb{P}(\|N^{-1}\mathbf{V}_t'\mathbf{V}_t\| \geq M) &\leq \mathbb{P}(\|N^{-1}\mathbf{V}_t'\mathbf{V}_t - E(N^{-1}\mathbf{V}_t'\mathbf{V}_t)\| + \|\mathbb{E}(N^{-1}\mathbf{V}_t'\mathbf{V}_t)\| \geq M) \\ &\leq \mathbb{P}(\|N^{-1}\mathbf{V}_t'\mathbf{V}_t - \mathbb{E}(N^{-1}\mathbf{V}_t'\mathbf{V}_t)\| + c \geq M) = o(N^{-\delta}).\end{aligned} \tag{A.117}$$

Let us now consider $\mathbb{P}(\|N^{-1/2}\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)\|^2 \leq c_t(r,r')/2)$, the second term on the right-hand side of (A.112). Because $\mathbb{E}(\|N^{-1/2}\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)\|^2) = c_t(r,r')$ for N large enough, we have

$$\mathbb{E}(\|N^{-1/2}\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)\|^2) \geq \frac{2c_t(r,r')}{3}. \tag{A.118}$$

Hence, since $\mathbb{P}(Y \leq -a) = \mathbb{P}(-Y \geq a) \leq \mathbb{P}(|Y| \geq a)$,

$$\begin{aligned}&\mathbb{P}\left(\|N^{-1/2}\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)\|^2 \leq \frac{c_t(r,r')}{2}\right) \\ &\leq \mathbb{P}\left(\|N^{-1/2}\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)\|^2 - \mathbb{E}(\|N^{-1/2}\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)\|^2) \leq -\frac{c_t(r,r')}{6}\right) \\ &\leq \mathbb{P}\left(|\|N^{-1/2}\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)\|^2 - \mathbb{E}(\|N^{-1/2}\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)\|^2)| \geq \frac{c_t(r,r')}{6}\right).\end{aligned} \tag{A.119}$$

Applying Lemma B.5 of Bonhomme and Manresa to this yields

$$\mathbb{P}\left(\|N^{-1/2}\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0 - \boldsymbol{\beta}_r^0)\|^2 \leq \frac{c_t(r,r')}{2}\right) = o(N^{-\delta}). \tag{A.120}$$

We move on to the first term on the right-hand side of (A.112). If we assume that $\eta \leq c_t(r,r')/(120M)$ and then apply $\mathbb{P}(Y \leq -a) \leq \mathbb{P}(|Y| \geq a)$, this term can be written as

$$
\begin{aligned}
&\mathbb{P}\left(2N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0-\boldsymbol{\beta}_r^0)\le -\frac{c_t(r,r')}{2}+30\eta M\right)\\
&\le \mathbb{P}\left(2N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0-\boldsymbol{\beta}_r^0)\le -\frac{c_t(r,r')}{4}\right)\\
&\le \mathbb{P}\left(|N^{-1}\boldsymbol{\varepsilon}_t'\mathbf{V}_t(\boldsymbol{\beta}_{r'}^0-\boldsymbol{\beta}_r^0)|\ge \frac{c_t(r,r')}{4}\right),
\end{aligned}
\tag{A.121}
$$

which is $o(N^{-\delta})$ by Bonhomme and Manresa's lemma. The fifth and final term on the right-hand side of (A.112) is of even lower order than this and is therefore dominated by the other terms. Hence, by putting everything together, (A.112) becomes

$$
\mathbb{P}(I_{1,t}(r,r')\le -I_{0,t}(r,r')+|I_{2,t}(r,r')|)=o(N^{-\delta}),
\tag{A.122}
$$

which in view of the fact that $G<\infty$ in turn implies

$$
\mathbb{P}(Z_t(\boldsymbol{\beta}_r)=1)\le \sum_{r'\neq g}\mathbb{P}(I_{1,t}(r,r')\le -I_{0,t}(r,r')+|I_{2,t}(r,r')|)=o(N^{-\delta}).
\tag{A.123}
$$

Hence, since $\mathbb{P}(A)=\mathbb{E}[1(A)]$, we arrive at

$$
\begin{aligned}
\mathbb{E}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}(\widehat{r}_t(\boldsymbol{\beta}_r)\neq r_t^0)\right)&\le \sum_{r=1}^{R}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(Z_t(\boldsymbol{\beta}_r))\\
&=\sum_{r=1}^{R}\frac{1}{T}\sum_{t=1}^{T}\mathbb{P}(Z_t(\boldsymbol{\beta}_r)=1)=o(N^{-\delta}).
\end{aligned}
\tag{A.124}
$$

This result holds uniformly in $\boldsymbol{\beta}\in\mathcal{N}$. The sought result is implied by this and Markov's inequality (see Bonhomme and Manresa, Proof of Lemma B.4). ■

Define

$$Q(\boldsymbol{\beta}) := Q(\boldsymbol{\beta}, \widehat{\mathbf{r}}) = \frac{1}{NT}\sum_{t=1}^{T} \|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{\widehat{r}_t})\|^2, \tag{A.125}$$

$$Q^{\inf}(\boldsymbol{\beta}) := Q(\boldsymbol{\beta}, \mathbf{r}^0) = \frac{1}{NT}\sum_{t=1}^{T} \|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{r_t^0})\|^2, \tag{A.126}$$

where $\widehat{r}_t = \widehat{r}_t(\boldsymbol{\beta}_r)$. Let $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}^{\inf}$ be the minimizers of $Q(\boldsymbol{\beta})$ and $Q^{\inf}(\boldsymbol{\beta})$, respectively.

Lemma A.5. *Suppose that Assumptions 1 and 2 hold. Then, as N, $T \to \infty$,*

$$\max_{r\in\mathcal{R}} \|\widehat{\boldsymbol{\beta}}_r^{\inf} - \widehat{\boldsymbol{\beta}}_r\|^2 = o_p(N^{-\delta}). \tag{A.127}$$

Proof: Because $Q(\boldsymbol{\beta}) - Q^{\inf}(\boldsymbol{\beta}) = 0$ if $\widehat{r}_t = r_t^0$, in the notation of the proof of Lemma A.4, we have

$$\begin{aligned}
Q(\boldsymbol{\beta}) - Q^{\inf}(\boldsymbol{\beta}) &= \frac{1}{NT}\sum_{t=1}^{T}[\|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{\widehat{r}_t})\|^2 - \|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{r_t^0})\|^2] \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}(\widehat{r}_t(\boldsymbol{\beta}_r) \neq r_t^0)N^{-1}[\|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{\widehat{r}_t})\|^2 - \|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta}_{r_t^0})\|^2] \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}(\widehat{r}_t(\boldsymbol{\beta}_r) \neq r_t^0)\Delta_t(\boldsymbol{\beta}_{r_t^0}, \boldsymbol{\beta}_{\widehat{r}_t}).
\end{aligned} \tag{A.128}$$

By Assumption 2, there is a constant $c < \infty$ such that $\|\mathbb{E}(N^{-1}\mathbf{V}_t'\mathbf{V}_t)\| \leq c$ for all t and N sufficiently large. By using this and some of the results reported in the proof of Lemma A.4,

$$|\Delta_t(\boldsymbol{\beta}_{r_t^0}, \boldsymbol{\beta}_{\widehat{r}_t})| \leq 40\eta^2\|N^{-1}\mathbf{V}_t'\mathbf{V}_t\| + o_p(1) \leq 40\eta^2 c + o_p(1). \tag{A.129}$$

Hence, by Lemma A.4,

$$|Q(\boldsymbol{\beta}) - Q^{\inf}(\boldsymbol{\beta})| \leq \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}(\widehat{r}_t(\boldsymbol{\beta}_r) \neq r_t^0)(40\eta^2 c + o_p(1)) = o_p(N^{-\delta}), \tag{A.130}$$

which holds uniformly in $\boldsymbol{\beta} \in \mathcal{N}$. We now make use of this last result to bound $|Q(\widehat{\boldsymbol{\beta}}) - Q^{\inf}(\widehat{\boldsymbol{\beta}})|$. For any $\epsilon > 0$,

$$\begin{aligned}
&\mathbb{P}(|Q(\widehat{\boldsymbol{\beta}}) - Q^{\inf}(\widehat{\boldsymbol{\beta}})| > \epsilon N^{\delta}) \\
&\leq \mathbb{P}(\widehat{\boldsymbol{\beta}} \notin \mathcal{N}) + \mathbb{P}(\widehat{\boldsymbol{\beta}} \in \mathcal{N}, |Q(\widehat{\boldsymbol{\beta}}) - Q^{\inf}(\widehat{\boldsymbol{\beta}})| > \epsilon N^{\delta}),
\end{aligned} \tag{A.131}$$

where, in view of $\mathbb{P}(A) = \mathbb{E}[\mathbb{1}(A)]$, $\mathbb{1}(A \cap B) = \mathbb{1}(A)\mathbb{1}(B)$ and $|\widehat{Q}(\boldsymbol{\beta}) - Q^{\inf}(\boldsymbol{\beta})| = o_p(N^{-\delta})$,

$$\begin{aligned}
&\mathbb{P}(\widehat{\boldsymbol{\beta}} \in \mathcal{N}, |Q(\widehat{\boldsymbol{\beta}}) - Q^{\inf}(\widehat{\boldsymbol{\beta}})| > \epsilon N^{\delta}) \\
&= \mathbb{E}[\mathbb{1}(\widehat{\boldsymbol{\beta}} \in \mathcal{N})\mathbb{1}(|Q(\widehat{\boldsymbol{\beta}}) - Q^{\inf}(\widehat{\boldsymbol{\beta}})| > \epsilon N^{\delta})] \\
&\leq \mathbb{E}\left[\mathbb{1}\left(\sup_{\boldsymbol{\beta} \in \mathcal{N}} |Q(\boldsymbol{\beta}) - Q^{\inf}(\boldsymbol{\beta})| > \epsilon N^{\delta}\right)\right] \\
&= \mathbb{P}\left(\sup_{\boldsymbol{\beta} \in \mathcal{N}} |Q(\boldsymbol{\beta}) - Q^{\inf}(\boldsymbol{\beta})| > \epsilon N^{\delta}\right) = o(1).
\end{aligned} \tag{A.132}$$

Since $\widehat{\boldsymbol{\beta}}$ is consistent by Theorem 1, we also have $P(\widehat{\boldsymbol{\beta}} \notin \mathcal{N}) = o(1)$. It follows that

$$\mathbb{P}(|Q(\widehat{\boldsymbol{\beta}}) - Q^{\inf}(\widehat{\boldsymbol{\beta}})| > \epsilon N^{\delta}) = o(1), \tag{A.133}$$

so that

$$|Q(\widehat{\boldsymbol{\beta}}) - Q^{\inf}(\widehat{\boldsymbol{\beta}})| = o_p(N^{-\delta}). \tag{A.134}$$

We similarly have $|Q(\widehat{\boldsymbol{\beta}}^{\text{inf}}) - Q^{\text{inf}}(\widehat{\boldsymbol{\beta}}^{\text{inf}})| = o_p(N^{-\delta})$. These results can be used to show that

$$0 \le Q^{\text{inf}}(\widehat{\boldsymbol{\beta}}) - Q^{\text{inf}}(\widehat{\boldsymbol{\beta}}^{\text{inf}}) = Q(\widehat{\boldsymbol{\beta}}) - Q(\widehat{\boldsymbol{\beta}}^{\text{inf}}) + o_p(N^{-\delta}) \le o_p(N^{-\delta}), \tag{A.135}$$

where the first inequality holds because $\widehat{\boldsymbol{\beta}}^{\text{inf}}$ minimizes $Q^{\text{inf}}(\cdot)$, and therefore $Q^{\text{inf}}(\widehat{\boldsymbol{\beta}}) - Q^{\text{inf}}(\widehat{\boldsymbol{\beta}}^{\text{inf}}) \ge 0$. Similarly, since $\widehat{\boldsymbol{\beta}}$ minimizes $Q(\cdot)$, $Q(\widehat{\boldsymbol{\beta}}) - Q(\widehat{\boldsymbol{\beta}}^{\text{inf}}) \le 0$, from which the last equality follows. We can therefore show that

$$Q^{\text{inf}}(\widehat{\boldsymbol{\beta}}) - Q^{\text{inf}}(\widehat{\boldsymbol{\beta}}^{\text{inf}}) = o_p(N^{-\delta}). \tag{A.136}$$

Consider the left-hand side of the above equation. Use of $\|\mathbf{A}+\mathbf{B}\|^2 = \|\mathbf{A}\|^2 + \|\mathbf{B}\|^2 + 2\text{tr}\,(\mathbf{A}'\mathbf{B})$ gives

$$\begin{aligned}
&Q^{\text{inf}}(\widehat{\boldsymbol{\beta}}) - Q^{\text{inf}}(\widehat{\boldsymbol{\beta}}^{\text{inf}}) \\
&= \frac{1}{NT}\sum_{t=1}^{T}[\|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t\widehat{\boldsymbol{\beta}}_{r_t^0})\|^2 - \|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t\widehat{\boldsymbol{\beta}}_{r_t^0}^{\text{inf}})\|^2] \\
&= \frac{1}{NT}\sum_{t=1}^{T}[\|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t\widehat{\boldsymbol{\beta}}_{r_t^0}^{\text{inf}}) + \mathbf{M}_{\widehat{F}}\mathbf{X}_t(\widehat{\boldsymbol{\beta}}_{r_t^0}^{\text{inf}} - \widehat{\boldsymbol{\beta}}_{r_t^0})\|^2 - \|\mathbf{M}_{\widehat{F}}(\mathbf{y}_t - \mathbf{X}_t\widehat{\boldsymbol{\beta}}_{r_t^0}^{\text{inf}})\|^2] \\
&= \frac{1}{NT}\sum_{t=1}^{T}\|\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\widehat{\boldsymbol{\beta}}_{r_t^0}^{\text{inf}} - \widehat{\boldsymbol{\beta}}_{r_t^0})\|^2 + 2\text{tr}\,[(\mathbf{y}_t - \mathbf{X}_t\widehat{\boldsymbol{\beta}}_{r_t^0}^{\text{inf}})'\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\widehat{\boldsymbol{\beta}}_{r_t^0}^{\text{inf}} - \widehat{\boldsymbol{\beta}}_{r_t^0})] \\
&= \frac{1}{NT}\sum_{t=1}^{T}\|\mathbf{M}_{\widehat{F}}\mathbf{X}_t(\widehat{\boldsymbol{\beta}}_{r_t^0}^{\text{inf}} - \widehat{\boldsymbol{\beta}}_{r_t^0})\|^2,
\end{aligned} \tag{A.137}$$

where the last equality holds because $\widehat{\boldsymbol{\beta}}_{r_t^0}^{\text{inf}}$ is an OLS estimator that satisfies the first order condition $\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{X}_t\widehat{\boldsymbol{\beta}}_{r_t^0}^{\text{inf}})'\mathbf{M}_{\widehat{F}}\mathbf{X}_t = \mathbf{0}_{1\times k}$. We now use the arguments provided in the proof of Lemma A.4 to take care of the estimation error coming from $\widehat{\mathbf{F}}$. We then apply the same steps as in Proof of Theorem 1 to bound $\|\mathbf{V}_t(\widehat{\boldsymbol{\beta}}_{r_t^0}^{\text{inf}} - \widehat{\boldsymbol{\beta}}_{r_t^0})\|^2$. It follows that w.p.a.1,

$$
\begin{aligned}
Q^{\text{inf}}(\widehat{\boldsymbol{\beta}}) - Q^{\text{inf}}(\widehat{\boldsymbol{\beta}}^{\text{inf}}) &= \frac{1}{NT}\sum_{t=1}^{T} \|\mathbf{V}_t(\widehat{\boldsymbol{\beta}}^{\text{inf}}_{r_t^0} - \widehat{\boldsymbol{\beta}}_{r_t^0})\|^2 + |O_p(\mathcal{C}_{NT}^{-2})| \\
&\geq \frac{1}{NT}\sum_{t=1}^{T} \|\mathbf{V}_t(\widehat{\boldsymbol{\beta}}^{\text{inf}}_{r_t^0} - \widehat{\boldsymbol{\beta}}_{r_t^0})\|^2 \\
&\geq \sum_{r=1}^{R} \lambda_{\min}(\mathbf{S}_{r,r})\|\widehat{\boldsymbol{\beta}}^{\text{inf}}_r - \widehat{\boldsymbol{\beta}}_r\|^2 \\
&\geq \rho_{\min}\sum_{r=1}^{R} \|\widehat{\boldsymbol{\beta}}^{\text{inf}}_r - \widehat{\boldsymbol{\beta}}_r\|^2 \\
&\geq \rho_{\min}\max_{r\in\mathcal{R}} \|\widehat{\boldsymbol{\beta}}^{\text{inf}}_r - \widehat{\boldsymbol{\beta}}_r\|^2, \qquad \text{(A.138)}
\end{aligned}
$$

where we have used that $\lambda_{\min}(\mathbf{S}_{r,r}) \geq \rho_{\min}$ w.p.a.1 by Assumption 1. Hence, since $Q^{\text{inf}}(\widehat{\boldsymbol{\beta}}) - Q^{\text{inf}}(\widehat{\boldsymbol{\beta}}^{\text{inf}}) = o_p(N^{-\delta})$ and $\rho_{\min} > 0$ by Assumption 1,

$$
\max_{r\in\mathcal{R}} \|\widehat{\boldsymbol{\beta}}^{\text{inf}}_r - \widehat{\boldsymbol{\beta}}_r\|^2 = o_p(N^{-\delta}), \qquad \text{(A.139)}
$$

which is what we wanted to show. ■

Proof of Theorem 2.

For part (a), we use

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{t\in\{1,\ldots,T\}} |\widehat{r}_t(\widehat{\boldsymbol{\beta}}_r) - r_t^0| > 0\right) \\
&\leq \mathbb{P}(\widehat{\boldsymbol{\beta}} \notin \mathcal{N}) + \mathbb{P}\left(\widehat{\boldsymbol{\beta}} \in \mathcal{N}, \sup_{t\in\{1,\ldots,T\}} |\widehat{r}_t(\widehat{\boldsymbol{\beta}}_r) - r_t^0| > 0\right) \\
&\leq \mathbb{P}(\widehat{\boldsymbol{\beta}} \notin \mathcal{N}) + \sum_{t=1}^{T} \mathbb{P}(\widehat{\boldsymbol{\beta}} \in \mathcal{N}, |\widehat{r}_t(\widehat{\boldsymbol{\beta}}_r) - r_t^0| > 0) \\
&\leq \mathbb{P}(\widehat{\boldsymbol{\beta}} \notin \mathcal{N}) + \sum_{t=1}^{T} \mathbb{P}(\widehat{\boldsymbol{\beta}} \in \mathcal{N}, \widehat{r}_t(\widehat{\boldsymbol{\beta}}_r) \neq r_t^0), \qquad \text{(A.140)}
\end{aligned}
$$

where, similarly to Proof of Lemma A.4,

$$\begin{aligned}\mathbb{P}(\widehat{\boldsymbol{\beta}} \in \mathcal{N}, \widehat{r}_t(\widehat{\boldsymbol{\beta}}_r) \neq r_t^0) &= \mathbb{E}[\mathbb{1}(\widehat{\boldsymbol{\beta}} \in \mathcal{N})\mathbb{1}(\widehat{r}_t(\widehat{\boldsymbol{\beta}}_r) \neq r_t^0)] \leq \mathbb{E}(\mathbb{1}(\widehat{r}_t(\widehat{\boldsymbol{\beta}}_r) \neq r_t^0)) \\ &\leq \sum_{r=1}^{R} \mathbb{E}(Z_t(\widehat{\boldsymbol{\beta}}_r)) = \sum_{r=1}^{R} \mathbb{P}(Z_t(\widehat{\boldsymbol{\beta}}_r) = 1) = o(N^{-\delta}). \end{aligned} \tag{A.141}$$

Hence, since $\mathbb{P}(\widehat{\boldsymbol{\beta}} \notin \mathcal{N}) = o(1)$ (Theorem 1), we can show that

$$\mathbb{P}\left(\sup_{t\in\{1,\ldots,T\}} |\widehat{r}_t(\widehat{\boldsymbol{\beta}}_r) - r_t^0| > 0\right) = o(1) + To(N^{-\delta}), \tag{A.142}$$

as required for (a).

Part (b) is a direct consequence of Lemmas A.5. The proof of the theorem is therefore complete. ∎

As a preparation for our next lemma and the proof of Theorem 3, we introduce the following notation:

$$\mathbf{Q}_{1,r} := \frac{1}{NT_r}\sum_{t=1}^{T} \mathbb{1}(r_t = r)(\mathbf{V}_t - \overline{\mathbf{U}}\overline{\mathbf{B}}_m\boldsymbol{\Gamma}_t)'\boldsymbol{\varepsilon}_t, \tag{A.143}$$

$$\mathbf{Q}_{2,r} := \frac{1}{NT_r}\sum_{t=1}^{T} \mathbb{1}(r_t = r)(\mathbf{V}_t - \overline{\mathbf{U}}\overline{\mathbf{B}}_m\boldsymbol{\Gamma}_t)'\mathbf{P}_{F^0}\boldsymbol{\varepsilon}_t, \tag{A.144}$$

$$\mathbf{Q}_{3,r} := \frac{1}{NT_r}\sum_{t=1}^{T} \mathbb{1}(r_t = r)(\mathbf{V}_t - \overline{\mathbf{U}}\overline{\mathbf{B}}_m\boldsymbol{\Gamma}_t)'(\mathbf{M}_{F^0} - \mathbf{M}_{\widehat{F}})\boldsymbol{\varepsilon}_t, \tag{A.145}$$

$$\mathbf{R}_{1,r} := \frac{1}{\sqrt{NT_r}}\sum_{t=1}^{T} \mathbb{1}(r_t = r)(\mathbf{V}_t' - \boldsymbol{\Gamma}_t'\overline{\mathbf{B}}_m'\overline{\mathbf{U}}')\overline{\mathbf{U}}\overline{\mathbf{B}}_m\boldsymbol{\gamma}_t, \tag{A.146}$$

$$\mathbf{R}_{2,r} := \frac{1}{\sqrt{NT_r}}\sum_{t=1}^{T} \mathbb{1}(r_t = r)(\mathbf{V}_t' - \boldsymbol{\Gamma}_t'\overline{\mathbf{B}}_m'\overline{\mathbf{U}}')\mathbf{P}_{F^0}\overline{\mathbf{U}}\overline{\mathbf{B}}_m\boldsymbol{\gamma}_t, \tag{A.147}$$

$$\mathbf{R}_{3,r} := \frac{1}{\sqrt{NT_r}}\sum_{t=1}^{T} \mathbb{1}(r_t = r)(\mathbf{V}_t' - \boldsymbol{\Gamma}_t'\overline{\mathbf{B}}_m'\overline{\mathbf{U}}')(\mathbf{M}_{F^0} - \mathbf{M}_{\widehat{F}})\overline{\mathbf{U}}\overline{\mathbf{B}}_m\boldsymbol{\gamma}_t. \tag{A.148}$$

We also need

$$\mathbf{b}_{1,r} := \lim_{T\to\infty} \frac{1}{T_r} \sum_{t=1}^{T} \mathbb{1}(r_t = r)(\boldsymbol{\Sigma}_{v,t} \begin{bmatrix} \boldsymbol{\beta}_r & \mathbf{I}_k \end{bmatrix} - \boldsymbol{\Gamma}_t' \mathbf{B}_m' \boldsymbol{\Omega}_u) \mathbf{P}_{-m} \boldsymbol{\Sigma}_u \mathbf{B}_m \gamma_t, \tag{A.149}$$

$$\mathbf{b}_{2,r} := \lim_{T\to\infty} \frac{1}{T_r} \sum_{t=1}^{T} \mathbb{1}(r_t = r)\sigma_{\varepsilon,t}^2 (\boldsymbol{\Sigma}_{v,t} \begin{bmatrix} \boldsymbol{\beta}_r & \mathbf{I}_k \end{bmatrix} - \boldsymbol{\Gamma}_t' \mathbf{B}_m' \boldsymbol{\Omega}_u) \mathbf{P}_{-m} \begin{bmatrix} 1 & \mathbf{0}_{1\times k} \end{bmatrix}', \tag{A.150}$$

$$\mathbf{b}_{3,r} := \lim_{T\to\infty} \frac{1}{T_r} \sum_{t=1}^{T} \mathbb{1}(r_t = r)(\boldsymbol{\Sigma}_{v,t} \begin{bmatrix} \boldsymbol{\beta}_r & \mathbf{I}_k \end{bmatrix} - \boldsymbol{\Gamma}_t' \mathbf{B}_m' \boldsymbol{\Omega}_u) \mathbf{B}_m \gamma_t, \tag{A.151}$$

$$\mathbf{b}_{4,r} := \lim_{T\to\infty} \frac{1}{T_r} \sum_{t=1}^{T} \mathbb{1}(r_t = r)\sigma_{\varepsilon,t}^2 \boldsymbol{\Gamma}_t' \mathbf{B}_m' \begin{bmatrix} 1 & \mathbf{0}_{k\times 1} \end{bmatrix}', \tag{A.152}$$

where $\mathbf{P}_{-m} := \mathbf{B}_{-m}(\mathbf{B}_{-m}'\boldsymbol{\Omega}_u\mathbf{B}_{-m})^{-1}\mathbf{B}_{-m}'$ and $\mathbf{B}_{-m}$ is as in the proof of Lemma A.1. Lemma A.6 below expresses the limits of $\mathbf{R}_{1,r}, \dots, \mathbf{R}_{3,r}$ and $\mathbf{Q}_{1,r}, \dots, \mathbf{Q}_{3,r}$ in terms of $\mathbf{b}_{1,r}, \dots, \mathbf{b}_{4,r}$.

Lemma A.6. *Suppose that Assumptions 1 and 2 hold. Then, as N, $T \to \infty$,*

(a) $\mathbf{R}_{3,r} = \sqrt{N} T_r^{-1/2} \mathbf{b}_{1,r} + O_p(T^{-1/2})$,

(b) $\mathbf{Q}_{3,r} = \sqrt{N} T_r^{-1/2} \mathbf{b}_{2,r} + O_p(T^{-1/2})$,

(c) $\|\mathbf{R}_{2,r}\| = O_p(N^{-1/2})$,

(d) $\mathbf{R}_{1,r} = \sqrt{N} T_r^{-1/2} \mathbf{b}_{3,r} + O_p(\mathcal{C}_{NT}^{-1})$,

(e) $\|\mathbf{Q}_{2,r}\| = O_p(N^{-1/2})$,

(f) $\mathbf{Q}_{1,r} = \frac{1}{\sqrt{NT_r}} \sum_{t=1}^{T} 1(r_t = r) \mathbf{V}_t' \boldsymbol{\varepsilon}_t - \sqrt{N} T_r^{-1/2} \mathbf{b}_{4,r} + O_p(\mathcal{C}_{NT}^{-1})$,

where $\mathcal{C}_{NT} := \max\{\sqrt{N}, \sqrt{T}\}$ as in Proof of Lemma A.4.

Proof: This proof is analogous to that of Lemma S.1 in Karabiyik et al. (2017) after interchanging the time and cross-sectional dimensions. We therefore omit the details.■

Proof of Theorem 3.

By using $\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r^0 = \widehat{\boldsymbol{\beta}}_r^{\text{inf}} - \widehat{\boldsymbol{\beta}}_r + \widehat{\boldsymbol{\beta}}_r^{\text{inf}} - \boldsymbol{\beta}_r^0$ and Lemma A.5,

$$\sqrt{NT_r}\|\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r^0 - (\widehat{\boldsymbol{\beta}}_r^{\text{inf}} - \boldsymbol{\beta}_r^0)\| = \sqrt{NT_r}\|\widehat{\boldsymbol{\beta}}_r^{\text{inf}} - \widehat{\boldsymbol{\beta}}_r\| = \sqrt{NT_r} o_p(N^{-\delta}). \tag{A.153}$$

Because $T_r = O(T)$ by Assumption 1, we have $\sqrt{NT_r} o_p(N^{-\delta}) = o(\sqrt{T}N^{1/2-\delta})$, which is $o(1)$ provided that $\sqrt{T}N^{1/2-\delta} = \sqrt{T}N^{-1/2}N^{1-\delta} = O(1)$, which it will be since $T^{-1}N \to \eta$ and $\delta > 1$ under the conditions of the theorem. Let us therefore consider $\sqrt{NT_r}(\widehat{\boldsymbol{\beta}}_r^{\text{inf}} - \boldsymbol{\beta}_r^0)$. The minimizer $\widehat{\boldsymbol{\beta}}^{\text{inf}}$ of $Q(\cdot)$ is simply the OLS solution that takes the regimes as known. The estimator $\widehat{\boldsymbol{\beta}}_r^{\text{inf}}$ of $\boldsymbol{\beta}_r^0$ is therefore given by

$$\widehat{\boldsymbol{\beta}}_r^{\text{inf}} := \left(\sum_{t=1}^T \mathbb{1}(r_t = r)\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t\right)^{-1} \sum_{t=1}^T \mathbb{1}(r_t = r)\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{y}_t. \tag{A.154}$$

From $\overline{\mathbf{C}\mathbf{B}}_m = \mathbf{I}_m$,

$$\mathbf{y}_t = \mathbf{X}_t\boldsymbol{\beta}_r^0 + \widehat{\mathbf{F}}\overline{\mathbf{B}}_m\boldsymbol{\gamma}_t - (\widehat{\mathbf{F}} - \mathbf{F}\overline{\mathbf{C}})\overline{\mathbf{B}}_m\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t = \mathbf{X}_t\boldsymbol{\beta}_r^0 + \widehat{\mathbf{F}}\overline{\mathbf{B}}_m\boldsymbol{\gamma}_t - \overline{\mathbf{U}\mathbf{B}}_m\boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t, \tag{A.155}$$

which in turn implies

$$\begin{aligned}\sqrt{NT_r}(\widehat{\boldsymbol{\beta}}_r^{\text{inf}} - \boldsymbol{\beta}_r^0) &= \left(\frac{1}{NT_r}\sum_{t=1}^T \mathbb{1}(r_t = r)\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t\right)^{-1} \\ &\quad \times \frac{1}{\sqrt{NT_r}}\sum_{t=1}^T \mathbb{1}(r_t = r)\mathbf{X}_t'\mathbf{M}_{\widehat{F}}(\boldsymbol{\varepsilon}_t - \overline{\mathbf{U}\mathbf{B}}_m\boldsymbol{\gamma}_t).\end{aligned} \tag{A.156}$$

From $T_r = O(T)$, we get $\mathcal{C}_{NT_r} = O(\mathcal{C}_{NT})$. By using this and some of the results obtained as a part of Proof of Lemma A.2,

$$\left\|\frac{1}{NT_r}\sum_{t=1}^T \mathbb{1}(r_t = r)\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t - \frac{1}{NT_r}\sum_{t=1}^T \mathbb{1}(r_t = r)\mathbf{V}_t'\mathbf{V}_t\right\| = O_p(\mathcal{C}_{NT}^{-2}), \tag{A.157}$$

where $(NT_r)^{-1}\sum_{t=1}^{T}\mathbb{1}(r_t = r)\mathbf{V}_t'\mathbf{V}_t \to \mathbf{\Theta}_r$ as $N, T \to \infty$ by Assumption 3. We can therefore show that

$$\frac{1}{NT_r}\sum_{t=1}^{T}\mathbb{1}(r_t = r)\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\mathbf{X}_t = \mathbf{\Theta}_r + o_p(1). \tag{A.158}$$

Let us now consider the numerator, which we write as

$$\frac{1}{\sqrt{NT_r}}\sum_{t=1}^{T}\mathbb{1}(r_t = r)\mathbf{X}_t'\mathbf{M}_{\widehat{F}}(\boldsymbol{\varepsilon}_t - \overline{\mathbf{U}}\overline{\mathbf{B}}_m\gamma_t) = \mathbf{Q}_r - \mathbf{R}_r \tag{A.159}$$

with obvious definitions of $\mathbf{Q}_r$ and $\mathbf{R}_r$. From $\mathbf{X}_t = \widehat{\mathbf{F}}\overline{\mathbf{B}}_m\mathbf{\Gamma}_t - \overline{\mathbf{U}}\overline{\mathbf{B}}_m\mathbf{\Gamma}_t + \mathbf{V}_t$, $\mathbf{M}_{\widehat{F}} = \mathbf{M}_{\widehat{F}^0}$ and Lemma A.6, we get

$$\begin{aligned}
\mathbf{Q}_r &:= \frac{1}{\sqrt{NT_r}}\sum_{t=1}^{T}\mathbb{1}(r_t = r)\mathbf{X}_t'\mathbf{M}_{\widehat{F}}\boldsymbol{\varepsilon}_t \\
&= \frac{1}{\sqrt{NT_r}}\sum_{t=1}^{T}\mathbb{1}(r_t = r)(\mathbf{V}_t - \overline{\mathbf{U}}\overline{\mathbf{B}}_m\mathbf{\Gamma}_t)'\mathbf{M}_{\widehat{F}^0}\boldsymbol{\varepsilon}_t \\
&= \frac{1}{NT_r}\sum_{t=1}^{T}\mathbb{1}(r_t = r)(\mathbf{V}_t - \overline{\mathbf{U}}\overline{\mathbf{B}}_m\mathbf{\Gamma}_t)'\boldsymbol{\varepsilon}_t \\
&- \frac{1}{NT_r}\sum_{t=1}^{T}\mathbb{1}(r_t = r)(\mathbf{V}_t - \overline{\mathbf{U}}\overline{\mathbf{B}}_m\mathbf{\Gamma}_t)'\mathbf{P}_{F^0}\boldsymbol{\varepsilon}_t \\
&- \frac{1}{NT_r}\sum_{t=1}^{T}\mathbb{1}(r_t = r)(\mathbf{V}_t - \overline{\mathbf{U}}\overline{\mathbf{B}}_m\mathbf{\Gamma}_t)'(\mathbf{M}_{F^0} - \mathbf{M}_{\widehat{F}})\boldsymbol{\varepsilon}_t \\
&= \mathbf{Q}_{1,r} - \mathbf{Q}_{2,r} - \mathbf{Q}_{3,r} \\
&= \frac{1}{\sqrt{NT_r}}\sum_{t=1}^{T}\mathbb{1}(r_t = r)\mathbf{V}_t'\boldsymbol{\varepsilon}_t - \sqrt{N}T_r^{-1/2}(\mathbf{b}_{4,r} + \mathbf{b}_{2,r}) + O_p(\mathcal{C}_{NT}^{-1}),
\end{aligned} \tag{A.160}$$

and

$$\begin{aligned}
\mathbf{R}_r &:= \frac{1}{\sqrt{NT_r}} \sum_{t=1}^{T} \mathbb{1}(r_t = r) \mathbf{X}_t' \mathbf{M}_{\widehat{F}} \overline{\mathbf{U}} \overline{\mathbf{B}}_m \gamma_t \\
&= \frac{1}{\sqrt{NT_r}} \sum_{t=1}^{T} \mathbb{1}(r_t = r) (\mathbf{V}_t' - \mathbf{\Gamma}_t' \overline{\mathbf{B}}_m' \overline{\mathbf{U}}') \overline{\mathbf{U}} \overline{\mathbf{B}}_m \gamma_t \\
&- \frac{1}{\sqrt{NT_r}} \sum_{t=1}^{T} \mathbb{1}(r_t = r) (\mathbf{V}_t' - \mathbf{\Gamma}_t' \overline{\mathbf{B}}_m' \overline{\mathbf{U}}') \mathbf{P}_{F^0} \overline{\mathbf{U}} \overline{\mathbf{B}}_m \gamma_t \\
&- \frac{1}{\sqrt{NT_r}} \sum_{t=1}^{T} \mathbb{1}(r_t = r) (\mathbf{V}_t' - \mathbf{\Gamma}_t' \overline{\mathbf{B}}_m' \overline{\mathbf{U}}') (\mathbf{M}_{F^0} - \mathbf{M}_{\widehat{F}}) \overline{\mathbf{U}} \overline{\mathbf{B}}_m \gamma_t \\
&= \mathbf{R}_{1,r} - \mathbf{R}_{2,r} - \mathbf{R}_{3,r} \\
&= \sqrt{N} T_r^{-1/2} (\mathbf{b}_{3,r} - \mathbf{b}_{1,r}) + O_p(\mathcal{C}_{NT}^{-1}),
\end{aligned} \tag{A.161}$$

which in turn implies

$$\begin{aligned}
&\frac{1}{\sqrt{NT_r}} \sum_{t=1}^{T} \mathbb{1}(r_t = r) \mathbf{X}_t' \mathbf{M}_{\widehat{F}} (\boldsymbol{\varepsilon}_t - \overline{\mathbf{U}} \overline{\mathbf{B}}_m \gamma_t) \\
&= \frac{1}{\sqrt{NT_r}} \sum_{t=1}^{T} \mathbb{1}(r_t = r) \mathbf{V}_t' \boldsymbol{\varepsilon}_t - \sqrt{N} T_r^{-1/2} \mathbf{b}_r + O_p(\mathcal{C}_{NT}^{-1}),
\end{aligned} \tag{A.162}$$

where $\mathbf{b}_r := \mathbf{b}_{2,r} + \mathbf{b}_{3,r} + \mathbf{b}_{4,r} - \mathbf{b}_{1,r}$.

By putting everything together, we obtain the following asymptotic representation for $\sqrt{NT_r}(\widehat{\boldsymbol{\beta}}_r^{\text{inf}} - \boldsymbol{\beta}_r^0)$:

$$\begin{aligned}
&\sqrt{NT_r}(\widehat{\boldsymbol{\beta}}_r^{\text{inf}} - \boldsymbol{\beta}_r^0) \\
&= \left(\frac{1}{NT_r} \sum_{t=1}^{T} \mathbb{1}(r_t = r) \mathbf{X}_t' \mathbf{M}_{\widehat{F}} \mathbf{X}_t \right)^{-1} \frac{1}{\sqrt{NT_r}} \sum_{t=1}^{T} \mathbb{1}(r_t = r) \mathbf{X}_t' \mathbf{M}_{\widehat{F}} (\boldsymbol{\varepsilon}_t - \overline{\mathbf{U}} \overline{\mathbf{B}}_m \gamma_t) \\
&= \mathbf{\Theta}_r^{-1} \left(\frac{1}{\sqrt{NT_r}} \sum_{t=1}^{T} \mathbb{1}(r_t = r) \mathbf{V}_t' \boldsymbol{\varepsilon}_t - \sqrt{N} T_r^{-1/2} \mathbf{b}_r \right) + o_p(1),
\end{aligned} \tag{A.163}$$

which together with (A.153), $T^{-1}T_r \to \pi_r$ (Assumption 1) and $T^{-1}N \to \eta$ imply

$$\begin{aligned}
&\sqrt{NT_r}(\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r^0) \\
&= \boldsymbol{\Theta}_r^{-1}\left(\frac{1}{\sqrt{NT_r}}\sum_{t=1}^{T}\mathbb{1}(r_t = r)\mathbf{V}_t'\boldsymbol{\varepsilon}_t - \sqrt{\eta}\pi_r^{-1/2}\mathbf{b}_r\right) + o_p(1), \qquad \text{(A.164)}
\end{aligned}$$

where we have again used that $\sqrt{T}N^{1/2-\delta} = O(1)$. The term $(NT_r)^{-1/2}\sum_{t=1}^{T}\mathbb{1}(r_t = r)\mathbf{V}_t'\boldsymbol{\varepsilon}_t$ is asymptotically normal by a cental limit law for independent processes. The limiting covariance matrix is given by

$$\begin{aligned}
&\lim_{N,T\to\infty}\mathbb{E}\left[\frac{1}{\sqrt{NT_r}}\sum_{t=1}^{T}\mathbb{1}(r_t = r)\mathbf{V}_t'\boldsymbol{\varepsilon}_t\left(\frac{1}{\sqrt{NT_r}}\sum_{t=1}^{T}\mathbb{1}(r_t = r)\mathbf{V}_t'\boldsymbol{\varepsilon}_t\right)'\right] \\
&= \lim_{N,T\to\infty}\frac{1}{NT_r}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathbb{1}(r_t = r)\mathbb{1}(r_s = r)\mathbb{E}(\mathbf{V}_t'\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_s'\mathbf{V}_s) \\
&= \lim_{N,T\to\infty}\frac{1}{T_r}\sum_{t=1}^{T}\mathbb{1}(r_t = r)N^{-1}\mathbb{E}(\mathbf{V}_t'\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{V}_t) \\
&= \lim_{T\to\infty}\frac{1}{T_r}\sum_{t=1}^{T}\mathbb{1}(r_t = r)\sigma_{\varepsilon,t}^2\boldsymbol{\Sigma}_{v,t} =: \boldsymbol{\Phi}_r, \qquad \text{(A.165)}
\end{aligned}$$

where we have used the fact that $\mathbf{V}_t$ and $\boldsymbol{\varepsilon}_t$ are serially and cross-sectionally uncorrelated (Assumption 2). They are also independent of each other with $\mathbb{E}(\varepsilon_{i,t}^2) =: \sigma_{\varepsilon,t}^2$, and therefore

$$\begin{aligned}
N^{-1}\mathbb{E}(\mathbf{V}_t'\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\mathbf{V}_t) &= N^{-1}\mathbb{E}[\mathbf{V}_t'\mathbb{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathbf{V}_t)\mathbf{V}_t] = N^{-1}\mathbb{E}[\mathbf{V}_t'(\sigma_{\varepsilon,t}^2\mathbf{I}_k)\mathbf{V}_t] \\
&= \sigma_{\varepsilon,t}^2 N^{-1}\mathbb{E}(\mathbf{V}_t'\mathbf{V}_t) \to \sigma_{\varepsilon,t}^2\boldsymbol{\Sigma}_{v,t} \qquad \text{(A.166)}
\end{aligned}$$

as $N \to \infty$. It follows that

$$\sqrt{NT_r}(\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r^0) \to_d N(\mathbf{0}_{k\times 1}, \boldsymbol{\Theta}_r^{-1}\boldsymbol{\Phi}_r\boldsymbol{\Theta}_r^{-1}) - \sqrt{\eta}\pi_r^{-1/2}\boldsymbol{\Theta}_r^{-1}\mathbf{b}_r \qquad \text{(A.167)}$$

as N, $T \to \infty$, which is what we wanted to show. ■

Proof of Theorem 4.

This proof uses the same basic steps as in the proof of Theorem 2 in Westerlund (2018). We begin by noting that under $\mathbf{\Sigma}_{v,t} = \mathbf{\Sigma}_{v,r}$ (Assumption 4), we have $\mathbf{\Theta}_r = T_r^{-1} \sum_{t=1}^{T} 1(r_t = r)\mathbf{\Sigma}_{v,r} = \mathbf{\Sigma}_{v,r}$, which does not depend on the subsample being used ("1" or "2"). By using this, the fact that the cardinality of $\mathcal{T}_\ell$ is given by $|\mathcal{T}_\ell| = T_r/2$ for $\ell \in \{1,2\}$, and the limiting representation of $\sqrt{NT_r}(\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r^0)$ provided in Proof of Theorem 3, we can show that

$$\sqrt{NT_r/2}(\widehat{\boldsymbol{\beta}}_{\ell,r} - \boldsymbol{\beta}_r^0) = \mathbf{\Theta}_r^{-1}(\mathbf{Q}_{\ell,r} - \mathbf{R}_{\ell,r}) + o_p(1), \tag{A.168}$$

where $\mathbf{Q}_{\ell,r}$ and $\mathbf{R}_{\ell,r}$ are defined analogously to $\mathbf{Q}_r$ and $\mathbf{R}_r$, respectively, in Lemma A.6 based on subsample ℓ. This implies

$$\begin{aligned}
&\sqrt{NT_r}(\widehat{\boldsymbol{\beta}}_r^{\text{spj}} - \boldsymbol{\beta}_r^0) \\
&= 2\sqrt{NT_r}(\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\beta}_r^0) - \frac{\sqrt{T_r}}{2\sqrt{T_r/2}}[\sqrt{NT_r/2}(\widehat{\boldsymbol{\beta}}_{1,r} - \boldsymbol{\beta}_r^0) + \sqrt{NT_r/2}(\widehat{\boldsymbol{\beta}}_{2,r} - \boldsymbol{\beta}_r^0)] \\
&= 2\mathbf{\Theta}_r^{-1}(\mathbf{Q}_r - \mathbf{R}_r) - \frac{1}{\sqrt{2}}[\mathbf{\Theta}_r^{-1}(\mathbf{Q}_{1,r} - \mathbf{R}_{1,r}) + \mathbf{\Theta}_r^{-1}(\mathbf{Q}_{2,r} - \mathbf{R}_{2,r})] + o_p(1).
\end{aligned} \tag{A.169}$$

We have already shown that under Assumption 4 $\mathbf{\Theta}_r$ does not depend on the subsample being used. The same is true for the bias terms. By using this, $|\mathcal{T}_\ell| = T_r/2$ and some of the results of the proof of Lemma A.6,

$$\mathbf{Q}_{\ell,r} - \mathbf{R}_{\ell,r} = \frac{1}{\sqrt{NT_r/2}} \sum_{t \in \mathcal{T}_{\ell,r}} \mathbb{1}(r_t = r)\mathbf{V}_t' \boldsymbol{\varepsilon}_t + \sqrt{2\eta}\pi_r^{-1/2}\mathbf{b}_r + o_p(1), \tag{A.170}$$

which in turn implies

$$\begin{aligned}
\sqrt{NT_r}(\widehat{\boldsymbol{\beta}}_r^{\text{spj}} - \boldsymbol{\beta}_r^0) &= \boldsymbol{\Theta}_r^{-1}[2(\mathbf{Q}_r - \mathbf{R}_r) - \frac{1}{\sqrt{2}}(\mathbf{Q}_{1,r} - \mathbf{R}_{1,r} + \mathbf{Q}_{2,r} - \mathbf{R}_{2,r})] + o_p(1) \\
&= \boldsymbol{\Theta}_r^{-1}\Bigg[\frac{2}{\sqrt{NT_r}}\sum_{t=1}^{T}\mathbb{1}(r_t = r)\mathbf{V}_t'\boldsymbol{\varepsilon}_t + 2\sqrt{\eta}\pi_r^{-1/2}\mathbf{b}_r \\
&\quad - \frac{1}{\sqrt{2}}\Bigg(\frac{1}{\sqrt{NT_r/2}}\sum_{t=1}^{T_r/2}\mathbb{1}(r_t = r)\mathbf{V}_t'\boldsymbol{\varepsilon}_t + \sqrt{2\eta}\pi_r^{-1/2}\mathbf{b}_r \\
&\quad + \frac{1}{\sqrt{NT_r/2}}\sum_{t=T_r/2+1}^{T_r}\mathbb{1}(r_t = r)\mathbf{V}_t'\boldsymbol{\varepsilon}_t + \sqrt{2\eta}\pi_r^{-1/2}\mathbf{b}_r\Bigg)\Bigg] + o_p(1) \\
&= \boldsymbol{\Theta}_r^{-1}\frac{1}{\sqrt{NT_r}}\sum_{t=1}^{T_r}\mathbb{1}(r_t = r)\mathbf{V}_t'\boldsymbol{\varepsilon}_t + o_p(1). \qquad \text{(A.171)}
\end{aligned}$$

The required result is a direct consequence of this and the asymptotic normality arguments of the proof of Lemma A.6. Because $\sigma_{\varepsilon,t}^2 = \sigma_{\varepsilon,r}^2$ and $\boldsymbol{\Sigma}_{v,t} = \boldsymbol{\Sigma}_{v,r}$ for all t such that $r_t = r$ under Assumption 4, we have

$$\boldsymbol{\Theta}_r = \lim_{T\to\infty}\frac{1}{T_r}\sum_{t=1}^{T}\mathbb{1}(r_t = r)\boldsymbol{\Sigma}_{v,t} = \boldsymbol{\Sigma}_{v,r}\lim_{T\to\infty}\frac{1}{T_r}\sum_{t=1}^{T}\mathbb{1}(r_t = r) = \boldsymbol{\Sigma}_{v,r}, \qquad \text{(A.172)}$$

$$\begin{aligned}
\boldsymbol{\Phi}_r &= \lim_{T\to\infty}\frac{1}{T_r}\sum_{t=1}^{T}\mathbb{1}(r_t = r)\sigma_{\varepsilon,t}^2\boldsymbol{\Sigma}_{v,t} = \sigma_{\varepsilon,r}^2\boldsymbol{\Sigma}_{v,r}\lim_{T\to\infty}\frac{1}{T_r}\sum_{t=1}^{T}\mathbb{1}(r_t = r) \\
&= \sigma_{\varepsilon,r}^2\boldsymbol{\Sigma}_{v,r}, \qquad \text{(A.173)}
\end{aligned}$$

and therefore the asymptotic covariance matrix simplifies to $\boldsymbol{\Theta}_r^{-1}\boldsymbol{\Phi}_r\boldsymbol{\Theta}_r^{-1} = \sigma_{\varepsilon,r}^2\boldsymbol{\Sigma}_{v,r}^{-1}$. ■