Weak (Proxy) Factors Robust Hansen-Jagannathan Distance For Linear Asset Pricing Models

Lingwei Kong *†

This version: November 15, 2019.
Please click here for latest version

Abstract

The Hansen-Jagannathan (HJ) distance statistic is one of the most dominant measures of model misspecification. However, the conventional HJ specification test procedure has poor finite sample performance, and we show that it can be size distorted even in large samples when (proxy) factors exhibit small correlations with asset returns. In other words, applied researchers are likely to falsely reject a model even when it is correctly specified. We provide two alternatives for the HJ statistic and two corresponding novel procedures for model specification tests, which are robust against the presence of weak (proxy) factors, and we also offer a novel robust risk premia estimator. Simulation exercises support our theory. Our empirical application documents the non-reliability of the traditional HJ test since it may produce counter-intuitive results, when comparing nested models, by rejecting a four-factor model but not the reduced three-factor model, while our proposed methods are practically more appealing and show support for a four-factor model for Fama French portfolios.

Keywords: asset pricing; identification robust statistics; reduced-rank models; model misspecification; rank test

---

*Tinbergen Institute & University of Amsterdam; personal website: sites.google.com/view/lingwei-kong; Email: l.kong@uva.nl

†Acknowledgment: Many thanks for my advisor, Frank Kleibergen, who not only helps me with his scientific advice and knowledge and many many insightful discussions and suggestions but also provides many useful tips for how to enjoy life along with research work. I am very grateful to Bart Keijsers, Eleni Aristodemou, Peter Boswijk, Yi He, and seminar participants at UvA for their comments on this project. I also want to thank Eva Janssens, and Emilie Berkhout for their organized brown-bag Ph.D. lunch seminars and my former colleague Merrick Li for his suggestions on writing well. I am also grateful for the comments from Laura Liu, Florian Gunsilius. I also thank Tim A. Kroencke, Martin Lettau, and Sydney Ludvigson for sharing their data, Nikolay Gospodinov, Raymond Kan, and Cesare Robotti for sharing their codes.

1 Introduction

Linear factor models have gained tremendous popularity in the empirical asset pricing literature, see e.g. Fama and French (1993), Lettau and Ludvigson (2001), Kan and Zhou (2004), Kan and Robotti (2008). The low dimensional factor structure in asset returns is well-documented (e.g., Kleibergen and Zhan (2015)), and Harvey et al. (2016) list hundreds of papers with factors that attempt to explain the cross-section of expected returns. Since so many factors are introduced, the asset pricing models are at best approximations, and proposed factors are at best proxies for some unobserved common factors.

Given the fact that there is a large pool of proposed factors and the linear asset pricing models are at best approximations, it is more appealing to determine whether or not the data reject a model, namely how good a model can approximate the data than to identify important factors (or factors with significant risk premia). The assessment of model performance is where specification tests play a role. To evaluate these factors and diagnose the model specifications, the HJ distance, proposed in Hansen and Jagannathan (1997), has emerged as one of the most dominant measures of model misspecification in empirical asset pricing literature (e.g., Jagannathan and Wang (1996), Kan and Zhou (2004)). However, the conventional HJ statistic can be unreliable.

We show that identical to the second pass R^2 studied in Kleibergen and Zhan (2015), the HJ statistic can be a misleading gauge of model fit and is not a satisfactory model selection tool. More importantly, we demonstrate that the specification test via the HJ statistic, which we refer to as the HJ specification test in the sequel , is not reliable when models are weakly identified. The lack of model identification can lead to spuriously significant risk premia (Kleibergen (2009), Bryzgalova (2016), Anatolyev and Mikusheva (2018)), and it affects specification tests concerning the full model as well. In an empirically relevant setting where proxy factors weakly correlate with the unobserved common factors, the HJ specification test can be size distorted even in large samples. The boundary, determined via the HJ specification test, between correct model specifications and misspecifications begins to blur in these so-called weak identification cases. Another potential issue would be the omitted-strong-factors problem. The resulting strong cross-sectional dependence in the error term can exaggerate all sorts of distortions when some included (proxy)[1] factors are weak (Kleibergen and Zhan (2015)).

One of the reasons for these failures is that sampling errors are no longer negligible asymptotically in the presence of weak (proxy) factors. Therefore, the conventional asymptotic justification may fail in empirically relevant settings, as weak (proxy) factors are commonly observed in many recent studies (e.g., Kleibergen (2009), Anatolyev and Mikusheva (2018)).

Recent papers have developed different techniques to incorporate some of these issues, most of which focus on the identification and inference of risk premia. Bryzgalova (2016) provides an estimation approach using shrinkage-based dimension-reduction technique which excludes weak/useless (proxy) factors. Anatolyev and Mikusheva (2018) propose an estimation procedure based on sample-splitting instrumental variables regression with proxies for the missing factor structure. Giglio and Xiu (2017) propose a three-pass estimation procedure and bypass the omitted factors bias by projecting risk premia of observed factors on those of strong factors extracted via principal components analysis (PCA). Alongside with these estimation techniques there are identification robust test statistics to correct for the overly optimistic statistical inference of the risk premia (e.g. Kleibergen (2009), Kleibergen and Zhan (2019), Kleibergen et al. (2018)). As for the specification tests of asset pricing models, Gospodinov et al. (2017) discuss the potential power loss of the $\mathcal{J}$ specification test when spurious/useless factors, which are completely uncorrelated with asset returns, are present.

This paper, instead of the spurious/useless factors, mainly focuses on problems resulting from weak (proxy) factors that are minorly correlated with asset returns, and aim to improve the performance of specification tests. To remedy the size distortion of the HJ test, we provide two alternatives for the HJ statistic, based on which we propose two easy-to-implement specification test procedures that are robust to the issues mentioned above.

[1]Sometimes researchers consider a latent factor structure in asset returns and regard included factors in empirical studies as proxies for priced latent common factors (e.g., Kleibergen and Zhan (2015), Giglio and Xiu (2017)), sometimes factors are assumed to be directly observed and priced weak common factors lead to problems (e.g., Kleibergen (2009), Anatolyev and Mikusheva (2018)). We mostly adopt the former idea, but our discussions and methods are also valid in the latter case, and we emphasize this by enclose the term proxy in brackets.

Our first proposed test procedure is a two-step Bonferroni-type method, and it is robust against identification issues when the number of asset returns is limited. This method takes into account the identification strength via a first-step confidence set, and we verify that it improves power compared with the $\mathcal{J}$ test.

Our second approach relies on a novel four-pass estimator, and the test procedure provides valid inference results in an asymptotic framework where the number of assets is comparable to the number of the observation periods. Our proposed four-pass estimator directly leads to a novel risk premia estimator, and thus we also contribute to the literature on estimation of risk premia in the presence of weak (proxy) factors and omitted factors. For linear asset pricing models, the conventional approach for estimating risk premia is known as the Fama-Macbeth (FM) two-pass estimation procedure (Fama and MacBeth (1973)), where risk premia estimates result from regressing average asset returns on first-pass estimated risk exposures (factor loadings βs). The two-pass procedure is easy to implement but can result in unreliable estimates and inference when some included factors are not strongly correlated with asset returns such that their risk exposures do not dominate corresponding sampling errors (Kleibergen (2009), Anatolyev and Mikusheva (2018)), which resembles the failure of the 2SLS estimator in instrumental variable regression when instruments are weak. Besides, Anatolyev and Mikusheva (2018) show that the missing factor structure exacerbates the weak (proxy) factor problem. We show our risk premia estimator is robust to the presence of both weak (proxy) factors and missing factors.

Our empirical application documents a strange behavior of the HJ test. Counter-intuitively, it can reject a four-factor model but not the corresponding three-factor model nested within the four-factor model. We attribute this behavior to the additional fourth factor being a weak proxy factor which leads to a undesirably high rejection rate of the HJ test. Our proposed procedures do not have this problem and reflect the factor structure in asset returns in a more informative way.

The paper is organized as follows: Section 2 reviews the basic model setting and shows the drawbacks of the HJ statistic; Section 3 and 4 introduce our proposed model specification test procedures, where Section 3 discusses our two-step Bonferroni-like method and Section 4 considers an approach that is valid with a double-asymptotic framework; Section 5 presents results of our empirical application.

We use the following notation: P_X stands for $X(X'X)^{-1}X'$ for a full column rank matrix X, M_X for $I - P_X$, $X^{\frac{1}{2}}$ for the upper triangular matrix from the Cholesky decomposition of the positive definite matrix X such that $X = (X^{\frac{1}{2}})'X^{\frac{1}{2}}$. Besides, in the following discussions, the notation would be more precise with N, T in the sub- or superscripts. For example to model the weak (proxy) factors, it might be reasonable to use notations such as $\beta_{T,N}, d_{g,T,N}, \gamma_{N,T}, \theta_{g,N,T}$ since the parameter values may change according to the sample dimensions in order to model the local to zero behavior. To avoid a cumbersome notation, we ignore these subscripts when there can be no misunderstanding.

2 Models and Problems

In this section, we introduce linear asset pricing models and the conventional model selection and specification test procedure based on the HJ distance. We use the term HJ statistic to denote the conventional squared HJ distance estimator, and to distinguish it from the other two estimators, our so-called HJN and HJS statistics, proposed in Sections 3 and 4.

We start by introducing our baseline model setting. We next derive asymptotic properties of the HJ statistic in the presence of weak (proxy) factors to clarify the problems we focus on.

2.1 Baseline model setting

We work with the linear asset pricing model because of its popularity in empirical studies. It imposes that all asset returns share common risk factors described by a small set of proposed factors. We regard the proposed factors in empirical studies as proxies for latent ones in the form as suggested in (Kleibergen and Zhan (2015)). Assumption 2.1 summarizes the baseline model settings:

Assumption 2.1. *For the $N \times 1$ vector of asset gross returns r_t, we assume that*

$$r_t = c + \beta f_t + u_t, \tag{1}$$

with f_t a $K \times 1$ vector of (possibly) unobserved zero-mean factors and u_t is an $N \times 1$ vector of *idiosyncratic components, and with a $K \times 1$ vector of proxy factors g_t*

$$f_t = d_g \left(g_t - \mu_g \right) + v_t, \tag{2}$$

where $\mu_g = \mathbb{E} g_t$, g_t is uncorrelated with u_t, v_t, d_g is of full rank and g_t, v_t, u_t are stationary with finite fourth moments. Furthermore,

$$c = \iota_N \lambda_0 + \beta \lambda_f, \tag{3}$$

where $\lambda_0 \neq 0$ is the zero-beta return, λ_f is a $K \times 1$ vector of risk premia, and the parameter space of λ_0, λ_f is compact.

Assumption 2.1 describes the beta representation of linear asset pricing models, and the DGP is similar to the one employed in Kleibergen and Zhan (2015). The moment conditions (or in other words the structural assumptions imposed on the constant term), $c = \iota_N \lambda_0 + \beta \lambda_f$, are commonly used in linear asset pricing model (e.g. Cochrane (2009)). If f_t is observed, then we would have perfect proxies with $d_g = I_K$, $v_t = 0$. Therefore, this model setting also embeds the model specification used in e.g. Kleibergen (2009), Anatolyev and Mikusheva (2018) where factors f_t are assumed to be observed. Using the observed factors g_t, model (1) can be rewritten as

$$r_t = c + \beta_g \bar{g}_t + \widetilde{u}_{g,t}, \tag{4}$$

with $\beta_g = \beta d_g, \bar{g}_t = g_t - \bar{g}, \bar{g} = \sum_{t=1}^T g_t / T, \widetilde{u}_{g,t} = u_{g,t} + \beta_g \left(\bar{g} - \mu_g \right) u_{g,t} = \beta v_t + u_t$. The estimation of the risk premia λ_g is usually accomplished by the FM two-pass estimator (Fama and MacBeth (1973), Shanken (1992)). In the first pass, the risk exposures β_g are estimated by regressing asset returns r_t on a constant and factors $\bar{g}_t$, and in the second pass the FM estimator $\widehat{\lambda}_g$ results from regressing average asset returns $\bar{r} = \sum_{t=1}^T r_t / T$ on an $N \times 1$ unity vector ι_N and the risk exposure estimates $\widehat{\beta}_g$.

Another well-known representation of asset pricing models is the stochastic discount factor (SDF) representation, based on which the HJ distance is defined. Cochrane (2009) shows that for linear asset pricing models, there is a corresponding SDF m_t that is linearly spanned by the latent factors

$$m_t(\theta) = F_t'\theta, \tag{5}$$

with $F_t = (1, f_t')'$, and re-scaled risk premia $\theta = (1/\lambda_0, -V_{ff}^{-1}\lambda_f/\lambda_0)$. The moment conditions (3) are then equivalent to the following ones

$$\mathbb{E}\left(m_t(\theta_0)r_t\right) = \iota_N, \tag{6}$$

with ι_N an $N \times 1$ unity vector with all entries equal to one. The population pricing errors which are the deviations from the moment conditions (6) are denoted by

$$e(\theta) = \iota_N - \mathbb{E}\left(m_t(\theta)r_t\right), \tag{7}$$

With a linear SDF (5) $e(\theta) = \iota_N - q\theta$, for $q = \mathbb{E}\left(r_t F_t'\right)$ a full column rank $N \times K$ matrix.

Hansen and Jagannathan (1997) (HJ) propose the minimum distance between the SDF of an asset pricing model and a set of correct SDFs as a measure of model misspecification. It also serves as a measure of goodness-of-fit. A smaller value of the HJ distance indicates a better model fit, and this is used for model selection. The population squared HJ distance δ has an explicit expression:

$$\delta^2 = \inf_{\theta} e(\theta)Q_r^{-1}e(\theta), \tag{8}$$

with $Q_r = \mathbb{E}\left(r_t r_t'\right)$ a full column rank $N \times N$ matrix. With a linear SDF, after some simple algebra, we can write the squared HJ distance explicitly as $\iota_N'\left(Q_r^{-1} - Q_r^{-1}q\left(q'Q_r^{-1}q\right)^{-1}q'Q_r^{-1}\right)\iota_N$, which is also numerically equal to $\iota_N'(Q_r^{-1} - Q_r^{-1}B(B'Q_r^{-1}B)^{-1}B'Q_r^{-1})\iota_N$ with $B = (c, \beta)$, and it is zero if and only if moment conditions (6) hold. Given the observed proxy factors g_t, the sample counterpart of the squared HJ distance, the HJ statistic, is

$$\widehat{\delta}_g^2 = \inf_{\theta_G} e_{g,T}(\theta_G)'\widehat{Q}_r^{-1}e_{g,T}(\theta_G). \tag{9}$$

In a linear asset pricing model, the sample pricing errors $e_{g,T}(\theta_G) = \sum_{t=1}^{T} e_{g,t}(\theta_G)/T = \iota_N -$

$q_{G,T}\theta_G$, $e_{g,t}(\theta_G) = \iota_N - r_t G_t'\theta_G$, $q_{G,T} = \sum_{t=1}^{T} r_t G_t'/T$, $\widehat{Q}_r = \sum_{t=1}^{T} r_t r_t'/T$, $G_t = (1, \bar{g}_t')'$, and the estimator resulting from this quadratic optimization problem is

$$\widehat{\theta}_G = \left(q_{G,T}'\widehat{Q}_r^{-1}q_{G,T}\right)^{-1} q_{G,T}'\widehat{Q}_r^{-1}\iota_N. \tag{10}$$

Jagannathan and Wang (1996) propose the HJ specification test by testing the moment conditions (6) via the HJ statistic. Under the null hypothesis that the moment conditions (6) hold, Jagannathan and Wang (1996) show that the asymptotic distribution of the HJ statistic follows a weighted sum of $\chi^2(1)$ random variables, which is because of the weighting matrix used in the HJ statistic. If we weight by the long-run covariance matrix of the sample pricing errors, we would have a regular chi-square-type limiting distribution. However, since we weight the HJ statistic differently with the second moment of asset returns as the weighting matrix, each of these $\chi^2(1)$ random variables has a weight different from one. Therefore, the critical values for the HJ statistic are obtained from the weighted sum of $\chi^2(1)$ random variables

$$\sum_{i=1}^{N-K-1} p_i x_i, \tag{11}$$

with x_i being independent $\chi^2(1)$ distributed random variables, and p_i being the positive eigenvalues of the matrix

$$\widehat{S}^{\frac{1}{2}}\left(\widehat{Q}_r^{-1} - \widehat{Q}_r^{-1}q_{G,T}\left(q_{G,T}'\widehat{Q}_r^{-1}q_{G,T}\right)^{-1}q_{G,T}'\widehat{Q}_r^{-1}\right)S_T^{\frac{1}{2}'},$$

with $\widehat{S} = S_T(\widehat{\theta}_G)$ and $S_T(\theta_G)$ a consistent estimator of the long-run variance matrix of the sample pricing errors $e_{g,T}(\theta_G)$ (for example, one may simple choose $S_T(\theta_G) = \frac{1}{T}\sum_{t=1}^{T} e_{g,t}(\theta_G)e_{g,t}(\theta_G)'$, $e_{g,t}(\theta_G) = \iota_N - r_t G_t'\theta_G$ provided that Assumptions 2.1-2.3 hold).

2.2 Problems and asymptotic properties

Our interest lies in the performance of the HJ statistic in the presence of weak identification issues, in particular, when observed proxies g_t are only weakly correlated with asset returns. Ahn and Gadarowski (2004) document the poor finite sample performances of the HJ specification test, and they argue that the size distortion is due to the critical value of the test which requires the estimation of the covariance matrix of $e_t(\theta)$ that performs badly with a limited number of observation periods. This is also consistent with the findings in Kleibergen and Zhan (2019), Kleibergen et al. (2018). In later parts, we show that not only in finite samples but also in large samples the HJ specification test can be severely size distorted in the presence of weak identification issues. We focus on *two issues* that can cause the potential deficiency of the HJ statistic.

1. Weak (proxy) factors. The HJ statistic depends on the estimator $\widehat{\theta}_G$. This is a GMM estimator based on the moment conditions (6) with weighting matrix $\widehat{Q}_r$. Similar to the FM risk premia estimator, this estimator can be constructed in two steps. In the first step the regressor in the SDF, q_g, is estimated via the estimator $q_{G,T}$, and in the second step $\widehat{\theta}_G$ results from regressing $\widehat{Q}_r^{-\frac{1}{2}}\iota_N$ on the first stage estimates $\widehat{Q}_r^{-\frac{1}{2}}q_{G,T}$. This close link with the FM estimator raises concerns for the quality of the θ_G estimator.

The FM estimator is unreliable under weak identification (e.g., Kan and Zhang (1999), Kleibergen (2009), Kleibergen and Zhan (2015), Kleibergen and Zhan (2019), Kleibergen et al. (2018), Anatolyev and Mikusheva (2018)). For linear asset pricing models, the identification strength is reflected by the rank of $B_g = (c, \beta_g)$ (e.g. Kleibergen and Zhan (2019)). The weak identification issues result from the empirical observation that the matrix B_g might be of reduced rank or near reduced rank. For example, this can happen when some (proxy) factors used in the estimation are weakly correlated with the asset returns. One way of modeling these weak (proxy) factors is to consider a sequence of models (or a sequence of parameter values) such that along the sequence, factor loadings are smaller and thus less informative for identifying risk premia. For example, suppose the β_g matrix is small, modeled by a drifting to zero sequence of order $O(1/\sqrt{T})$, then the sampling errors in the first stage estimator $\widehat{\beta}_g$, which are of the same order, are no longer negligible. These non-negligible sampling errors lead to the asymptotic invalidity of the FM estimator under weak (proxy) factors.

Following the same reasoning, the asymptotic justification for the estimator $\widehat{\theta}_G$ fails when q_g is small and comparable to its sampling error (see Theorem B.2). Since $q_g = \beta_g V_g$ with $V_g = \mathbb{E} g_t g_t'$, we model weak (proxy) factors using drifting to zero risk exposures (see Assumption 2.3) to mimic the behavior of a small q_g, which is in line with the literature on weak factors (e.g. Kleibergen

(2009)).

2. The missing factor structure. Omitted factors have received attention in recent studies (e.g., Kleibergen and Zhan (2015), Giglio and Xiu (2017), Anatolyev and Mikusheva (2018)). When we work with observed factors g_t in a latent factor setting, equation (4) suggests that the omitted factors v_t contribute to the error term $u_{g,t}$, and we also allow that unobserved factors explain most of the cross-sectional dependence in $u_{g,t}$. Similar to the discussion of the FM two-pass risk premia estimator in Anatolyev and Mikusheva (2018) , the missing factor structure could exacerbate the problem caused by the weak (proxy) factors and enlarge the bias in the estimator $\widehat{\theta}_g$ (see Theorem B.2), as the presence of an unobserved (missing) factor structure in the error terms creates the classical omitted-variables problem in the second step regression of the $\widehat{\theta}_g$ estimator when some (proxy) factors are weak.

Therefore, the HJ statistic may use an estimate that is potentially far away from the true value, and thus selection and inference based on the HJ statistic can be misleading. Before we continue to verify this, we make two assumptions.

Assumption 2.2. *Suppose u_t can be decomposed into two parts: a missing factor structure with a $K_z \times 1$ ($K_z \geq 0$) vector of unobserved strong factors z_t and weakly cross-sectional correlated noise e_t;*

$$u_t = \gamma z_t + e_t, \tag{12}$$

where (i) e_t are independent from $e_s, s \neq t$ and $g_{t'}, \forall t'$ with mean zero and bounded fourth moments $\sup_i \mathbb{E} e_{it}^4 < L < \infty$; (ii) denote $\Omega_e = \mathbb{E} e_t e_t'$, then $\lim_{N,T} tr(\Omega_e)/N = a > 0$ and $0 < l < \liminf_{N,T} \lambda_{\min}(\Omega_e) < \limsup_{N,T} \lambda_{\max}(\Omega_e) < L < \infty$ with $\lambda_{\min}(X)$ the smallest eigenvalues of matrix X and $\lambda_{\max}(X)$ the largest eigenvalues of matrix X; (iii) $\mathbb{E}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N \left(e_{it}^2 - \mathbb{E}e_{it}^2\right)\right|^4 < L < \infty$.

Assumption 2.3. *Denote Q_r by the second moments of r_t, $\eta = \left(\eta_{B_g}, \gamma\right)$ with $\eta_{B_g} = \left(c, \eta_{\beta_g}\right), \eta_{\beta_g} = \left(\beta_{g,1}, \sqrt{T}\beta_{g,2}\right), \beta_{g,i} = \beta d_{g,i}, i = 1,2$ being of dimension $N \times K_{g,i}, i = 1,2$ ($K_{g,1} + K_{g,2} = K, K_{g,2} \geq 0$) and full column rank matrices. (i) for fixed N, we assume $\eta'\eta$ is a $(1+K+K_z) \times (1+K+K_z)$ positive definite matrix; (ii) as N, T approach to infinity, $N^{-1}\eta'\eta$ converges to a $(1+K+K_z) \times (1+K+K_z)$ positive definite matrix.*

Our framework involves the observed factors g_t, and the omitted ones v_t, z_t. They have factor loadings β_g, β, γ respectively. Assumption 2.3 specifies the strengths of these factors. The loadings, $\beta_{g,2}$, of the $K_{g,2}$ proxy factors $g_{2,t}$ are modeled as drifting to zero sequences, so we call $g_{2,t}$ weak proxy factors. We do not restrict the strength of the priced latent factors f_t, and allow for weak priced latent factors. This assumption resembles the factor loading assumption in Anatolyev and Mikusheva (2018), but our risk exposure matrix (β_g, β, γ) is of reduced rank.

Assumption 2.2, which is similar to assumptions in Onatski (2012) and Anatolyev and Mikusheva (2018), does not fully rule out the cross-sectional dependence in the idiosyncratic error term e_t. This assumption allows the explanatory power of the cross-sectional variation in e_t to be comparable to the weak proxy factors when N,T increase proportionally, and the weak identification issue appears when the explanatory power of the proxy factors are roughly of the same order as e_t. When the cross sectional size N is fixed, the assumptions 2.2.(ii)-(iii) imposed on the noise term e_t hold naturally as long as Assumption 2.2.(i) holds and we can not really distinguish weak and strong factors. In later parts when N is fixed, we do not make use of the assumptions imposed on e_t but only the independence assumption (Assumption 2.2.(i)). We assume that e_t is independent across periods which is consistent with the efficient market hypothesis, and since empirical studies mostly use monthly or even less frequent data this is not a very unrealistic assumption.

Lemma 2.1. *Suppose Assumption 2.1, B.1 hold, let T increase to infinity then*

$$\widehat{B}_g =_d (c, \beta_g) + \psi_{B_g}/\sqrt{T}$$

where $\widehat{B}_g = q_{G,T}\widehat{Q}_G^{-1}, \widehat{Q}_G = \sum_{t=1}^T G_t G_t'/T, \psi_{B_g} = \left(\psi_c, \psi_{\beta_{g,1}}, \psi_{\beta_{g,2}}\right)$ and vec(ψ_{B_g}) being zero-mean normal random vectors.

Proof: see Appendix B.

Theorem 2.2. *Suppose Assumptions B.1 and 2.1-2.3 hold, let T increase to infinity with fixed N then the behavior of the HJ statistic is characterized by:*

$$T\widehat{\delta}_g^2 \to_d \widetilde{\psi}'_{B_g} M_{Q_r^{-\frac{1}{2}}\left(\eta_{B_g} + (0 \vdots \psi_{\beta_{g,2}})\right)} \widetilde{\psi}_{B_g}$$

with $\widetilde{\psi}_{B_g} = \psi_{B_g} Q_G \theta_G$.

Proof: see Appendix B.

Theorem 2.2 is derived assuming that the linear model is correctly specified, and it simply suggests that with strong factors, $K_{g,2} = 0$, the weighted sum of χ^2's provides a reasonable approximation (Corollary 2.2.1).

Corollary 2.2.1. *Suppose Assumption B.2 and the assumptions in Theorem 2.2 hold with* $K_{g,2} = 0$, *then*

$$T\widehat{\delta}_g^2 \sim_d \sum_{i=1}^{N-K-1} p_i x_i$$

with x_i *being independently* $\chi^2(1)$ *distributed random variables and* p_i *being the positive eigenvalues of the matrix* $\widehat{S}^{\frac{1}{2}} \left(\widehat{Q}_r^{-1} - \widehat{Q}_r^{-1} q_{G,T} \left(q_{G,T}' \widehat{Q}_r^{-1} q_{G,T} \right)^{-1} q_{G,T}' \widehat{Q}_r^{-1} \right) \widehat{S}^{\frac{1}{2}\prime}$.

Proof: this is a direct result of Theorem 2.2.

However, the conventional specification test procedure can be unreliable and suffer from severe size distortion even in large samples due to the irregular distribution of the HJ statistic (Corollary 2.2.2).

Corollary 2.2.2. *Suppose the assumptions in Corollary 2.2.1 hold with* $0 < K_{g,2} \leq K$,

$$\limsup_{N} \limsup_{T} \; \mathbb{P}\left(T\widehat{\delta}_g \geq \widehat{c}_{1-\alpha} \right) = 1$$

where $\widehat{c}_{1-\alpha}$ *is the conventional critical value derived from the distribution (11).*

Proof: see Appendix B.

Corollary 2.2.2 shows that under certain conditions the conventional specification test rejects the model specification with probability converging to one even when the moment conditions hold. Thus the conventional specification testing procedure based on the HJ statistic may mistake the "weak identification" resulting from the weak (proxy) factors for model misspecification, and leads to over-rejection when models are correctly specified.

2.3 Simulation exercises

We conduct simulation exercises to show that the HJ distance statistic as a model selection criterion might favor the presence of useless factors, and the HJ specification test suffers from severe size distortions.

In the first simulation exercise (Figures 1, 2), we calibrate the data generating process to match the data set of monthly gross asset returns on 25 size and book to market sorted portfolios from 1963 to 1998 and the three Fama French (FF) factors used by Lettau and Ludvigson (2001). The data is simulated in the following way: we simulate three proxy factors $g_t \sim \ i.i.d\ N(0, V_F)$, three omitted factors $v_t\ i.i.d\ \sim N(0.99V_F)$ and three strong factors are then generated by $f_t = 0.1 * g_t + v_t,$. V_F is calibrated to the sample covariance of the FF factors. We also generate three completely useless factors $w_t \sim i.i.d.N(0, V_F)$. We then generate returns via $r_t = \iota_N + \beta\lambda + \beta f_t + u_t$, $u_t \sim \ i.i.d\ N(0, V_u)$, where we set λ to be the sample risk premia estimated via the FM two-pass estimator, β is the sample slope parameter between the assets returns and FF factors, and V_u is the sample covariance of the residuals resulting from regressing asset returns on a constant and FF factors from the data.

Figure 1 compares the density functions of the simulated HJ statistics evaluated with various combinations of the factors g_t, f_t, w_t. For example, the black solid curve is drawn using three strong factors f_t, the black dashed curve is drawn with two strong factors f_{1t}, f_{2t}. Ideally, the black solid curve should be the most left, since the model with three strong factors should be most likely to be selected by the HJ statistic. However, comparing the red solid and black solid curves shows that adding additional useless factors leads to a shift of the distribution to the left, so it reduces the HJ statistic and leads to a "preferred model". The blue sold curve illustrates the density function of the HJ statistic of the model with three weak proxy factors. By construction, the moment conditions (6) are satisfied by the three weak proxy factors, and this model is correctly specified. If we compare the blue solid curve with the blacked dashed one which is constructed with only two strong factors, then the misspecified model with two strong factors is more likely to be selected. These observations imply that the HJ statistic is not a satisfying model selection tool.

The observations in Figure 1 show that values of the HJ statistic can not properly distinguish between weakly identified models and misspecified models. This is what motivates us to look further into the HJ specification test. Figure 2 compares two different approaches for approximating the

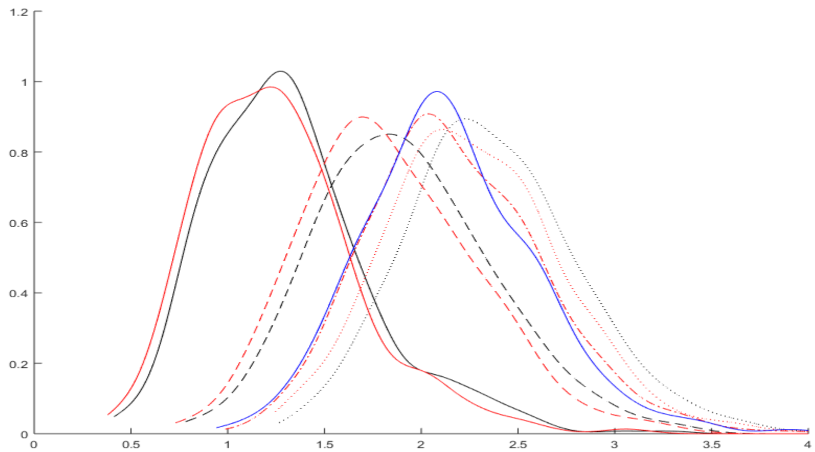

Figure 1: Density functions of the HJ statistic: (1) black solid: three strong factors; (2) black dashed: two strong factors; (3) black dotted: one strong factors; (4) blue solid: three weak proxy factors; (5) red solid: three strong factors and one useless factor; (6) red dashed: two strong factors and one useless factor; (7) red dotted: one strong factors and one useless factor; (8) red dash-dotted:one strong factors and two useless factors

distribution of the HJ statistic: one uses the conventional weighted sum of χ^2's, from which the critical values of the HJ statistic result, and another one uses the infeasible distribution from Theorem 2.2. The left-hand side panels of Figure 2 use three strong factors, while the right-hand side panels use three week ones.

The upper panels of Figure 2 show that both approximations for the distribution of the HJ statistic (the conventional weighted sum of $\chi^2(1)$s and the infeasible one from Theorem 2.2) are bad when T is small, and shift to the left compared with the density function of the HJ statistic. This observation is consistent with the one in Ahn and Gadarowski (2004) that the HJ specification test over-rejects correct model specifications in small samples. With a limited number of observation periods, not only sampling errors in the q_G estimators need to be taken into account but also those of other estimators such as the covariance estimator $\widehat{S}$ (e.g. Kleibergen and Zhan (2019), Kleibergen et al. (2018)). The infeasible distribution improves slightly by taking into account the sampling errors in the q_G estimators. When T is large, the randomness in the covariance estimators becomes small, but the sampling errors in $q_{G,T}$ still matter when proxy factors are weak. As shown in the lower panels of Figure 2, with a larger sample size the conventional approximation works fine when factors are strong but not when weak proxy factors are present. With weak proxy factors, the distribution of the HJ statistic is not properly approximated by the weighted sum of χ^2's even in large samples, and the HJ specification test is still likely to over-reject models when moment conditions do hold (Corollary 2.2.2).

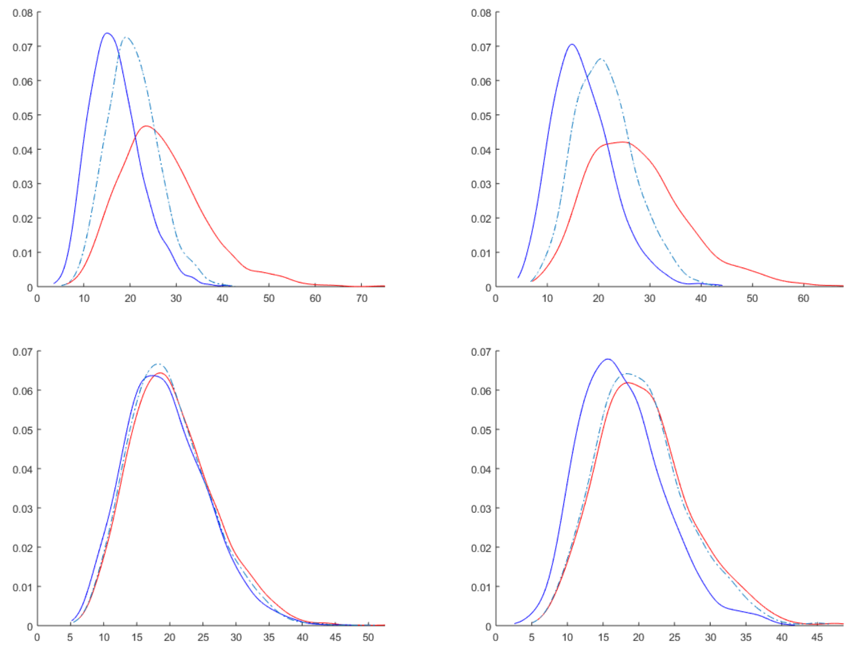

Figure 2: Density functions of the HJ statistic (red solid curve), weighted sum of χ^2's approximation (blue solid curves), asymptotic distribution based on Theory 2.2 (dot-dashed curve). Top left-hand side panel: $T = 100$, three strong factors f_t; top right-hand side panel: $T = 100$, three weak factors g_t; bottom left-hand side panel: $T = 1000$, three strong factors f_t; bottom right-hand side panel: $T = 1000$, three weak factors g_t.

Our second simulation exercise considers a simple single factor model in order to further illustrate the size distortion of the HJ specification test. We calibrated parameters to the data set from Kroencke (2017). We simulate the proxy factor $g_t \sim\ i.i.d.\ N(0, V_f/4)$, the omitted factor $v_t \sim\ i.i.d.\ N(0, Vff - d_g * (V_f/4) * d_g')$, the latent factor $f_t = d_g g_t + v_t$ and thus the variance of the f_t remain unchanged to different values of the d_g. We calibrate V_f to match the sample variance of the consumption growth factor from the data. The factor proposed in Kroencke (2017) has been shown to be weak (e.g. Kleibergen and Zhan (2019)). Therefore, we choose β to be of $10\widehat{\beta}$ with $\widehat{\beta}$ the sample regression parameter from Kroencke (2017), and thus r_t is generated with one single strong factor f_t via $r_t = \iota_N + \beta\lambda + \beta f_t + u_t$, $u_t \sim\ i.i.d\ N(0, V_u)$, where we match λ to the estimated risk premium from Kroencke (2017). We arbitrarily choose $d_g = 1.9; 0.9$ to mimic a strong and weak proxy factor.

Table 1 shows that the HJ specification tests have poor finite sample performances, size distortions increase with the number of assets and with relative weak proxy factors the distortion is more severe, and these observations support Corollary 2.2.2.

β	$N = 5$	$N = 10$	$N = 15$	$N = 31$
T=100, $d_g = 1.9$	0.5032	0.8824	0.9711	0.9992
T=10000, $d_g = 1.9$	0.1210	0.1486	0.2298	0.2238
T=100, $d_g = 0.9$	0.7132	0.9330	0.9814	1
T=10000, $d_g = 0.9$	0.5174	0.8906	0.8834	0.9978

Table 1: Rejection frequency table under the null (at 0.05 significance level) via the HJ specification test with critical values drawn from weighted sum of χ^2's

3 Specification test with limited N: HJS

As in previous discussions, the HJ specification test can not provide valid inference when weak (proxy) factors are present, and this is because the HJ specification test procedure ignores some non-negligible sampling errors in the estimates of parameters that can not be properly identified in the presence of weak identification issues. In this section, we suggest a numerically simple and identification robust test procedure which replace the estimates of these parameters with potential identification issues by those lying in a robust confidence set. This approach is related to the widely studied weak instrument problem where confidence sets with asymptotically correct coverage can be constructed for parameters with potential identification issues (e.g. Kleibergen (2005), Mikusheva (2010)).

3.1 HJS specification test

Our proposed HJS specification test procedure is conducted in three steps:

Step (1): Construct an identification robust confidence set, CS_{r,α_1}, for θ_G by inverting an Anderson-

Rubin (AR) type test statistic (e.g., Kleibergen (2009), Gospodinov et al. (2017)):

$$CS_{r,\alpha_1} = \{\theta \in \Theta : AR(\theta) \leq c_{1-\alpha_1}\}\,; \quad AR(\theta) = Te_T(\theta)'S_T^{-1}(\theta)e_T(\theta). \tag{13}$$

with $c_{1-\alpha_1}$ the $100(1-\alpha_1)\%$ percentile of the $\chi^2(N)$ distribution.

Step (2): Compute the HJS statistic:

$$\widehat{\delta}_g^* = \inf_{\theta \in CS_{r,\alpha_1}} \delta_{g,T}(\theta), \tag{14}$$

with $\delta_{g,T} = (\theta)e_{g,T}(\theta_G)'\widehat{Q}_r^{-1}e_{g,T}(\theta_G)$. To complete the construction of the HJS statistic we set $\widehat{\delta}_g^* = \infty$ when the confidence set CS_{r,α_1} is empty.

Step (3): This test would then reject the null hypothesis that moment conditions (6) hold if

$$T\widehat{\delta}_g^* > c_{1-\alpha}^*,$$

and the critical value is:

$$c_{1-\alpha}^* = \sup_{\theta \in CS_{r,\alpha_1}} c_{1-\alpha_2}^*(\theta), \tag{15}$$

where $c_{1-\alpha}^*(\theta)$ is the $100(1-\alpha_1)\%$ percentile of the weighted sum of N $\chi^2(1)$ random variables with weights being the non-zero eigenvalues of $S_T^{\frac{1}{2}}(\theta)\widehat{Q}_r^{-1}S_T^{\frac{1}{2}\prime}(\theta)$, α_1, α_2 are chosen such that $\alpha_1 > 0$, $\alpha_2 > 0$ and $(1-\alpha_1)(1-\alpha_2) = 1-\alpha$ with α the overall significance level.

Our HJS specification test procedure combines a less powerful but robust statistic (AR) with a non-robust one (HJ) to incorporate the model identification strength in our testing procedure, and in later discussion we show that this test improves performance in size (compared with the HJ test) and power (compared with the $\mathcal{J}$ test).

Before we proceed to show the size and power performances of the HJS specification test (Theory 3.2 and Theory 3.3), we first discuss the properties of the robust confidence set $CS_{r,\alpha}$. Kleibergen and Zhan (2019) study a similar robust risk premia confidence set using the GRS-FAR statistic. They show that this kind of set can be unbounded in certain cases. Therefore, for practical reason we restrict the parameter space Θ to be a compact set (Assumption 2.1), of which the robust confidence set is a subset. By construction, when the model is strongly identified we would expect the confidence set to shrink to a point as sample size grows, and when the model is weakly identified or even unidentified the diameter of this set can be arbitrarily large.

Lemma 3.1. *Suppose the assumptions in Corollary 2.2.1 hold, then*

$$\liminf_{T} \mathbb{P}\left(\theta_G \in CS_{r,\alpha}\right) \geq 1 - \alpha$$

proof: See Appendix B.

Lemma 3.1 implies that the confidence set covers the true value with the requested probability asymptotically even in the presence of weak (proxy) factors, which is essential for the correct size performance of the HJS test. This result holds under more general cases, for example it holds even when the model is not identified, and this correct coverage probability of the confidence set directly results from the correct size of the identification robust AR test statistic.

Theorem 3.2. *Suppose the assumptions in Lemma 3.1 hold,*

$$\limsup_{T} \mathbb{P}\left(T\widehat{\delta}_g^* \geq c_{1-\alpha}^*\right) \leq \alpha$$

proof: See Appendix B.

Theorem 3.2 shows that $c_{1-\alpha}^*$ provides a upper bound for the HJS statistic, which is also a upper bound for the HJ statistic as the HJ statistic is smaller than the HJS statistic by construction, and that the HJS specification test is size correct in the presence of weak (proxy) factors. The proof of it implies that the size property of the HJS specification test is a direct result of Lemma 3.1, and given that the lemma holds for more general conditions, we know that the HJS specification test can be extended to more general cases as well. Theorem 3.2 also implies that the HJS specification test is conservative, which is understandable as we use the infimum to construct the HJS statistic instead of the supremum. However given the diameter of the robust confidence set can be arbitrarily large, using the supremum can lead to size distortion (see Example B.1).

Even though it is conservative, the HJS specification test has better power performance compared with another well-know specification test, the $\mathcal{J}$ specification test. The $\mathcal{J}$ specification test statistic is also constructed based on the AR statistic such that

$$\mathcal{J} = \inf_{\theta} AR(\theta)$$

Gospodinov et al. (2017) show that the $\mathcal{J}$ specification test is size correct in the presence of spurious/useless factors, which means q_G is of reduced rank and the model is not identified, but it has a complete power loss in such cases. We extend their results to weakly identified models. Theory 3.3 shows that in both unidentified and weakly identified models, the $\mathcal{J}$ specification test suffers from power loss, while our HJS test still maintains proper power performance.

Theorem 3.3. *Suppose Assumptions B.1, 2.1, 2.2 hold, but instead of the correct proxy factors g_t, proxy factors $\widetilde{g}_t$ are used such that $\widetilde{g}_t$ is a $\widetilde{K} \times 1$ vector, $\left\|\iota_N - q_{\widetilde{G}}\theta_{\widetilde{G}}\right\| > a > 0, \forall\theta_{\widetilde{G}} \in \Theta$ with $q_{\widetilde{G}} = \mathbb{E}(r_t\widetilde{G}_t'), \widetilde{G}_t = (1, \widetilde{g}_t)$ and the model is misspecified. In addition, assume that the ψ_{B_g}, when we replace G_t with $\widetilde{G}_t$, in Lemma 2.1 satisfies that $\psi_{B_g} \sim N(0, Q_{\widetilde{G}} \otimes \Sigma)$ with $Q_{\widetilde{G}} = \mathbb{E}(\widetilde{G}_t\widetilde{G}_t')$ and Σ the covariance matrix of $\widetilde{u}_{g,t}$. Let $H = \left(\iota_N, q_{\widetilde{G}}\right)$.*

(i) (Gospodinov et al. (2017) Theorem 2, unidentified model under misspecification) Suppose H has a column rank $K+1-k$ for an integer $k \geq 1$, then we have

$$\mathcal{J} \preceq_d w_{k,i}$$

where $w_{k,i}$ is the smallest eigenvalue of $\mathcal{W}_k \sim \mathcal{W}_k\left(N-K-1+k, I_k\right)$ and $\mathcal{W}_k\left(N-K-1+k, I_k\right)$ denotes the Wishart distribution with $N-K-1+k$ degrees of freedom and a scaling matrix I_k. Furthermore,

$$\limsup_T \mathbb{P}\left(\mathcal{J} \geq c_{\chi^2_{N-K},1-\alpha}\right) \leq \alpha,$$
$$\liminf_T \mathbb{P}\left(T\widehat{\delta}^*_g \geq c^*_{1-\alpha}\right) = 1,$$

with $c_{\chi^2_{N-K},1-\alpha}$ the $1-\alpha$ quantile of χ^2_{N-K}.

(ii) (weakly identified model under misspecification) Suppose $(HQ_x)'*(HQ_x)$ *with* $Q_x = diag(I_{K+1-k}, \sqrt{T}I_k)$ *converges to a positive definite matrix, then we have*

$$\mathcal{J} \to_d w_{k,ii}$$

where $w_{k,ii}$ *is the smallest eigenvalue of* $\mathcal{W}_k \sim \mathcal{W}_k(\mu, N-K-1+k, I_k)$ *and* $\mathcal{W}_k(\mu, N-K-1+k, I_k)$ *denotes the non-central Wishart distribution with* $N-K-1+k$ *degrees of freedom and a scaling matrix* I_k*, a location parameter* μ *(*μ *is specified in the proof). Furthermore,*

$$\liminf_T \mathbb{P}\left(\mathcal{J} \geq c_{\chi^2_{N-K},1-\alpha}\right) < 1,$$
$$\liminf_T \mathbb{P}\left(T\widehat{\delta}^*_g \geq c^*_{1-\alpha}\right) = 1,$$

with $c_{\chi^2_{N-K},1-\alpha}$ *the* $1-\alpha$ *quantile of* χ^2_{N-K}*.*

proof: See Appendix B.

3.2 Simulation exercise

In this section, we conduct a simple simulation exercise with a single-factor model to evaluate the empirical rejection rates (the size and power performance) of our proposed HJS specification test. We calibrate the data generating process in our simulations to match the data set from Kroencke (2017). We simulate the factor $f_t \sim$ *i.i.d.* $N(0, V_f)$, where we set V_f to match the sample variance of the consumption growth factor. r_t is generated with one factor f_t via $r_t = \iota_N + \beta\lambda + \beta_\perp d + \beta d_g f_t + u_t$, $u_t \sim$ *i.i.d* $N(0, V_u)$, where we match λ to the estimated risk premium, β is the sample slope parameter between the assets returns and consumption growth factor, and V_u is the sample covariance of the residuals resulting from regressing asset returns on a constant and the consumption growth factor. $\beta_\perp$ is a vector which is orthogonal to ι_N, β and $\left\|\sqrt{T}\beta_\perp\right\| = 1$. We set $T = 100$. We use d_g to tune the identification strength of the factors in our simulation exercise where a larger d_g means a stronger factor, and d to tune the model misspecification level where a lager d means a larger deviation from the moment conditions (6) for our simulated data.

For the size performance comparison, we set $d = 0$ and thus moment conditions (6) hold for our simulated data. Figure 3 shows that the HJ specification test is highly size distorted and the distortion only drops down slightly when we increase the identification strength of the factor, while the HJS specification test has a better finite sample behavior and remains size correct.

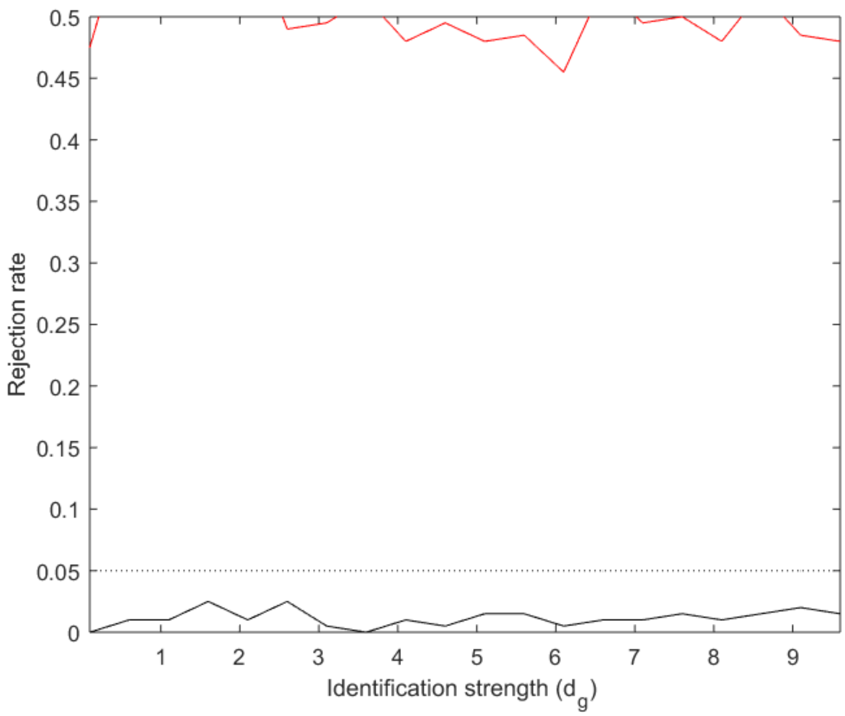

Figure 3: Size against strength of the factor (d_g): rejection frequency of the HJ specification test (red) and rejection frequency of the HJS specification test (black)

For the power performance comparison, we set $d_g = 0$ which means f_t only serves as a spurious factor. Figure 4 shows that the rejection frequency of the HJS specification test increases much faster compared with the one of the $\mathcal{J}$ specification test when the level of model misspecification (d) increases. The rejection frequency of the $\mathcal{J}$ specification test remains relatively small even when the HJS specification test rejection frequency is close to one, and this implies the HJS specification test has better power performance.

These observations support our theory and show that the HJS specification test has good performance in both size and power.

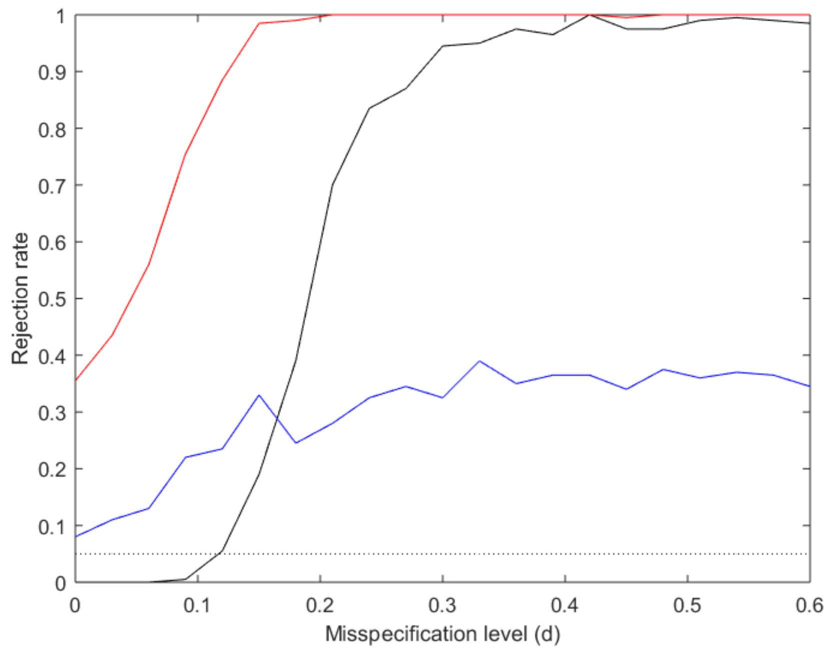

Figure 4: Power against the level of the model misspecificaion (d): rejection frequency of the HJ specification test (red), rejection frequency of the HJS specification test (black) and rejection frequency of the $\mathcal{J}$ specification test (blue)

4 Specification testing with large N: HJN

In the previous section, we construct the HJS statistic using a robust confidence set of θ_G since it is only weakly identified with a limited number of asset returns. The HJS specification testing procedure involves optimization steps, which is commonly done in practice through a grid search procedure. In this section, we provide another novel valid specification test statistic, the HJN statistic, which does not involve any time consuming optimization procedure. The construction of the HJN statistic uses a consistent θ_G estimator, and thus we first introduce our θ_G estimator and then the HJN statistic.

4.1 Four-pass estimator

When we work within a double-asymptotic framework such that both the number of time periods and the number of asset returns grow, weak (proxy) factors do not necessarily lead to a weak identification problem (Anatolyev and Mikusheva (2018)), which is similar to the case of many weak instruments that information about some parameters though limited aggregates slowly. Even though $\widehat{\theta}_G$ is not consistent (Theory B.2), another consistent estimator for θ_G can be constructed. With extended number of asset returns, we can estimate θ_G consistently by removing the missing factor structure via PCA and using an IV-type technique to correct for the remaining issues. The consistent estimator gives another way to construct a statistic for the HJ distance, based on which we propose a novel specification test statistic, our HJN statistic. In the following, we first introduce our four-pass θ_G estimator with extended number of asset returns, and thereafter provide the motivation for this four-pass procedure.

We propose the following steps to estimate θ_G with N base portfolios of gross returns r_t:

Step (1): Estimate $\widehat{c}, \widehat{\beta}_g$ in the linear observed-(proxy)-factor model (4) via OLS with N base portfolios of returns.

Step (2): Determine the omitted factor structure using the following two steps:

(2.1) Determine the number of factors, K_{vz}, in $\widehat{u}_{g,t} = r_t - \widehat{c} - \widehat{\beta}_g \bar{g}_t$ by

$$\widehat{K}_{vz} = \arg \min_{0 \leq j \leq K_{vz,\max}} \left(N^{-1} T^{-1} \lambda_j \left(\widehat{u}_g \widehat{u}_g' \right) + j\phi(N,T) \right) - 1$$

where $\lambda_j(A)$ is the j-th largest eigenvalue of a given matrix A, $\widehat{u}_g$ is $T \times N$ matrix stacked with the OLS residuals $\widehat{u}_{g,t}$, $K_{vz,\max}$ is an arbitrary upper bound for K_{vz} and $\phi(N,T)$ is a penalty function with the properties $\phi(N,T) \to 0, \phi(N,T)/(N^{-\frac{1}{2}} + T^{-\frac{1}{2}}) \to \infty$ (e.g., in later simulation exercise and empirical application we simply choose $\phi(N,T) = N^{-\frac{1}{4}} + T^{-\frac{1}{4}}$);

(2.2): Estimate the $T \times N$ common component matrix $cc = xb'$ stacked with the common components cc_t, $\widetilde{cc} = \widehat{x}\widehat{b}$, such that $\widehat{x}$ is equal to $\sqrt{T}$ times the eigenvector associated with the $\widehat{K}_{vz}$ largest eigenvalues of the matrix $\widehat{u}_g \widehat{u}_g'$, and $\widehat{b} = \widehat{x}' \widehat{u}_g / T$ corresponds with the OLS estimator regressing $\widehat{u}_g$ on $\widehat{x}$:

$$\left(\widehat{x}, \widehat{b} \right) = \arg \min_{b_i, x_t \ \ s.t. \ \sum_{t=1}^{T} x_t x_t'/T = I_{\widehat{K}_{vz}}} \sum_{i,t} \left(\widehat{u}_{g,it} - b_i' x_t \right)^2$$

Step (3): Split the sample into two non-overlapping subsamples along the time index and remove the missing factor structure from the regressors in the SDF of both subsamples:

$$\widetilde{q}_{G,T}^{(i)} = q_{G,T}^{(i)} - \widetilde{cc}_{G,T}^{(i)\prime}, i = 1, 2$$

where

$$q_{G,T}^{(1)} = \frac{1}{\lfloor \frac{T}{2} \rfloor} \sum_{t=1}^{\lfloor \frac{T}{2} \rfloor} r_t G_t'; \; q_{G,T}^{(1)} = \frac{1}{T - \lfloor \frac{T}{2} \rfloor} \sum_{t=\lfloor \frac{T}{2} \rfloor + 1}^{T} r_t G_t'$$
$$\widetilde{cc}_{G,T}^{(2)} = \frac{1}{\lfloor \frac{T}{2} \rfloor} \sum_{t=1}^{\lfloor \frac{T}{2} \rfloor} \widetilde{cc}_t G_t'; \; \widetilde{cc}_{G,T}^{(2)} = \frac{1}{T - \lfloor \frac{T}{2} \rfloor} \sum_{t=\lfloor \frac{T}{2} \rfloor + 1}^{T} \widetilde{cc}_t G_t'.$$

Step (4): We then use IV regression to derive two estimators, $\widetilde{\theta}_G^{(i)}, i = 1, 2$, where we use $\widetilde{q}_{G,T}^{(1)}$ as instrument for $\widetilde{q}_{G,T}^{(2)}$ and vice versa. Thereafter, our proposed four-pass estimator is then derived by taking average of both estimators:

$$\widetilde{\theta}_G = \sum_{i=1}^{2} \widetilde{\theta}_G^{(i)} / 2,$$

with $\widetilde{\theta}_G^{(1)} = \left(\widetilde{q}_{G,T}^{(1)\prime} P_{\widetilde{q}_{G,T}^{(2)}} \widetilde{q}_{G,T}^{(1)} \right)^{-1} \widetilde{q}_{G,T}^{(1)\prime} P_{\widetilde{q}_{G,T}^{(2)}} \iota_N$ and $\widetilde{\theta}_G^{(2)} = \left(\widetilde{q}_{G,T}^{(2)\prime} P_{\widetilde{q}_{G,T}^{(1)}} \widetilde{q}_{G,T}^{(2)} \right)^{-1} \widetilde{q}_{G,T}^{(2)\prime} P_{\widetilde{q}_{G,T}^{(1)}} \iota_N$.

Our estimation approach for θ_G resolves the problems of the missing factor structure and the weak (proxy) factors simultaneously. We make use of the results from Bai and Ng (2002), Bai (2003) and Giglio and Xiu (2017) in step (2) to recover the common components in the error terms using principal component analysis, and we use the instrument variable idea applied for the factor models, which is used in Anatolyev and Mikusheva (2018), in step (4) to solve potential endogeneity issues. Compared with the estimator proposed in Anatolyev and Mikusheva (2018), our proposed estimator relaxes the restrictions on the number of omitted factors and the restrictions on the rank of the loadings of all the factors present in the model.

To illustrate why our proposed procedure is robust against weak (proxy) factors and a missing factor structure, we start by comparing it with the conventional θ_G estimator. To do so, we first rewrite equation (6) ($\iota_N = \mathbb{E}q_{G,T}\theta_G$) as

$$\iota_N = q_{G,T}\theta_G - \epsilon_{q_G}\theta_G,$$

with $\epsilon_{q_G} = (q_{G,T} - \mathbb{E}q_{G,T})$ which is correlated with $q_{G,T}$. The term ϵ_{q_G} vanishes asymptotically, and so it is dominated by $q_{G,T}$ when all proxy factors are strong. The conventional estimator, which results from regressing ι_N on $q_{G,T}$, is then valid in large samples, since ϵ_{q_G} becomes negligible. However, if some (proxy) factors are weak, some columns of $q_{G,t}$ are of the same order as ϵ_{q_G}, then there would be a classic endogeneity problem if we simply regress ι_N on $q_{G,T}$.

To solve the endogeneity problem, a valid instrument can be constructed in our framework with a sample splitting technique and this idea is also employed in Anatolyev and Mikusheva (2018). Given the independence of the e_t from non-overlapping sub-samples, $q_{G,T}^{(1)}$ can serve as an instrument for $q_{G,T}^{(2)}$ and vice versa when there is no missing factor structure ($K_{vz} = 0$) and this is the starting point of our proposed procedure. When there is a missing factor structure with factors that might be correlated across time, $q_{G,T}^{(1)}$ is no longer a valid instrument for $q_{G,T}^{(2)}$. Therefore, we use $\widetilde{q}_{G,T}^{(i)}, i = 1, 2$ which results from removing the missing factor structure from $q_{G,T}^{(i)}, i = 1, 2$. By doing so, $\widetilde{q}_{G,T}^{(1)}$ is asymptotically uncorrelated with $\widetilde{\epsilon}_{q_G}^{(2)}$, and is a valid instrument.

As shown in Theorem 4.1, our estimation procedure provides $\min\{\sqrt{T}, \sqrt{N}\}$-consistent results for θ_G, of which a non-linear transformation leads to a consistent risk premia estimator (Corollary 4.1).

Theorem 4.1. *Suppose Assumptions 2.1 - 2.3, C.1 - C.9 hold, and $N/T \to c$.*

$$\sqrt{NT} Q_{B_g,T}^{-1} \left(\widetilde{\theta}_G - \theta_G \right) \to O_p(1),$$

with $Q_{B_g,T} = diag(I_{1+K_{g,1}}, \sqrt{T} I_{1+K_{g,1}})$.[2]

Proof: See Appendix C.3 .

[2] With some additional regular assumptions, we can construct $\widehat{\Sigma}_{\theta_G}$ such that $\widehat{\Sigma}_{\theta_G}^{1/2} \left(\widetilde{\theta}_g - \theta_G \right) \to_d N(0, I)$. See Appendix C.3.

Corollary 4.1.1. *Suppose the assumptions in Theorem 4.1 hold, then*

$$\sqrt{NT}Q_x^{-1}\left(\widetilde{\lambda}_g - \lambda_g\right) \to O_p(1),$$

with $\widetilde{\lambda}_g = -V_g diag(0_{K\times 1}, I_K)\widetilde{\theta}_G/\widetilde{\theta}_{G,1}$. *Proof: this is a direct result of Theorem 4.1.*

4.2 HJN specification test

Kleibergen and Zhan (2018) study risk premia on mimicking portfolios by projecting non-traded factors on traded base portfolios, and then carry out identification robust tests using a set of testing portfolios. We use a similar idea to construct the HJN specification test:

Step (1) Estimate θ_G from a set of N *base portfolios* r_t of asset returns using our proposed four-pass estimator $\widetilde{\theta}_G$

Step (2) Estimate q_G from a set of *testing portfolios* R_t of n asset returns and n is fixed such that $\widetilde{q}_G = \frac{1}{T}\sum_{t=1}^{T} R_t G_t'$

Step (3) The HJN statistic is

$$\widetilde{\delta}_g^2 = \widetilde{e}_T' \widehat{Q}_R^{-1} \widetilde{e}_T,$$

with sample pricing errors $\widetilde{e}_T = \iota_n - \widetilde{q}_G\widetilde{\theta}_G$, $\widehat{Q}_R = \frac{1}{T}\sum_{t=1}^{T} R_t R_t'$.

Step (4) This test would reject the null hypothesis that moment conditions (6) hold if

$$T\widetilde{\delta}_g > \widetilde{c}_{1-\alpha},$$

where $\widetilde{c}_{1-\alpha}$ is the $1-\alpha$ quantile of the weighted sum of N $\chi^2(1)$ random variables with weights being the positive eigenvalues of the matrix $\widetilde{S}^{\frac{1}{2}}(\widehat{Q}_R)^{-1}\widetilde{S}^{\frac{1}{2}\prime}$ with $\widetilde{S}$ a consistent estimator of the long-run variance matrix of the sample pricing errors $e_{T,R}(\theta_G) = \iota_n - \widetilde{q}_G\theta_G$.

Remark: the HJN specification test does not require the base portfolios and testing portfolios to be non-overlapping.

Corollary 4.1.2. *Suppose the assumptions in Theorem 4.1 hold, then*

$$T\widetilde{\delta}_g^2 \sim_d \sum_{i=1}^{n} p_i x_i$$

with x_i being independently $\chi^2(1)$ distributed random variables and p_i being the positive eigenvalues of the matrix $\widetilde{S}^{\frac{1}{2}}(\widehat{Q}_R)^{-1}\widetilde{S}^{\frac{1}{2}\prime}$ with $\widetilde{S}$ a consistent estimator of the long-run variance matrix of the sample pricing errors $e_{T,R}(\theta_G)$ (for example, one may simple choose $\widetilde{S} = \frac{1}{T}\sum_{t=1}^{T} e_{g,t,R}(\widetilde{\theta}_G)e_{g,t,R}(\widetilde{\theta}_G)', e_{g,t,R}(\theta_g) = \iota_N - R_t G_t' \theta_g$).

Proof: See Appendix C.3.

Theorem 4.2 shows that our HJN specification test is size correct even with weak (proxy) factors.

Theorem 4.2. *Suppose the assumptions in Theorem 4.1 hold,*

$$\limsup_T \mathbb{P}\left(T\widetilde{\delta}_g \geq \widetilde{c}_{1-\alpha}\right) = \alpha$$

with $\widetilde{c}_{1-\alpha}$ the $1-\alpha$ quantile of the weighted sum of N $\chi^2(1)$ random variables with being the positive eigenvalues of the matrix $\widetilde{S}^{\frac{1}{2}}(\widehat{Q}_R)^{-1}\widetilde{S}^{\frac{1}{2}\prime}$ with $\widetilde{S}$ a consistent estimator of the long-run variance matrix of the sample pricing errors $e_{T,R}(\theta_G)$ (for example, one may simple choose $\widetilde{S} = \frac{1}{T}\sum_{t=1}^{T} e_{g,t,R}(\widetilde{\theta}_G)e_{g,t,R}(\widetilde{\theta}_G)', e_{g,t,R}(\theta_g) = \iota_N - R_t G_t' \theta_g$).

proof: This is a direct result from Corollary 4.1.2.

4.3 Simulation exercises and empirical applications

Similar to section 3.2, we again evaluate the empirical rejection rates of our HJS specification test via simulation exercises.

We calibrate to the data set used in Anatolyev and Mikusheva (2018): the monthly returns on 100 Fama-French portfolios sorted by size and book-to-market and three Fama-French factors (g_t). From the portfolio returns we obtain the first four principal components (PC), and we regard the first three PCs as priced latent common factors, f_t, and the fourth one as the omitted factor, z_t. With normalization $(f,z)'(f,z)/T = I_4$, we set the variance of these factors to be 1. We regress demeaned returns on f_t, z_t for their risk exposures, and calculate the sample mean $\mu_{\beta\gamma}$ and the sample variance $V_{\beta\gamma}$ of the risk exposures. We compute the sample variance $\sigma_e^2 I_N$ of the residuals after regressing returns on f_t, z_t. To maintain the relation between observed factors g_t and PCs f_t, we regress f_t on the three Fama French factors to obtain the slope d_g and residual covariance matrix V_v, and d_g captures the quality of the proxy factors.

We then simulate our data in the following way. In the first step we simulate observed factors from i.i.d $N(0, I_3)$ and latent factors f_t are generated by $Ad_g g_t + v_t$ with v_t simulated as i.i.d $N\left(0, (I-A)d_g d_g'(I-A) + V_v\right)$. A is a diagonal matrix which we use to adjust the strengths of our (proxy) factors, and we set $A = \text{diag}(I_2, d_\alpha)$ in our simulations with d_α tuning the strength of the simulated factors. As for the corresponding risk exposures, we use $(\beta_i', \gamma_i')' \sim$ i.i.d $N(\mu_{\beta\gamma}, V_{\beta\gamma})$. Then in the end, we generate $r_t = \iota_N + \beta_\perp d + \beta\lambda + \beta f_t + \gamma z_t + e_t$, $e_t \sim$ $i.i.d$ $N(0, \sigma_e^2 I_N)$, where we match λ to the estimated risk premia resulting from the data. $\beta_\perp$ is a vector which is orthogonal to ι_N, β and $\|\beta_\perp\| = 1$. Similar to the previous simulation setting in Figure 4 we use d to tune the model misspecification level and $d = 0$ when we simulate size curves.

In our simulations, we fix $N = 100$. For the HJN specification test we use all the simulate 100 asset gross returns to form the base portfolios and the first 25 to form the testing portfolios, and we use the testing portfolios for the conventional HJ specification test.

Figure 5 compares the size curves of the HJ specification test and of the HJN specification test. It shows that the HJ specification test is highly size distorted even when d_α is large (left hand side panel of Figure 5) and the size distortion increases when proxy factors become weaker (smaller d_α), while the HJN specification test roughly remains size correct. Even with relatively large number of time periods, the conventional HJ specification test still over-reject. Observations from Figure 5 also seem to imply that our HJN specification test tends to under-reject in finite samples, and to show it performs well when the model is misspecified we also simulate power curves in Figure 6.

Figure 6 shows power curves of the HJ specification test and the HJN specification test respectively. The left hand side panel of Figure 6 uses $d_\alpha = 0.5$ to mimic one weak proxy factor, while all proxy factors in the right hand side panel are strong with a larger value of d_α. Figure 6 shows that the HJN specification test has proper power performance regardless of the presence of weak (proxy) factors, and rejection frequency increases faster when the proxy factor is stronger.

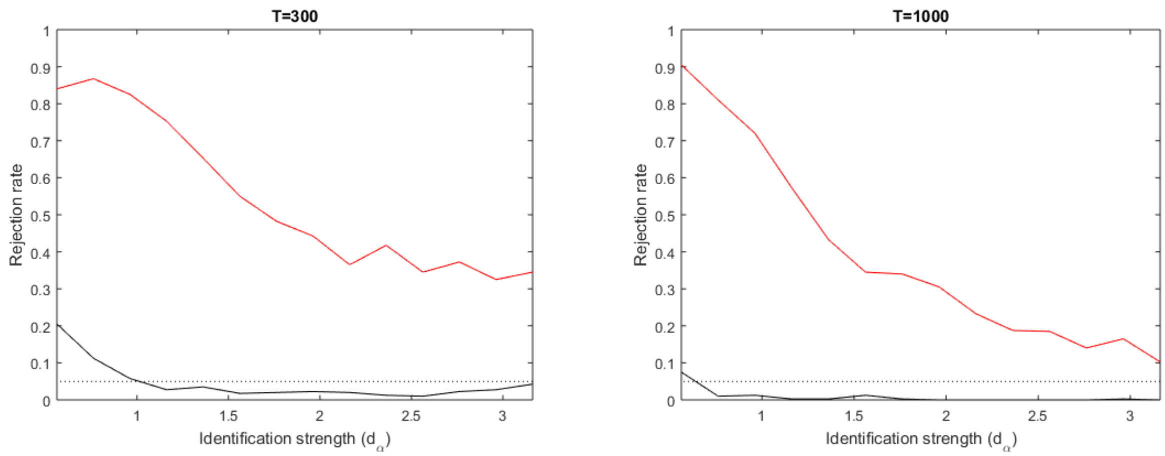

Figure 5: Size against strength of the factor (d_α): rejection frequency of the HJ specification test (red) and rejection frequency of the HJN specification test (black). Left hand side panel: T=300; right hand side panel: T=1000

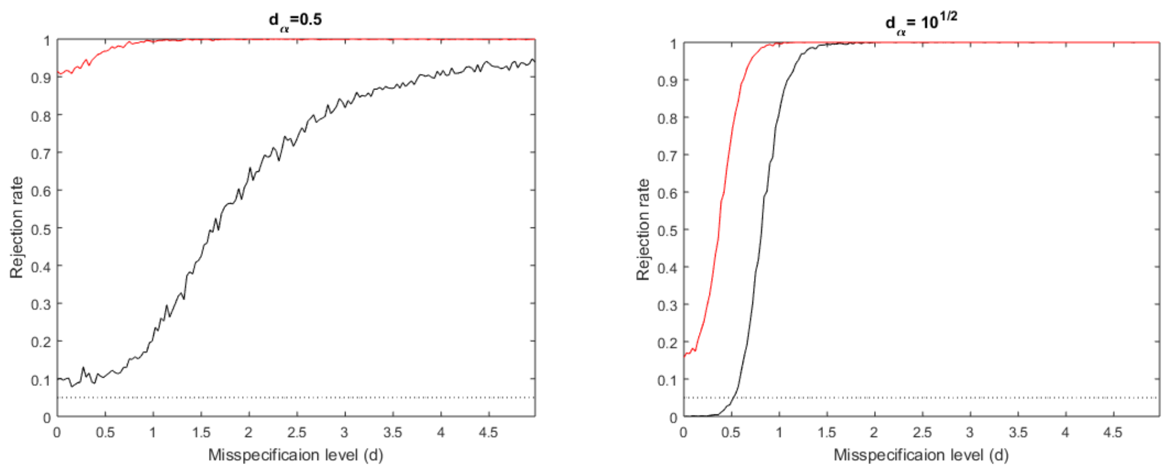

Figure 6: Power against the level of the model misspecificaion (d): rejection frequency of the HJ specification test (red) and rejection frequency of the HJS specification test (black). Left hand side panel: $d_\alpha = 0.5$; Right hand side panel: $d_\alpha = \sqrt{10}$.

5 Empirical Application

We apply our proposed test procedures on the data set of monthly returns on 100 Fama-French portfolios sorted by size and book-to-market and the three Fama-French factors (market, SmB, HmL) and the momentum factor.

An intuitive measure for the factor structure in asset returns is total variation of the asset returns explained by the principal components[3] (Kleibergen and Zhan (2015)). We construct the

[3]This corresponds with the nuclear norm of the demeaned asset returns

spectral decomposition of the sample covariance matrix of the 100 portfolio returns, and denote $\lambda_1 > \lambda_2 > \cdots$ the characteristic roots (or eigenvalues of the PCs of asset returns) in descending order. We use the characteristic roots ratios (CRRs) $\lambda_i / \left(\sum_j \lambda_j \right), i = 1, 2, 3, 4$, which represent the total variation of the portfolio returns explained by the first four PCs respectively, to check the factor structure of portfolio returns (see Figures 7 and 8).

Figures 7 and 8 also report the p-values of specification tests (HJ and HJN) with respect to a three-FF-factor model and a four-factor (adding the momentum) model from 1963-09 to 2019-08 using rolling windows of 240 and 120 months respectively[4]. For the HJN specification test we use all the 100 asset returns to form the base portfolios and the first 25 to form the testing portfolios, and we use the testing portfolios for the conventional HJ specification test. They also report measures for the presence of a factor structure in the asset returns: the fraction of the total variation of the portfolio returns that is explained by their principal components.

Figures 7 shows that when comparing nested models, the HJ test can produce counter-intuitive results by rejecting a four-factor model but not the reduced three-factor model (see points near the coordinate '2015-01'). This is an unfortunate outcome since the four-factor mode apparently embeds the three-factor model and if the four-factor model is rejected we would expect the three-factor model to be rejected as well. We attribute this strange behavior to the momentum factor having only weak correlation with the returns and thus inducing a larger rejection rate of the HJ test, while our HJN specification test does not have such problem.

The HJN specification test also captures changes in the factor structure of asset returns in a more sensible way compared with the HJ test. As shown in Figure 7, when CRRs vary in different time periods (e.g., the total variation of the portfolio returns is mostly explained by the first PC for points near the coordinate '2000-01' while the other PCs only account for a much lower percentage of the variation), the HJN specification tests reflect the changes in the factor structure of asset returns with variations in p-values of tests of a four-factor model, while the HJ specification tests reject both three-factor and four-factor models for most time periods and is not informative for the factor structure of asset returns.

[4]We first choose the window size of 240 months in Figure 7, because our simulations suggest sample size around 300 seems to be enough for carrying out our tests properly.

Both of the HJ and the HJN tests in Figure 7 seem to have larger p values near the coordinate '2015-01' while the patterns of characteristic roots ((a) and (b)) seem to be rather stable. This is because a 240-month window size is a bit too long, and some changes in the factor structure might be averaged out and thus not detected by CRRs. We choose a smaller rolling window size (120 months) in Figure 8, and it shows the change in the factor structure (Figure 8.(a)) after the coordinate '2010-01'. Similar to what we observe in Figure 7, Figure 8 also shows our HJN specification test respond to the factor structure in the asset returns in a more informative way, while the HJ tests only report small p values for most time periods for both three- and four-factor models.

We observe in Figures 7 and 8 that the HJN specification tests in some rolling windows do not reject a four-factor model. To further study this observation, Table 2 reports results based on the data from 1977-08 to 2019-08. We see in Table 2 that both the HJ and HJN specification tests reject the three-factor model, and while the HJ specification test rejects the four-factor model, the HJN specification test does not reject it. Our HJS specification test seems to be a bit conservative and does not reject both models in this application. The estimates for the four-factor model using our proposed approach indicate a larger change in values corresponding to the momentum factor, and this might result from the momentum factor being weak. Our specification tests support a four-factor model for Fama French portfolios, and observations show that the momentum factor might only serve as a weak proxy factor which can explain the difference between the HJ and the HJN specification test results and the differences in estimated parameter values.

	HJ(p-val)	HJN(p-val)	HJS(rejected)
n=25			
Three factor	0.000	0.000	No
Four factor	**0.000**	**0.0694**	No

Table 2: Tests of specification using monthly returns on 100 portfolios sorted by size and book-to-market and the three Fama-French factors and the momentum factor from 1977-08 to 2019-08

6 Conclusions

We show that the HJ statistic is not a valid model selection tool and model specification test statistic when weak (proxy) factors are present. We propose two novel approaches that provide size-correct model specification tests, alongside with which we also propose novel weak (proxy) factors robust risk premia estimators. Our empirical application supports a four factor structure for Fama French portfolios despite that the momentum factor is a weak proxy factor.

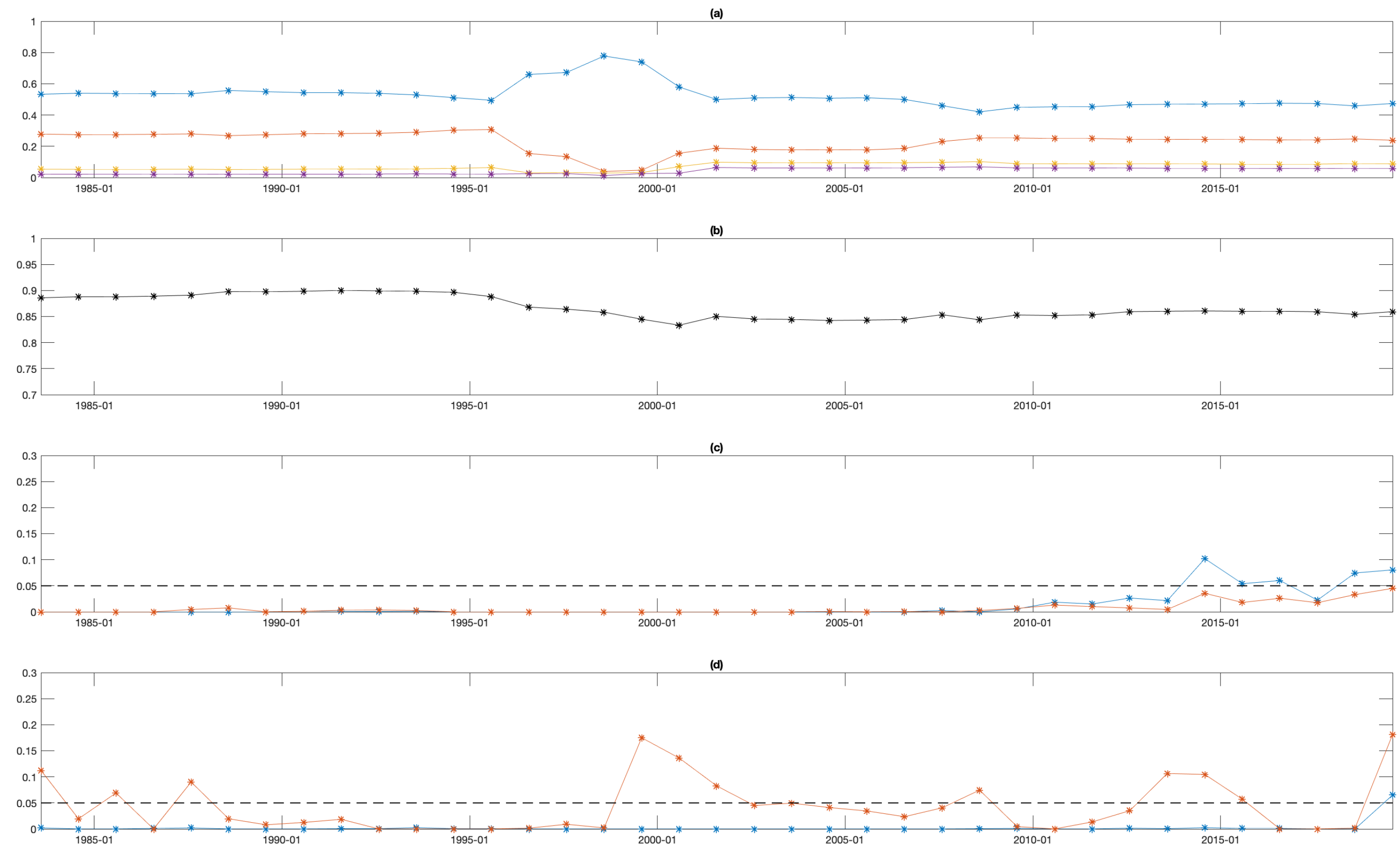

Figure 7: The time series of estimates from 1963-09 to 2019-08 with rolling windows of size T = 240 (20 years) and the number of increments between successive rolling windows being 12 (1 year) (x-axis is labeled with the ending period of each rolling window, for example, the first rolling window uses data from 1963-09 to 1983-08, so x-axis should label the first point with '1983-08'). (a) The fraction of the total variation of the portfolio returns that is explained by their four largest principal components respectively (CRRs) ($\lambda_i/(\sum_{j=1}^N \lambda_j), i = 1, \cdots, 4$); (b) The fraction of the total variation of the portfolio returns that is explained by the sum of the four largest principal components ($\sum_{i=1}^4 \lambda_i/(\sum_{j=1}^N \lambda_j)$); (c) p-values of the HJ specification test of the three-FF-factor model (blue), p-values of the HJ specification test of the four-factor (three FF and momentum factors) model (red);(d) p-values of the HJN specification test of the three-FF-factor model (blue), p-values of the HJN specification test of the four-factor model (red).

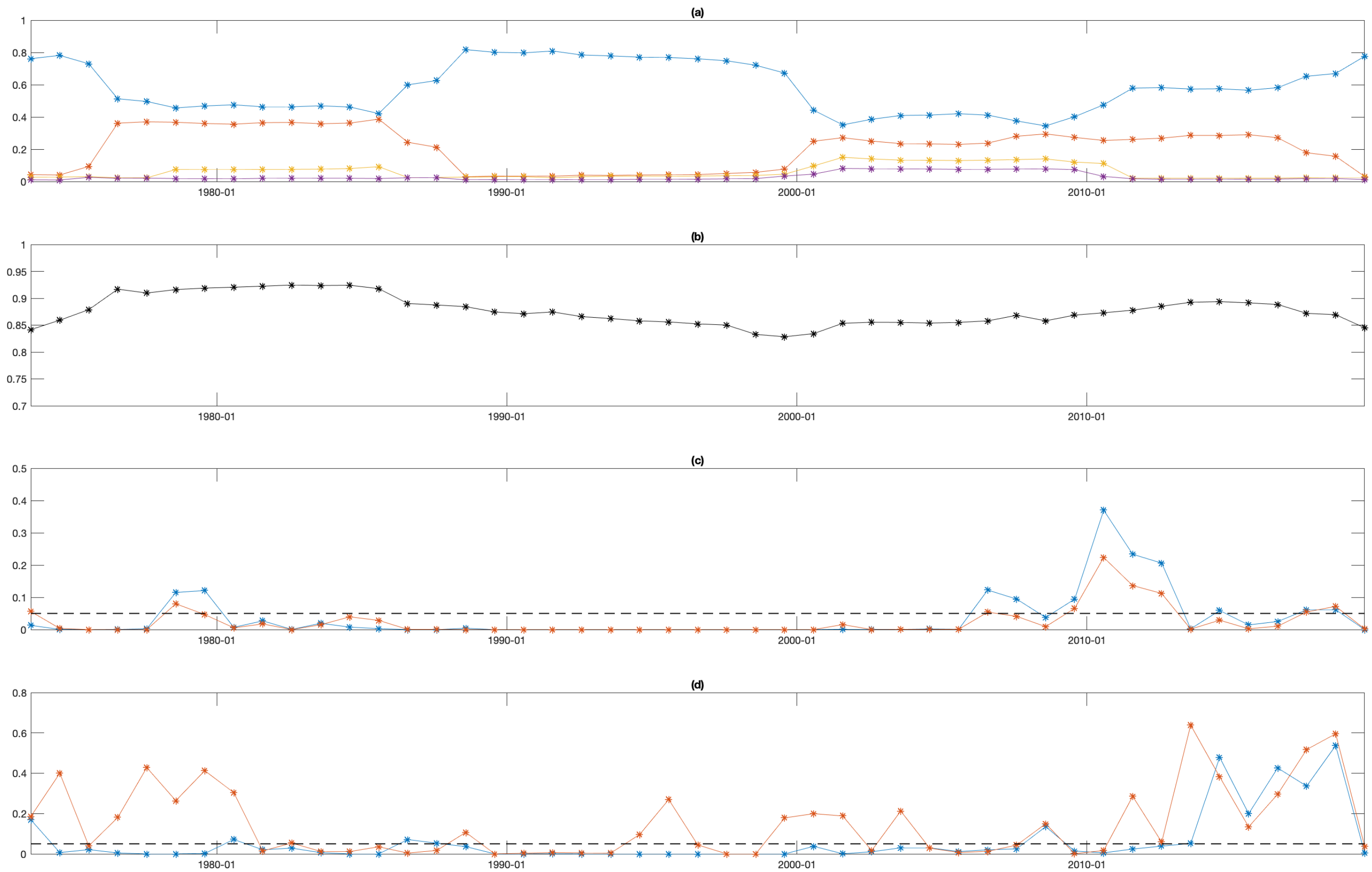

Figure 8: The time series of estimates from 1963-09 to 2019-08 with rolling windows of size T = 240 (20 years) and the number of increments between successive rolling windows being 12 (1 year) (x-axis is labeled with the ending period of each rolling window). (a) The fraction of the total variation of the portfolio returns that is explained by their four largest principal components respectively (CRRs) ($\lambda_i/(\sum_{j=1}^{N}\lambda_j), i = 1, \cdots, 4$); (b) The fraction of the total variation of the portfolio returns that is explained by the sum of the four largest principal components ($\sum_{i=1}^{4}\lambda_i/(\sum_{j=1}^{N}\lambda_j)$); (c) p-values of the HJ specification test of the three-FF-factor model (blue), p-values of the HJ specification test of the four-factor (three FF and momentum factors) model (red);(d) p-values of the HJN specification test of the three-FF-factor model (blue), p-values of the HJN specification test of the four-factor model (red).

References

Seung C Ahn and Christopher Gadarowski. Small sample properties of the gmm specification test based on the hansen–jagannathan distance. *Journal of Empirical Finance*, 11(1):109–132, 2004.

Stanislav Anatolyev and Anna Mikusheva. Factor models with many assets: strong factors, weak factors, and the two-pass procedure. *arXiv preprint arXiv:1807.04094*, 2018.

Jushan Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171, 2003.

Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

Svetlana Bryzgalova. Spurious factors in linear asset pricing models. *Working paper*, 2016.

John H Cochrane. *Asset pricing: Revised edition*. Princeton university press, 2009.

Eugene F Fama and Kenneth R French. Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56, 1993.

Eugene F Fama and James D MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636, 1973.

Yuguang Fang, Kenneth A Loparo, and Xiangbo Feng. Inequalities for the trace of matrix product. *IEEE Transactions on Automatic Control*, 39(12):2489–2490, 1994.

Stefano Giglio and Dacheng Xiu. Inference on risk premia in the presence of omitted factors. Technical report, National Bureau of Economic Research, 2017.

Nikolay Gospodinov, Raymond Kan, and Cesare Robotti. Spurious inference in reduced-rank asset-pricing models. *Econometrica*, 85(5):1613–1628, 2017.

Lars Peter Hansen and Ravi Jagannathan. Assessing specification errors in stochastic discount factor models. *The Journal of Finance*, 52(2):557–590, 1997.

Campbell R Harvey, Yan Liu, and Heqing Zhu. ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68, 2016.

Roger A Horn, Roger A Horn, and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2013.

Ravi Jagannathan and Zhenyu Wang. The conditional capm and the cross-section of expected returns. *The Journal of finance*, 51(1):3–53, 1996.

Raymond Kan and Cesare Robotti. Model comparison using the hansen-jagannathan distance. *The Review of Financial Studies*, 22(9):3449–3490, 2008.

Raymond Kan and Chu Zhang. Two-pass tests of asset pricing models with useless factors. *the Journal of Finance*, 54(1):203–235, 1999.

Raymond Kan and Guofu Zhou. *Hansen-Jagannathan distance: Geometry and exact distribution*. School of Economics and Finance, University of Hong Kong, 2004.

Frank Kleibergen. Testing parameters in gmm without assuming that they are identified. *Econometrica*, 73(4):1103–1123, 2005.

Frank Kleibergen. Tests of risk premia in linear factor models. *Journal of econometrics*, 149(2): 149–173, 2009.

Frank Kleibergen and Richard Paap. Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics*, 133(1):97–126, 2006.

Frank Kleibergen and Zhaoguo Zhan. Unexplained factors and their effects on second pass r-squared's. *Journal of Econometrics*, 189(1):101–116, 2015.

Frank Kleibergen and Zhaoguo Zhan. Identification-robust inference on risk premia of mimicking portfolios of non-traded factors. *Journal of Financial Econometrics*, 16(2):155–190, 2018.

Frank Kleibergen and Zhaoguo Zhan. Robust inference for consumption-based asset pricing. 2019.

Frank Kleibergen, Lingwei Kong, and Zhaoguo Zhan. Identification robust testing of risk premia in finite samples. *Working paper*, 2018.

Tim A Kroencke. Asset pricing without garbage. *The Journal of Finance*, 72(1):47–98, 2017.

Martin Lettau and Sydney Ludvigson. Consumption, aggregate wealth, and expected stock returns. *the Journal of Finance*, 56(3):815–849, 2001.

Anna Mikusheva. Robust confidence sets in the presence of weak instruments. *Journal of Econometrics*, 157(2):236–247, 2010.

Alexei Onatski. Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2):244–258, 2012.

Magda Peligrad, Sergey Utev, et al. Central limit theorem for stationary linear processes. *The Annals of Probability*, 34(4):1608–1622, 2006.

Jay Shanken. On the estimation of beta-pricing models. *The review of financial studies*, 5(1):1–33, 1992.

A Additional figures and tables

This section provides additional figures and tables. Mainly, we provide more results for our empirical application in Section 5, and further illustrate how the conventional tests may fail in some empirically relevant settings.

Figure 9 illustrates the density function of the HJS statistic under same settings as in Figure 1. We use our previous simulation experiment calibrated to data from Lettau and Ludvigson (2001) to illustrate properties of the HJS statistic. The black dotted curve in Figure 9 suggests the value of the HJS statistic could be small when the model is not correctly specified. These smaller values are due to much smaller confidence sets CS_r when model is not correctly specified. Therefore, using the HJS statistic as a model selection tool may not be a good idea even though it provides size correct specification test.

Figures 10 and 11 follow the same settings as in Figures 7 and 8, and they provide additional results by adding the p-values of the rank test (Kleibergen and Paap (2006)) of q_G (the rank of q_G reflects the identification strength of the model, e.g., Kleibergen and Zhan (2019), Kleibergen et al. (2018), as it shows whether the sample pricing errors vary enough as a function of θ), and the p-values of the $\mathcal{J}$ specification test.

Figures 10 and 11 show that the $\mathcal{J}$ specification tests tend to give larger p-values, and thus it is less informative. When we use 10-year window size (Figure 11) instead of the 20-year window size (Figure 10), the p-values of the rank test increase and the lack of identification strength also increases the p-values of the $\mathcal{J}$ test. For points near the coordinate '1990-01' in Figure 11, the rank test can not reject that the q_G is of reduced rank (lack of identification strength), and the corresponding $\mathcal{J}$-test p-values increase while the HJ tests tend to have larger p-values for testing the reduced three-factor model than the four-factor model. In short, in the presence of weak (proxy) factors, both well-know test statistics can not provide satisfying inference results.

Tables 4 and 3 follow the same settings as in Table 2 and reports additional results: tests with different value of n and parameter estimates. If n is getting too large, the HJN test also suffers from finite sample issues and tends to have smaller p-values, as its validity requires $n/N \to 0$. We leave the construction of a high-dimensional robust test statistic for further study.

	HJ(p-val)	HJN(p-val)	HJS(rejected)
n=30			
Three factor	0.000	0.000	No
Four factor	**0.000**	**0.054**	No
n=35			
Three factor	0.000	0.000	No
Four factor	**0.000**	**0.027**	No
n=40			
Three factor	0.000	0.000	No
Four factor	**0.000**	**0.026**	No

Table 3: Tests of specification using monthly returns on 100 portfolios sorted by size and book-to-market and the three Fama-French factors and the momentum factor from 1977-08 to 2019-08

	Market	SMB	HML	MOM
(1)				
$\widehat{\theta}_G$	0.0480	-0.0013	-0.0127	–
$\widetilde{\theta}_G$	0.3081	-0.1296	0.1370	–
$\widehat{\lambda}_g$	-5.2108	0.4037	-0.0347	–
$\widetilde{\lambda}_g$	-5.4740	0.3250	-0.2261	–
(2)				
$\widehat{\theta}_G$	0.0472	0.0174	-0.0172	0.0240
$\widetilde{\theta}_G$	0.1146	-0.4108	0.2233	-0.8945
$\widehat{\lambda}_g$	-3.7893	0.6822	-0.3995	1.5882
$\widetilde{\lambda}_g$	-2.8368	0.9104	-0.6567	3.6223

Table 4: Estimates with Fama-French factors and the momentum factor using monthly returns on 100 portfolios sorted by size and book-to-market from 1977-08 to 2019-08

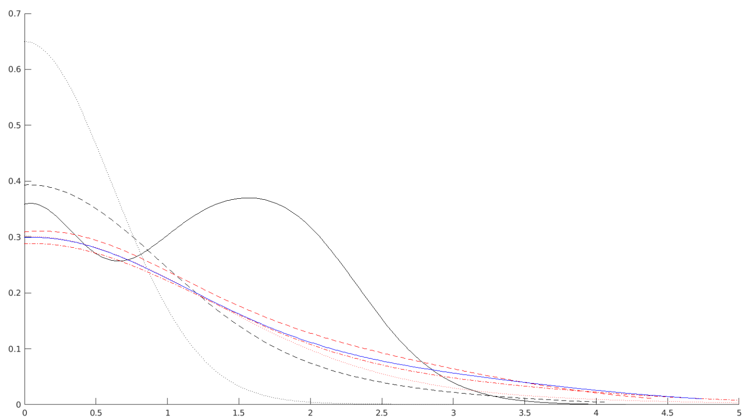

Figure 9: Densities of the HJS statistic: (1) black solid: three strong factors; (2) black dashed: two strong factors; (3) black dotted: one strong factors; (4) blue solid: three weak proxy factors; (5) red solid: three strong factors and one useless factor; (6) red dashed: two strong factors and one useless factor; (7) red dotted: one strong factors and one useless factor; (8) red dash-dotted:one strong factors and two useless factors

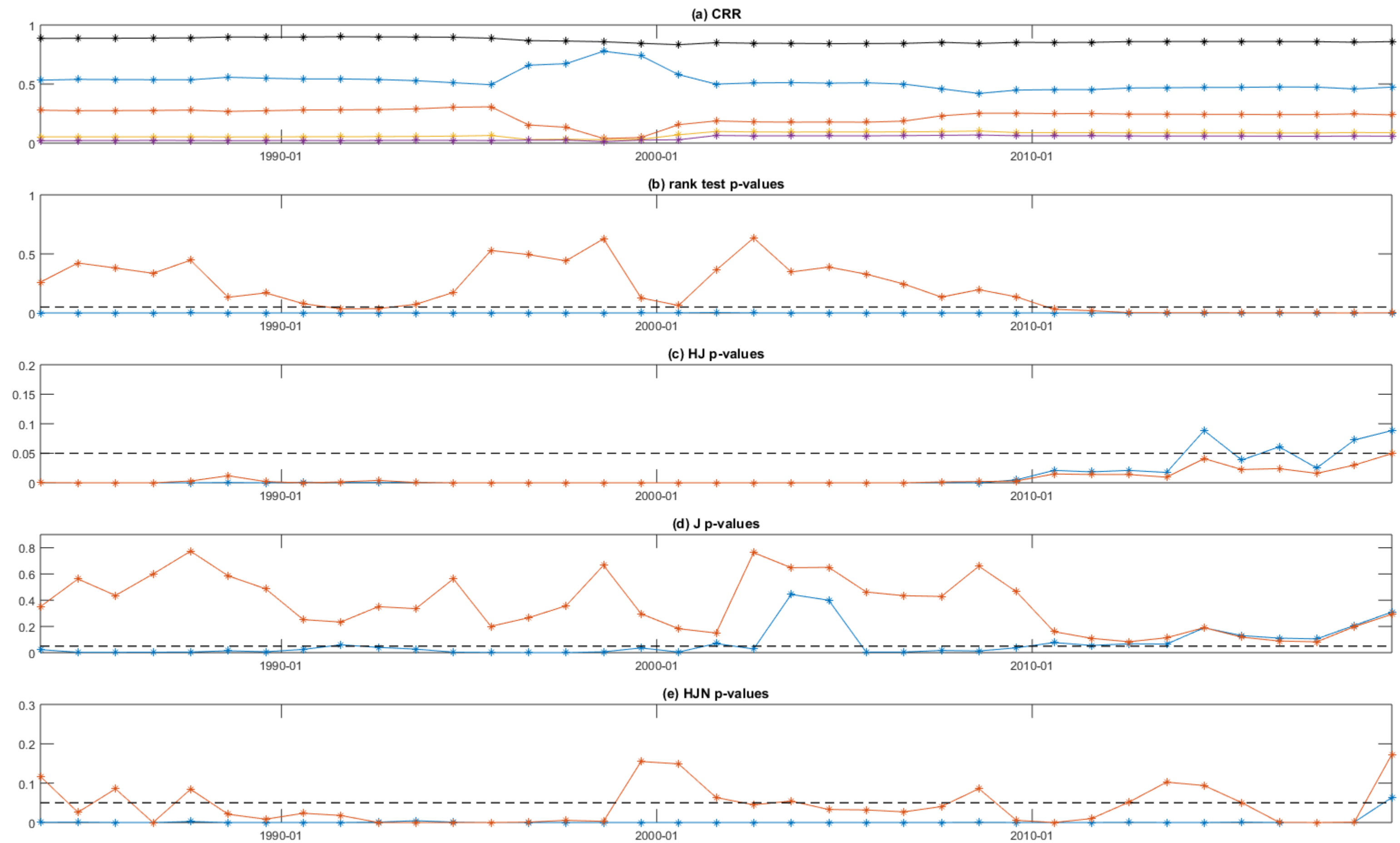

Figure 10: The time series of estimates from 1963-09 to 2019-08 with rolling windows of size T = 240 (20 years) and the number of increments between successive rolling windows being 12 (1 year) (x-axis is labeled with the ending period of each rolling window, for example, the first rolling window uses data from 1963-09 to 1983-08, so x-axis should label the first point with '1983-08'). (a) The fraction of the total variation of the portfolio returns that is explained by their four largest principal components respectively (CRRs) $(\lambda_i/(\sum_{j=1}^{N}\lambda_j), i = 1, \cdots, 4)$; the black curve is the fraction of the total variation of the portfolio returns that is explained by the sum of the four largest principal components (sum of the first four CRRs, $\sum_{i=1}^{4}\lambda_i/(\sum_{j=1}^{N}\lambda_j)$); (b) p-values of the rank test (Kleibergen and Paap (2006)) of q_G using the estimator $q_{G,T}$ with G constructed by the three FF factors (blue), and the four factors (three FF and momentum factors) (red) respectively; (c) p-values of the HJ specification test of the three-FF-factor model (blue), p-values of the HJ specification test of the four-factor (three FF and momentum factors) model (red); (d) p-values of the $\mathcal{J}$ specification test of the three-FF-factor model (blue), p-values of the HJN specification test of the four-factor (three FF and momentum factors) model (red); (e) p-values of the HJN specification test of the three-FF-factor model (blue), p-values of the HJN specification test of the four-factor (three FF and momentum factors) model (red).

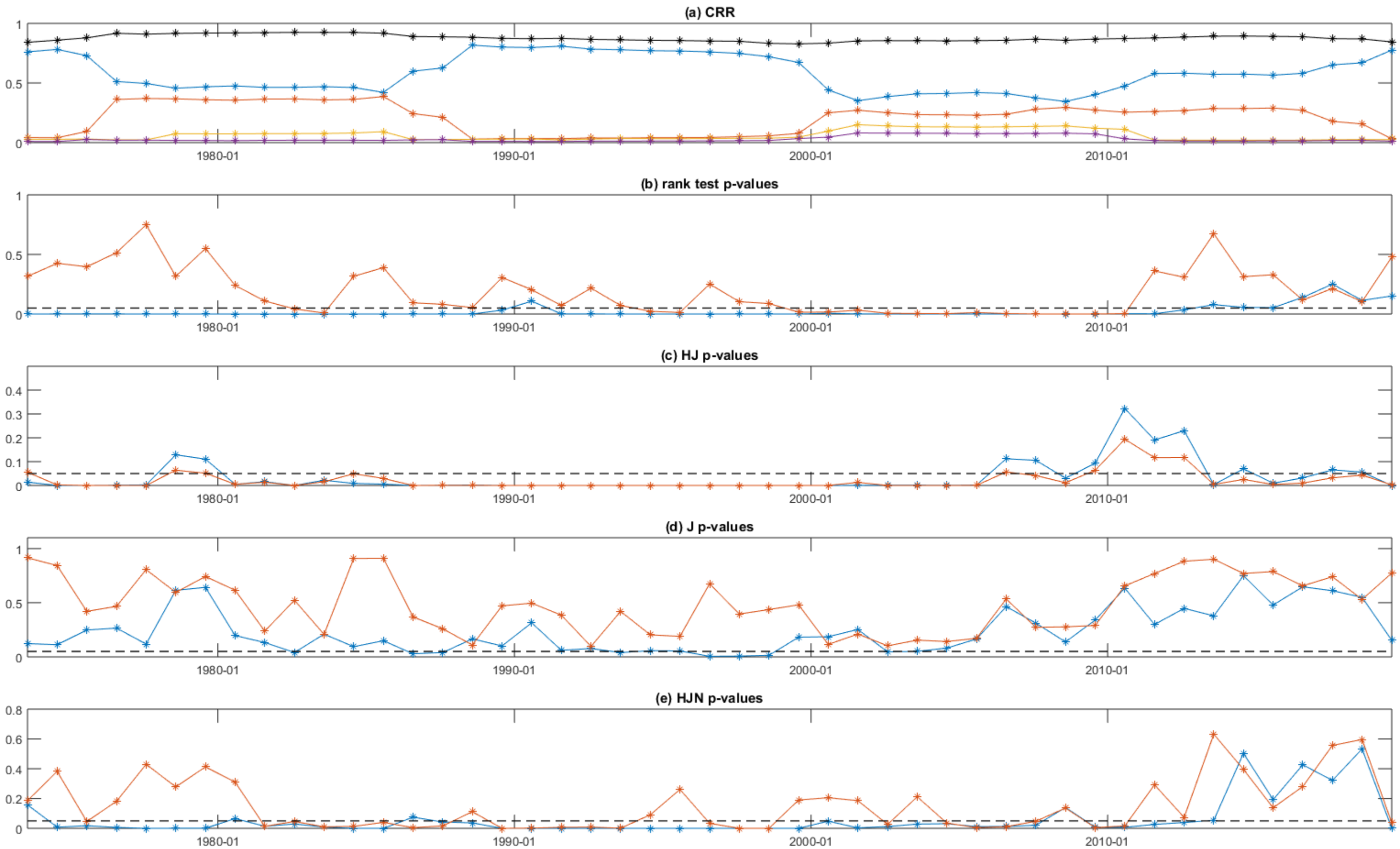

Figure 11: The time series of estimates from 1963-09 to 2019-08 with rolling windows of size T = 240 (20 years) and the number of increments between successive rolling windows being 12 (1 year) (x-axis is labeled with the ending period of each rolling window). (a) The fraction of the total variation of the portfolio returns that is explained by their four largest principal components respectively (CRRs) $(\lambda_i/(\sum_{j=1}^N \lambda_j), i = 1, \cdots, 4)$; the black curve is the fraction of the total variation of the portfolio returns that is explained by the sum of the four largest principal components (sum of the first four CRRs, $\sum_{i=1}^4 \lambda_i/(\sum_{j=1}^N \lambda_j)$); (b) p-values of the rank test (Kleibergen and Paap (2006)) of q_G using the estimator $q_{G,T}$ with G constructed by the three FF factors (blue), and the four factors (three FF and momentum factors) (red) respectively; (c) p-values of the HJ specification test of the three-FF-factor model (blue), p-values of the HJ specification test of the four-factor (three FF and momentum factors) model (red); (d) p-values of the $\mathcal{J}$ specification test of the three-FF-factor model (blue), p-values of the HJN specification test of the four-factor (three FF and momentum factors) model (red); (e) p-values of the HJN specification test of the three-FF-factor model (blue), p-values of the HJN specification test of the four-factor (three FF and momentum factors) model (red).

B Proofs related to sections 2 and 3

In appendix, we use $\lambda_j(A)$ to denote the jth largest eigenvalue of a given matrix A, $\lambda_{\min}(A), \lambda_{\max}(A)$ the minimum and the maximum eigenvalues. With $A = (a_{ij})$, multiple matrix norms are denoted as $\|A\| = \sqrt{\lambda_{\max}(A'A)}$, $\|A\|_1 = \max_j \sum_i |a_{ij}|$, $\|A\|_\infty = \max_i \sum_j |a_{ij}|$, $\|A\|_F = \sqrt{tr(A'A)}$, $\|A\|_{\max} = \max_{i,j} |a_{ij}|$.

Assumption B.1. *The following asymptotic distributions hold jointly:* $(\xi_{g,T}, \xi_{gg,T}, \xi_{vG,T}, \xi_{zG,T}, \xi_{eG,T},) \to_d (\xi_g, \xi_V, \xi_{vG}, \xi_{zG}, \xi_{eG})$, *where* $\xi_{g,T} = \frac{1}{\sqrt{T}}\left(\sum_{t=1}^T g_t - \mu_g\right)$, $\xi_{gg,T} = \frac{1}{\sqrt{T}}\sum_{t=1}^T \bar{g}_t\bar{g}_t'$, $\xi_{vG,T} = \frac{1}{\sqrt{T}}\sum_{t=1}^T v_t G_t'$,$\xi_{zG,T} = \frac{1}{\sqrt{T}}\sum_{t=1}^T z_t G_t'$, $\xi_{eG,T} = \frac{1}{\sqrt{T}}\sum_{t=1}^T e_t G_t'$, *and* ξ_g, $vec(\xi_{vG}), vec(\xi_{zG})$, $vec(\xi_{eG})$ *are zero-mean normal random vectors,* ξ_V *is a Gaussian random matrix.*

Assumption B.1 is a central limit theorem for the different components in equation (4) interacted with a constant and the proxy factors. Some of the statements such as $\xi_{g,T} \to_d \xi_g$ would hold if proxy factors are stationary with finite fourth moments and satisfy some strong mixing conditions (see e.g. Peligrad et al. (2006)). We specify a relatively strong assumption here instead of dealing with heavy technical details, but our results can be extended to general cases.

Proof of Lemma 2.1. Assumption 2.1 implies $r_t = c + \beta_g\bar{g}_t + \beta_g(\bar{g} - \mu_g) + \beta v_t + u_t$ then

$$
\begin{aligned}
\widehat{B}_g &= \sum_{i=1}^T r_i G_i' \left(\sum_{t=1}^T G_t G_t'\right)^{-1} = \sum_{i=1}^T \left(c + \beta_g\bar{g}_t + \beta_g(\bar{g} - \mu_g) + \beta v_t + u_t\right) G_i' \left(\sum_{t=1}^T G_t G_t'\right)^{-1} \\
&= (c, \beta_g) + \frac{1}{\sqrt{T}}(\beta_g \xi_{g,T}, 0) + \frac{1}{\sqrt{T}}(\beta, \gamma)\,\xi_{vzG,T}\widehat{Q}_G^{-1} + \frac{1}{\sqrt{T}}\xi_{eG,T}\widehat{Q}_G^{-1}
\end{aligned}
$$

The conclusion is then a direct result from Assumption B.1. □

Proof of Theorem 2.2. Assumption 2.1 and B.1 imply that

$$\widehat{Q}_r = Q_r + O_p(1/\sqrt{T}) \tag{16}$$

$$\iota_N = B_g Q_G \theta_G \tag{17}$$

Rewrite $T\widehat{\delta}_g^2$ as

$$\sqrt{T}\left(\iota_N-\widehat{B}_gQ_G\theta_G\right)'W\sqrt{T}\left(\iota_N-\widehat{B}_gQ_G\theta_G\right) \tag{18}$$

where

$$\begin{aligned}W&=(\widehat{Q}_r^{-1}-\widehat{Q}_r^{-1}\widehat{B}_gQ_{B_g,T}(Q_{B_g,T}\widehat{B}_g'\widehat{Q}_r^{-1}\widehat{B}_gQ_{B_g,T})^{-1}Q_{B_g,T}\widehat{B}_g'\widehat{Q}_r^{-1})\\Q_{B_g,T}&=\operatorname{diag}\left(I_{1+K_{g,1}},\sqrt{T}I_{K_{g,2}}\right)\end{aligned}$$

Lemma 2.1 and equation (17) imply that

$$\sqrt{T}\left(\iota_N-\widehat{B}_gQ_G\theta_G\right)\rightarrow_d\widetilde{\psi}_{B_g} \tag{19}$$

Lemma 2.1 and Assumption 2.3 imply the following equation holds jointly with equation (19)

$$\widehat{B}_gQ_{B_g,T}\rightarrow_d\eta_{B_g}+(0\vdots\psi_{\beta_{g,2}}) \tag{20}$$

Plug equations (16)(19)(20) in equation (18), then we would derive the conclusion. □

Lemma B.1. *Suppose Assumption B.1 and 2.1-2.3 hold, let N,T increase and then the restrictions on e_t from Assumption 2.2 (i)(ii)(iii) also hold for $\widetilde{e}_t$ with $\widetilde{e}_t=Q_r^{-\frac{1}{2}}e_t$ with $Q_r^{-\frac{1}{2}}=A\Lambda^{-\frac{1}{2}}A'$ such that $Q_r=A\Lambda A'$ with $A'A=I_N$ and we assume $Q_r^{-\frac{1}{2}}$ is a row diagonally-dominant matrix[5].*

Proof. $\widetilde{e}_t,t=1,\cdots,T$ are i.i.d. mean zero random vectors by construction. with finite fourth moments by construction. Next we show $\sup_i\mathbb{E}\widetilde{e}_{it}^4$ is bounded. Assumption 2.1 implies $Q_r=c'c+\beta V_f\beta'+\gamma V_z\gamma'+\Omega_e$ and thus we have the following results by eigenvalue inequalities (see e.g.

[5]In the proof of this lemma we need that $Q_r^{-\frac{1}{2}}$ has bounded absolute row sum. This is not a wild assumption if we consider the Gershgoring-type eigenvalue inclusion theorem and all eigenvalues of $Q_r^{-1/2}$ are bounded.

7.3.P16 of Horn et al. (2013)):

$$\lambda_{\max}\left(Q_r\right)=\lambda_{\max}\left(c'c+\beta V_f\beta'+\gamma V_z\gamma'+\Omega_e\right)\leq\lambda_{\max}\left(c'c+\beta V_f\beta'+\gamma V_z\gamma'\right)+L \tag{21}$$

$$\lambda_{\min}\left(Q_r\right)=\lambda_{\min}\left(c'c+\beta V_f\beta'+\gamma V_z\gamma'+\Omega_e\right)\geq\lambda_{\min}\left(c'c+\beta V_f\beta'+\gamma V_z\gamma'\right)+l=l \tag{22}$$

Then by the assumption that $Q_r^{-\frac{1}{2}}$ is a row diagonally-dominant matrix we know any row sums of $Q_r^{-\frac{1}{2}}$ would be upper bounded by $2l^{-1/2}$ and thus $\sup_i \mathbb{E}e_{it}^4 \leq L$ implies that $\sup_i \mathbb{E}\widetilde{e}_{it}^4$ is bounded. Therefore, Assumption 2.2 (i) holds.

The term $c'c+\beta V_f\beta'+\gamma V_z\gamma'$ in Q_r is a positive semi-definite matrix and we can rewrite that term as $A_{Q_r}\Lambda_{Q_r}A'_{Q_r}$ such that Λ_{Q_r} is a diagonal matrix containing all positive eigenvalues of $c'c+\beta V_f\beta'+\gamma V_z\gamma'$ and A_{Q_r} are the corresponding eigenvectors. Therefore, $Q_r=A_{Q_r}\Lambda_{Q_r}A'_{Q_r}+\Omega_e$ and thus

$$Q_r^{-1}=\Omega_e^{-1}-\Omega_e^{-1}A_{Q_r}\left(\Lambda_{Q_r}^{-1}+A'_{Q_r}\Omega_e^{-1}A_{Q_r}\right)^{-1}A'_{Q_r}\Omega_e^{-1} \tag{23}$$

which then implies that

$$\operatorname{tr}\left(\mathbb{E}Q_r^{-1/2}e_te_t'Q_r^{-1/2}\right)/N=\operatorname{tr}\left(\Omega_eQ_r^{-1}\right)/N=1-\operatorname{tr}\left(\left(\Lambda_{Q_r}^{-1}+A'_{Q_r}\Omega_e^{-1}A_{Q_r}\right)^{-1}\right)/N$$

From the fact that eigenvalues of Λ_{Q_r} explode by Assumption 2.3 and eigenvalues of Ω_e are bounded by Assumption 2.2 , Courant-Fischer minimax principle implies $\operatorname{tr}\left(\left(\Lambda_{Q_r}^{-1}+A'_{Q_r}\Omega_e^{-1}A_{Q_r}\right)^{-1}\right)/N\to 0$. Furthermore, Assumption 2.2 and 2.3 imply that $\liminf_{N,T}\lambda_{\min}\left(Q_r^{-1/2}\right)$ and $\limsup_{N,T}\lambda_{\min}\left(Q_r^{-1/2}\right)$ are bounded, and thus we also have

$$0<l<\liminf_{N,T}\lambda_{\min}\left(Q_r^{-1/2}\Omega_eQ_r^{-1/2}\right)<\limsup_{N,T}\lambda_{\max}\left(Q_r^{-1/2}\Omega_e\right)Q_r^{-1/2}<L<\infty.$$

Therefore, Assumption 2.2 (ii) holds for $\widetilde{e}_t$. As for Assumption 2.2 (iii), inequalities for the trace of matrix product (e.g. Fang et al. (1994)) suggest

$$\begin{aligned}&\lambda_{\min}\left(Q_r^{-1}\right)\operatorname{tr}\left(e_te_t'\right)-\lambda_{\max}\left(Q_r^{-1}\right)\operatorname{tr}\left(\mathbb{E}\left(e_te_t'\right)\right)\\ \leq&\operatorname{tr}\left(Q_r^{-1}\left(e_te_t'-\mathbb{E}\left(e_te_t'\right)\right)\right)\leq\lambda_{\max}\left(Q_r^{-1}\right)\operatorname{tr}\left(e_te_t'\right)-\lambda_{\min}\left(Q_r^{-1}\right)\operatorname{tr}\left(\mathbb{E}\left(e_te_t'\right)\right)\end{aligned}$$

which implies that $\mathbb{E}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\widetilde{e}_{it}\widetilde{e}_{is}-\mathbb{E}\widetilde{e}_{it}\widetilde{e}_{is}\right)\right|^4<L<\infty$ given all eigenvalues of Q_r^{-1} are bounded. □

Theorem B.2 illustrates how the weak (proxy) factors can affect the asymptotic properties of the estimator $\widehat{\theta}_G$. This theorem resembles theorem one in Anatolyev and Mikusheva (2018), and indeed we observe that the asymptotic behavior of the estimator $\widehat{\theta}_G$ is similar to the one of the two-pass FM risk premia in Anatolyev and Mikusheva (2018).

Theorem B.2.

Case (1): Suppose Assumption B.1 and 2.1-2.3 hold, N is fixed and T increases to infinity:

$$\sqrt{T}Q_{B_g,T}^{-1}\left(\widehat{Q}_G(\widehat{\theta}_G-\theta_G)-bias_e-bias_m\right)=O_p(1)$$

where

$$\begin{aligned}&bias_e=-\left(\widehat{B}_g'\widehat{Q}_r^{-1}\widehat{B}_g\right)^{-1}\widehat{B}_{g,e}'\widehat{Q}_r^{-1}\widehat{B}_{g,e}\widehat{Q}_G\theta_G\\&bias_m=-\left(\widehat{B}_g'\widehat{Q}_r^{-1}\widehat{B}_g\right)^{-1}\left(B_g+\widehat{B}_{g,m}\right)'\widehat{Q}_r^{-1}\widehat{B}_{g,m}\widehat{Q}_G\theta_G\\&\widehat{B}_{g,e}=\frac{1}{\sqrt{T}}\xi_{eG,T}\widehat{Q}_G^{-1};\ \widehat{B}_{g,m}=\frac{1}{\sqrt{T}}\left(\beta,\gamma\right)\xi_{vzG,T}\widehat{Q}_G^{-1};\ Q_{B_g,T}=diag\left(I_{1+K_{g,1}},\sqrt{T}I_{K_{g,2}}\right)\end{aligned}$$

and

$$\sqrt{T}Q_{B_g,T}^{-1}bias_e = \begin{pmatrix} \sqrt{T}bias_{e,1} \\ bias_{e,2} \end{pmatrix} \to_d WQ_{B_g}\widetilde{e}'Q_r^{-1}\widetilde{e}Q_G\theta_G$$

$$\sqrt{T}Q_{B_g,T}^{-1}bias_m = \begin{pmatrix} \sqrt{T}bias_{m,1} \\ bias_{m,2} \end{pmatrix} \to_d W\left(\begin{pmatrix} I_{1+K_g} \\ 0 \end{pmatrix} + \widetilde{\xi}Q_{B_g}\right)'\eta'Q_r^{-1}\eta\widetilde{\xi}Q_G\theta_G$$

$$W = -\left(\left(\eta\left(\begin{pmatrix} I_{1+K_g} \\ 0 \end{pmatrix} + \widetilde{\xi}Q_{B_g}\right) + \widetilde{e}Q_{B_g}\right)'Q_r^{-1}\left(\eta\left(\begin{pmatrix} I_{1+K_g} \\ 0 \end{pmatrix} + \widetilde{\xi}Q_{B_g}\right) + \widetilde{e}Q_{B_g}\right)\right)^{-1}$$

Case (2): Suppose Assumption B.1 and 2.1-2.3 hold, let N, T increases to infinity $N/T \to c$ and Q_r is a known row diagonally-dominant matrix such that $\eta'Q_r^{-1}\eta/N \to \widetilde{D}$:

$$\sqrt{NT}Q_{B_g,T}^{-1}\left(\widehat{Q}_G(\widehat{\theta}_G - \theta_{G,T}) - bias_e - bias_m\right) = O_p(1)$$

where

$$\theta_{G,T} = \theta_G + \widehat{Q}_G^{-1}\left(\widehat{B}_g'Q_r^{-1}\widehat{B}_g\right)^{-1}\widehat{B}_g'Q_r^{-1}\left(q_G\left(\widehat{Q}_G^{-1} - Q_G^{-1}\right) - \frac{1}{\sqrt{T}}(\beta_g\xi_{g,T}, 0)\right)\widehat{Q}_G\theta_G$$

$$\sqrt{T}Q_{B_g,T}^{-1}\widehat{Q}_G(\theta_{G,T} - \theta_G) = O_p(1)$$

$$\sqrt{T}Q_{B_g,T}^{-1}bias_e = \begin{pmatrix} \sqrt{T}bias_{e,1} \\ bias_{e,2} \end{pmatrix} \to_d WQ_{B_g}\widetilde{\Sigma}_eQ_G\theta_G$$

$$\sqrt{T}Q_{B_g,T}^{-1}bias_m = \begin{pmatrix} \sqrt{T}bias_{m,1} \\ bias_{m,2} \end{pmatrix} \to_d W\left(\begin{pmatrix} I_{1+K_g} \\ 0 \end{pmatrix} + \widetilde{\xi}Q_{B_g}\right)'\widetilde{D}\widetilde{\xi}Q_G\theta_G$$

$$W = -\left(\left(\begin{pmatrix} I_{1+K_g} \\ 0 \end{pmatrix} + \widetilde{\xi}Q_{B_g}\right)'\widetilde{D}\left(\begin{pmatrix} I_{1+K_g} \\ 0 \end{pmatrix} + \widetilde{\xi}Q_{B_g}\right) + Q_{B_g}\widetilde{\Sigma}_eQ_{B_g}\right)^{-1}$$

Proof. Case (1): After simple algebra, we can express the term $\widehat{Q}_G\left(\widehat{\theta}_G - \theta\right) - bias_e - bias_m$ in the following way

$$\widehat{Q}_G\left(\widehat{\theta}_G-\theta\right)-bias_e-bias_m=\left(\widehat{B}_g'\widehat{Q}_r^{-1}\widehat{B}_g\right)^{-1}\left\{\widehat{B}_g'\widehat{Q}_r^{-1}\left(q_G\left(\widehat{Q}_G^{-1}-Q_G^{-1}\right)-\frac{1}{\sqrt{T}}\left(\beta_g\xi_{g,T},0\right)\right)\right.$$
$$\left.-\left(\frac{1}{\sqrt{T}}\left(\beta_g\xi_{g,T},0\right)+\widehat{B}_{g,e}\right)'\widehat{Q}_r^{-1}\widehat{B}_{g,m}-\left(B_g+\frac{1}{\sqrt{T}}\left(\beta_g\xi_{g,T},0\right)+\widehat{B}_{g,m}\right)'\widehat{Q}_r^{-1}\widehat{B}_{g,e}\right\}\widehat{Q}_G\theta_G \tag{24}$$

From the proofs of Lemma 2.1 and Theorem 2.2 we have

$$\widehat{B}_g=B_g+\frac{1}{\sqrt{T}}\left(\beta_g\xi_{g,T},0\right)+\widehat{B}_{g,m}+\widehat{B}_{g,e} \tag{25}$$
$$\widehat{B}_gQ_{B_g,T}=\left(c+O_p(1/\sqrt{T}),\eta_{\beta_g}\right)+\left(\frac{1}{\sqrt{T}}\widetilde{e}_T+\frac{1}{\sqrt{T}}\eta\widetilde{\xi}_T\right)Q_{B_g,T} \tag{26}$$

where $\widetilde{\xi}_T=\sqrt{T}\widehat{B}_{g,m}=\begin{pmatrix}0 & 0\\ \left(d_gQ_{\beta_g}\right)^{-1} & 0\\ 0 & I_{K_z}\end{pmatrix}\xi_{vzG,T}\widehat{Q}_G^{-1}$, $\widetilde{e}_T=\sqrt{T}\widehat{B}_{g,e}=\xi_{eG,T}\widehat{Q}_G^{-1}$,$\widetilde{\xi}_T\to_d\widetilde{\xi},\widetilde{e}_T\to_d\widetilde{e},Q_{B_g,T}/\sqrt{T}\to Q_{B_g}=\mathrm{diag}\left(0_{1+K_{g,1}},I_{K_{g,2}}\right)$ and $\widetilde{\xi},\widetilde{e}$ are Gaussian random matrices. Equation (26) implies that

$$\widehat{B}_gQ_{B_g,T}\to_d\ \eta\left(\begin{pmatrix}I_{1+K_g}\\ 0\end{pmatrix}+\widetilde{\xi}Q_{B_g}\right)+\widetilde{e}Q_{B_g} \tag{27}$$

With the above intermediate results, we prove our statement for the term $bias_e$ and the rests follow the same steps. Assumption B.1 implies that the above asymptotic results hold jointly and thus if we plug these into the equation below we would derive the asymptotic distribution of the term $bias_e$

$$\sqrt{T}Q_{B_g,T}^{-1}bias_e=-\left((\widehat{B}_gQ_{B_g,T})'\widehat{Q}_r^{-1}(\widehat{B}_gQ_{B_g,T})\right)^{-1}(\widehat{B}_{g,e}Q_{B_g,T})'\widehat{Q}_r^{-1}(\sqrt{T}\widehat{B}_{g,e})\widehat{Q}_G\theta_G \tag{28}$$

Case (2): Next we discuss the case where N,T both grows to infinity. We first show the following two results:

$$(2.1)\ \ T\widetilde{e}_T'Q_r^{-1}\widetilde{e}_T/N \to_p \tilde{\Sigma}_e \tag{29}$$

$$(2.2)\ \ \sqrt{T}\widetilde{e}_T'Q_r^{-1}\eta/N \to_p 0 \tag{30}$$

To prove the statement (2.1), let $\widetilde{e}_t = Q_r^{-\frac{1}{2}}e_t$ and $\rho(s,t) = \frac{1}{\sqrt{N}}\sum_{i=1}^N \widetilde{e}_{it}'\widetilde{e}_{is}$. By construction, $\mathbb{E}\rho^2(s,t) = \frac{1}{N}\text{tr}\left(\Omega_{\widetilde{e}}\Omega_{\widetilde{e}}\right) \leq \lambda_{\max}^2(\Omega_{\widetilde{e}})$ with $\Omega_{\widetilde{e}} = \mathbb{E}\widetilde{e}_t\widetilde{e}_t'$ and thus Lemma B.1 implies that $\mathbb{E}\rho^2(s,t)$ is bounded. Assumption 2.1 (finite fourth moments of proxy factors), Assumption 2.2 (i) and the bounded $\mathbb{E}\rho^2(s,t)$ deliver the following inequality

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{T\sqrt{N}}\sum_{i=1}^N\left(\sum_{t=1}^T\sum_{s\neq t}^T G_{tm}G_{sn}\widetilde{e}_{it}'\widetilde{e}_{is}\right)\right)^2\\
=&\mathbb{E}\left(\frac{1}{T}\left(\sum_{t=1}^T\sum_{s\neq t}^T G_{tm}G_{sn}\rho(s,t)\right)\right)\left(\frac{1}{T}\left(\sum_{t'=1}^T\sum_{s'\neq t}^T G_{t'm}G_{s'n}\rho(s',t')\right)\right)\\
=&\mathbb{E}\left(\frac{1}{T^2}\left(\sum_{t=1}^T\sum_{s\neq t}^T\left(G_{tm}^2G_{sn}^2 + G_{tm}G_{sn}G_{sm}G_{tn}\right)\rho^2(s,t)\right)\right) < L
\end{aligned} \tag{31}$$

which via Chebyshev's inequality gives

$$\frac{1}{T\sqrt{N}}\sum_{i=1}^N\left(\sum_{t=1}^T\sum_{s\neq t}^T G_tG_s'\widetilde{e}_{it}'\widetilde{e}_{is}\right) = O_P(1) \tag{32}$$

Finally we arrive at (2.1):

$$T\widetilde{e}_T'Q_r\widetilde{e}_T/N = \widehat{Q}_g^{-1}\left(\frac{1}{NT}\sum_{t=1}^T G_t\widetilde{e}_t'\widetilde{e}_tG_t'\right)\widehat{Q}_g^{-1} + O_p(1/\sqrt{N}) = \tilde{\Sigma}_e + O_p(1/\sqrt{N}) \tag{33}$$

where the first equality is due to equation (32) and the last equality is guaranteed by Lemma B.1.

Next, we prove the statement (2.2). We first look at the second moments

$$\mathbb{E}\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}G_te_t'Q_r^{-1}\eta/\sqrt{N}\right\|^2\right)\leq\mathbb{E}\left(\operatorname{tr}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}G_t\widetilde{e}_t'Q_r^{-1/2}\eta/\sqrt{N}\right)\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}G_t\widetilde{e}_t'Q_r^{-1/2}\eta/\sqrt{N}\right)'\right)$$

$$=\frac{1}{T}\sum_{t=1}^{T}\operatorname{tr}\left(\mathbb{E}\left(G_t\widetilde{e}_t'Q_r^{-1/2}\eta\eta'Q_r^{-1/2}\widetilde{e}_tG_t'\right)/N\right)\leq L\operatorname{tr}\left(\Omega_{\widetilde{e}}Q_r^{-1/2}\eta\eta'Q_r^{-1/2}/N\right)\leq L\lambda_{\max}\left(\Omega_{\widetilde{e}}\right)\left\|Q_r^{-1/2}\right\|_1^2\left(\|\eta\|_F^2/N\right)$$

where the first inequality is because factors g_t have finite fourth moments. The proof of Lemma B.1 shows that $\lambda_{\max}\left(\Omega_{\widetilde{e}}\right)$ and $\left\|Q_r^{-1/2}\right\|_1^2$ are bounded, Assumption 2.3 implies that $\|\eta\|_F^2/N$ is bounded. Therefore, we know

$$\mathbb{E}\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}G_te_t'Q_r^{-1}\eta/\sqrt{N}\right\|^2\right)\leq L<\infty \tag{34}$$

and thus (2.2) holds since $\sqrt{T}\widetilde{e}_T'Q_r^{-1}\eta/N=\frac{1}{\sqrt{N}}Q_g^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}G_te_t'Q_r^{-1}\eta/\sqrt{N}\right)=O_p(1/\sqrt{N})$. In the end, $\eta'Q_r^{-1}\eta/N\to\widetilde{D}$ and the result (2.2) imply the following term is of order $O_p(1)$ when $N/T\to c$:

$$\frac{\sqrt{T}}{\sqrt{N}}Q_{B_g,T}\left\{-\left(\frac{1}{\sqrt{T}}\left(\beta_g\xi_{g,T},0\right)+\frac{1}{\sqrt{T}}\widetilde{e}_T\right)'Q_r^{-1}\frac{1}{\sqrt{T}}\widetilde{\xi}_T-\left(B_g+\frac{1}{\sqrt{T}}\left(\beta_g\xi_{g,T},0\right)+\frac{1}{\sqrt{T}}\widetilde{\xi}_T\right)'Q_r^{-1}\frac{1}{\sqrt{T}}\widetilde{e}_T\right\}=O_p(1)$$

□

Assumption B.2. *Let $e_{g,t}(\theta_G)=\iota_N-r_tG_t'\theta_G$, and the restrictions on e_t from Assumption 2.2 (i)(ii)(iii) also hold for $e_{g,t}(\theta_G)$.*

Proof of Corollary 2.2.2. Theorem 2.2 suggest for given N, $T\widehat{\delta}_g^2\to_d d_{1,N}$ with $d_{1,N}=\widetilde{\psi}_{B_g}'M_{Q_r^{-\frac{1}{2}}\left(\eta_{B_g}+(0\vdots\psi_{\beta_{g,2}})\right)}\widetilde{\psi}_{B_g}$, $\widetilde{\psi}_{B_g}\sim N(0,S_{B_g})$ and $S_{B_g}=p\lim S_{B_g,T}$, $S_{B_g,T}=\frac{1}{T}\sum_{t=1}^{T}e_{g,t}(\theta_G)e_{g,t}(\theta_G)'$. From the construction of the $\widehat{c}_{1-\alpha}$, we know it is drawn from the distribution of the $d_{2,N}=\psi_S'M_{Q_r^{-\frac{1}{2}}\left(\eta_{B_g}+(0\vdots\psi_{\beta_{g,2}})\right)}\psi_S$ with $\psi_S\sim N(0,S)$, $S=p\lim\widehat{S}$, $\widehat{S}=\frac{1}{T}\sum_{t=1}^{T}e_{g,t}(\widehat{\theta}_G)e_{g,t}(\widehat{\theta}_G)'$, $\widehat{e}_{g,t}=\iota_N-r_tG_t'\widehat{\theta}_G$ and ψ_S independent from ψ_{B_g}. The proof of Theorem B.2 suggest that $\widetilde{\psi}_{B_g}'P_{Q_r^{-\frac{1}{2}}\left(\eta_{B_g}+(0\vdots\psi_{\beta_{g,2}})\right)}\widetilde{\psi}_{B_g}=O_p(1)$ for any given N, and the same for the term $\psi_S'P_{Q_r^{-\frac{1}{2}}\left(\eta_{B_g}+(0\vdots\psi_{\beta_{g,2}})\right)}\psi_S$. Now we look at the difference

$S_{B_g,T} - S_T$.

$$
\begin{aligned}
S_T =& \frac{1}{T}\sum_{t=1}^{T}\left(\iota_N - r_t G_t'\left(\widehat{\theta}_G - \theta_G + \theta_G\right)\right)\left(\iota_N - r_t G_t'\left(\widehat{\theta}_G - \theta_G + \theta_G\right)\right)' \\
=& S_{B_g,T} + \frac{1}{T}\sum_{t=1}^{T}\left(r_t G_t'\left(\widehat{\theta}_G - \theta_G\right)\right)\left(r_t G_t'\left(\widehat{\theta}_G - \theta_G\right)\right)' \\
& - \frac{1}{T}\sum_{t=1}^{T}\left(r_t G_t'\left(\widehat{\theta}_G - \theta_G\right)\right)e_{g,t}' - \frac{1}{T}\sum_{t=1}^{T} e_{g,t}\left(r_t G_t'\left(\widehat{\theta}_G - \theta_G\right)\right)' \qquad (35)
\end{aligned}
$$

Assumption B.2 and proofs of Theorem B.2 then suggest the last two terms be negligible in large samples, and thus for fixed N when T is large $S_T \approx S_{B_g,T} + \frac{1}{T}\sum_{t=1}^{T}\left(r_t G_t'\left(\widehat{\theta}_G - \theta_G\right)\right)\left(r_t G_t'\left(\widehat{\theta}_G - \theta_G\right)\right)'$. This then lead to the conclusion. □

Proof of Lemma 3.1. Assumption B.2 implies that for θ_G, $AR(\theta_G) \to_d \chi^2(N)$ which then implies the result. □

Proof of Theorem 3.2. Notice if $\theta_G \in CS_{r,\alpha_1}$ then $T\widehat{\delta}_g^* \leq Te_{g,T}(\theta_G)'Q_r^{-1}e_{g,T}(\theta_G)$ and $c_{1-\alpha_2}(\theta_G) \leq c_{1-\alpha}^*$. Assumption B.2 implies in large samples

$$
\liminf_T \mathbb{P}\left(Te_{g,T}(\theta_G)'Q_r^{-1}e_{g,T}(\theta_G) \leq c_{1-\alpha_2}(\theta_G)\right) = 1 - \alpha_2
$$

The Lemma 3.1 and $(1-\alpha_1)(1-\alpha_2) = 1-\alpha$ lead to conclusion. □

Proof of Theorem 3.3. We first provide the proof related with the $\mathcal{J}$ statistic, which essentially results from the proof of Theorem 2 in Gospodinov et al. (2017).

Denote $W = TL'\widehat{B}_g'P_1\left(P_1'\Sigma P_1\right)^{-1}P_1'\widehat{B}_g L$, where P_1 is an $N\times(N-1)$ orthogonal matrix whose columns are orthogonal to ι_N such that $P_1'P_1 = I_{N-1}, P_1P-1' = M_{\iota_N}$; L is an lower triangular matrix such that $Q_{\widetilde{G}} = LL'$ and Σ is the covariance matrix. Define $Z = (P_1'\Sigma P_1)^{-1/2}P_1'\widehat{B}_g L$ and $M = (P_1'\Sigma P_1)^{-1/2}P_1'B_g L$, and then

$$
\sqrt{T}\mathrm{vec}\left(Z - M\right) \to_d N(0, I_{(N-1)K})
$$

From the assumption on H, we know there exists $K\times k$ and $K\times K-k$ matrices C_1, C_2 where (C_1, C_2) is a $K\times K$ orthogonal matrix and $\widetilde{M}_1 = MC_1, \widetilde{M}_2 = MC_2$ are of orders $O(1/\sqrt{T}), O(1)$ respectively.

Let $\sqrt{T}\widetilde{M}_1 \to \widetilde{\mu}$, then in case (1) $\widetilde{\mu} = 0$ and in case (2) $\widetilde{\mu}$ is bounded. Let $\widetilde{Z} = (\widetilde{Z}_1, \widetilde{Z}_2) = (ZC_1, ZC_2)$ we would have

$$\sqrt{T}\begin{pmatrix} \text{vec}\left(\widetilde{Z}_1\right) \\ \text{vec}\left(\widetilde{Z}_2 - \widetilde{M}_2\right) \end{pmatrix} \to_d N\left(\begin{pmatrix} \text{vec}\left(\widetilde{\mu}\right) \\ 0 \end{pmatrix}, \quad I_{(N-1)K}\right)$$

The proof of theorem 2 in Gospodinov et al. (2017) shows that: (i) the asymptotic distribution of the $\mathcal{J}$ statistic is the same as the one of the largest eigenvalue, w_k, of $\widetilde{W}^{-1}$ with

$$\widetilde{W} = T\widetilde{Z}_1' M_{\widetilde{Z}_2}\widetilde{Z}_1, \tag{36}$$

and thus (ii) in case (1) where H is of reduced rank and $\widetilde{\mu} = 0$, $\widetilde{W} \to_d \mathcal{W}_k\left(N - K - 1 + k, I_r\right)$ and

$$\mathbb{P}\left(w_k \leq a\right) \leq \mathbb{P}\left(x_k \leq a\right), \quad x_k \sim \chi^2_{N-k},$$

In case (2), where $\widetilde{\mu} \neq 0$, $\widetilde{W}$ follows a non-central Wishart distribution $W_k(N-K-1+k, I_r, \mu)$ with $\mu = \widetilde{\mu}' M_{\widetilde{M}_2}\widetilde{\mu}, \|\mu\| \leq L < \infty$ asymptotically, which then implies the inconsistency of the $\mathcal{J}$ test. The consistency of the HJS test is obvious since $\left\|\iota_N - q_{\widetilde{G}}\theta_{\widetilde{G}}\right\| > a > 0, \forall\theta_{\widetilde{G}} \in \Theta$ implies that $\left\|\iota_N - q_{\widetilde{G},T}\theta_{\widetilde{G}}\right\| = O_p(\sqrt{T}), \forall\theta_{\widetilde{G}} \in \Theta$.

□

Example B.1. *We use an specific example, where we suppose Assumptions 2.1, - 2.3 hold with* $Q_{\widetilde{G}} = \mathbb{E}(G_t G_t')$, $K \geq K_{g,2} \geq 1$, *to show that over the supreme,* $T\sup_{\theta\in CS_{r,\alpha_1}} \delta_{g,T}(\theta)$, *is not properly bounded by* $c^*_{1-\alpha}$ *in the sense that there would be* $\alpha > 0$ *such that*

$$\liminf_T \mathbb{P}\left(\left(T\sup_{\theta\in CS_{r,\alpha_1}} \delta_{g,T}(\theta)\right) \geq c^*_{1-\alpha}\right) > \alpha.$$

We prove this by discussing elements in the confidence set. We group θ*s in* CS_{r,α_1} *into two classes: (1)* θ*s with entries corresponding to strong proxy factors deviating from their true values; (2)* θ*s with entries corresponding to weak proxy factors deviating from their true values. Notice for any* θ*s belong to class (1),* $T\delta_{g,T}(\theta)$ *is of order* $O_p(T)$*, while for* θ*s belong to class (2) we have* $T\delta_{g,T}(\theta) = T\delta_{g,T}(\theta_G) + O_p(1)$*, and* $S_T(\theta) = S_T(\theta_G) + O_p(1/\sqrt{T})$.

Now we only need to show that the confidence set in the presence of weak proxy factors contains some θ*s in class-(2) with positive probability* $\widetilde{\alpha}_1 \geq l > 0$ *in large samples, which then leads to the conclusion. The proof of Theorem 1 in Gospodinov et al. (2017) implies that*

$$\mathcal{J}(\theta) = CD(\theta) + T\frac{\left(\iota_N' S_T(\theta)^{-1}\left(\iota_N - q_{G,T}\theta\right)\right)^2}{\iota_N' S_T(\theta)^{-1}\iota_N} + O_p(\frac{1}{\sqrt{T}})$$

where $CD(\theta) = T\theta' q_{G,T}' P_1 \left(\left(\theta' \otimes P_1'\right) \widehat{V}\left(\sqrt{T} vec(q_{G,T})\right)\left(\theta \otimes P_1\right)\right) P_1' q_{G,T}\theta$*. Notice* $\inf_\theta CD(\theta)$ *is a rank test (Kleibergen and Paap (2006)), and in the presence of weak factors* $CD\left(\theta_G + (0_{K\times 1}, 1)\right)$ *converges to a non-central chi-square distribution which would then implies our claim.*

C Consistent θ estimator

Proofs related to section 4 relies heavily on the properties of our four-pass estimator, and thus we first discuss our four-pass estimator before we provide proofs. This section contains the following results: (1) we first show that the number of the strong factors can be estimated consistently; (2) we then show common component can be estimated consistently when $\sqrt{N}/T \to 0$; (3) we show that with proxies for the common component, θ can be consistently derived even in the presence of weak identification issues via our proposed estimator and we also derive its asymptotic distribution.

Assumption C.1. $\|(\beta;\gamma)_i\| = \|c_{\beta\gamma,i}\| \leq L$. $Q_{(\beta;\gamma)} = \lim_{N\to\infty} (\beta;\gamma)'(\beta;\gamma)/N$ *with* $Q_{(\beta;\gamma)}$ *a* $K_{vz} \times K_{vz}$ *positive definite matrix with* $0 < l \leq \lambda_{K_g}\left(Q_{(\beta;\gamma)}\right) \leq \lambda_1\left(Q_{(\beta;\gamma)}\right) < L < \infty$. *(Assumption 2.3 implies this assumption via the Ostrowski theorem and some extra mild assumptions on* Q_r*.) such that*

$$\left\|N^{-1}c'_{\beta\gamma}c_{\beta\gamma} - Q_{\beta\gamma}\right\| = o_p(1)$$

Assumption C.2. *Let* $\gamma_N(t,t') = \mathbb{E}\left(N^{-1}\sum_{i=1}^N e_{it}e_{it'}\right)$*, there exists a positive constant* L *such that*

$$(1) T^{-1}\sum_{t=1}^T\sum_{t'=1}^T |\gamma_N(t,t')| \leq L; \max_t |\gamma_N(t,t')| \leq L$$
$$(2) T^{-2}\sum_{t=1}^T\sum_{t'=1}^T \mathbb{E}\left(\sum_{i=1}^N e_{it}e_{it'} - \mathbb{E}\left(\sum_{i=1}^N e_{it}e_{it'}\right)\right)^2 = T^{-2}\sum_{t=1}^T\sum_{t'=1}^T\left(\mathbb{E}\left(\sum_{i=1}^N e_{it}e_{it'}\right)^2 - N^2\gamma_N(t,t')^2\right) \leq LN$$

Assumption C.2 is implies by Assumption 2.2.

Assumption C.3. $\mathbb{E}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N\left(e_{it}e_{is} - \mathbb{E}e_{it}e_{is}\right)\right|^4 < L < \infty$

Assumption C.4. *This is implied by Assumptions 2.1 and 2.2 (Weak dependence between proxy* *factors and Idiosyncratic Errors from Bai and Ng (2002))*

$$(i)\quad \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_t e_{it}\right\|^2\right)\leq L$$
$$(ii)\quad \mathbb{E}\left(\frac{1}{K_{vz}}\sum_{i=1}^{K_{vz}}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}g_t u_{vz,it}\right\|^2\right)\leq L$$
$$(iii)\quad \mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{vz,t} e_{it}\right\|^2\right)\leq L$$

Assumption C.5. *For all N and T,*

$$\mathbb{E}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}c_{\beta\gamma,ij}e_{it}\right)^2 \leq LNT$$
$$\mathbb{E}\left(\sum_{t=1}^{T}\sum_{i=1}^{N}c_{\beta\gamma,ij}e_{it}\right)^2 \leq LNT \tag{37}$$

Assumption C.6.

$$\left\|T^{-1}u_{vz}'u_{vz}-\Sigma_{vz}\right\|_{\max}=O_p(T^{-\frac{1}{2}})$$

with Σ_{vz} a positive definite matrix.

Assumption C.7.

$$\sup_t \mathbb{E}\left\|u_{vz,t}\right\|^4\leq L$$

This is one assumption identical to one imposed on factors (Assumption A) in Bai and Ng (2002) and Anatolyev and Mikusheva (2018) impose slightly stronger assumption.

Assumption C.8.

$$(1)\sum_{t}^{T}|\gamma_N(s,t)|\leq L$$
$$(2)\sum_{i=1}^{N}|\tau(i,j)|\leq L$$

with $\mathbb{E}e_{it}e_{jt} = \tau_t(i,j), |\tau_t(i,j)| \le |\tau(i,j)|$.

Assumption C.9. *(i)*

$$\frac{1}{NT}\mathbb{E}\left\|\sum_{s=1}^{T}\sum_{n=1}^{N}u_{vz,s}\left(e_{is}e_{it}-\mathbb{E}(e_{is}e_{it})\right)\right\|^2 \le L$$

(ii) for each t*, as* $N \to \infty$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}c_{\beta\gamma,i}e_{ti} \to_d N(0,\Pi_t)$$

with $\Pi_t = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}c_{\beta\gamma,i}c'_{\beta\gamma,j}\mathbb{E}e_{ti}e_{tj}$. *And for all* $j = 1,\cdots,N$

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}c_{\beta\gamma,i}\left(e_{ti}e_{tj}-\mathbb{E}e_{ti}e_{tj}\right) = O_p(1)$$

(iii)

$$\mathbb{E}\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}u_{vz,t}c'_{\beta\gamma,i}e_{it}\right\|_F^2 \le M$$

Assumption C.9 is identical to the Assumption F in Bai (2003).

C.1 (1)

In this section we prove that we can estimate the number of the strong factors in the $u_{g,t}$ consistently. Here we only provide one way to estimate the number, the estimation approach is not unique. Bai and Ng (2002) propose multiple consistent estimators for the number of strong factors with different penalty functions. Here we use the one employed in Giglio and Xiu (2017).

Under Assumption 2.1, we know

$$r_t = \widetilde{c} + \beta d_g \bar{g}_t + u_{g,t} \tag{38}$$

with $u_{g,t} = \beta v_t + \gamma z_t + e_t, \widetilde{c} = c + \beta d_g(\bar{g} - \mu_g)$. v_t is assumed to be of a $K_g \times 1$ vector, but the dimension of z_t is unknown. We estimate the number K_{vz} of the omitted strong factors by

$$\widehat{K}_{vz} = \arg \min_{K_g \leq j \leq K_{vz,\max}} \left(N^{-1}T^{-1}\lambda_j \left(\widehat{u}_g \widehat{u}_g' \right) + j\phi(N,T) \right) - 1 \tag{39}$$

where $\widehat{u}_g$ is $T \times N$ matrix stacked with the residuals $\widehat{u}_{g,t}$, $K_{vz,\max}$ is an arbitrary upper bound for K_{vz} and $\phi(N,T)$ is a penalty function with the properties $\phi(N,T) \to 0, \phi(N,T)/(N^{-\frac{1}{2}} + T^{-\frac{1}{2}}) \to \infty$. Now we show this estimator is consistent.

Theorem C.1. *Suppose Assumptions 2.1, C.1 - C.6 hold, let N, T increase then*

$$\widehat{K}_{vz} \to_p K_{vz}$$

Proof. We basically follow the steps in Giglio and Xiu (2017) with small changes in the middle.

(1) We first prove the claim such that for $1 \leq j \leq K_{vz}$

$$|N^{-1}T^{-1}\lambda_j \left(\widehat{u}_g \widehat{u}_g' \right) - \lambda_j \left((Q_{(\beta;\gamma)})^{\frac{1}{2}} \Sigma_{(v;z)} (Q_{(\beta;\gamma)})^{\frac{1}{2}} \right)| = o_p(1) \tag{40}$$

with $Q_{(\beta;\gamma)} = \lim_{N\to\infty} (\beta;\gamma)'(\beta;\gamma)/N$.

For convenience, in this proof, denote $c_{\beta\gamma} = (\beta;\gamma), u_{vz} = (v;z)$.

(1.1) Notice

$$\widehat{u}_g \widehat{u}_g' - M_{\bar{G}} u_{vz} c_{\beta\gamma}' c_{\beta\gamma} u_{vz}' M_{\bar{G}} = M_{\bar{G}} u_{vz} c_{\beta\gamma}' e' M_{\bar{G}} + M_{\bar{G}} e c_{\beta\gamma} u_{vz}' M_{\bar{G}} + M_{\bar{G}} e e' M_{\bar{G}}$$

with $\widehat{u}_g = M_{\bar{G}} r$. We show in the following steps that the three terms on the right is negligible when divided by NT.

For the term $M_{\bar{G}} e e' M_{\bar{G}}$, we have

$$\left\| M_{\bar{G}} e e' M_{\bar{G}} - n\Gamma_u \right\| \leq \left\| e e' - n\Gamma_u \right\|_F + 2 \left\| P_{\bar{G}} e e' \right\|_F + \left\| P_{\bar{G}} e e' P_{\bar{G}} \right\|_F \tag{41}$$

Assumption C.2.(1) implies that

$$\mathbb{E}\left\|e\right\|_F^2 = \mathbb{E}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2 \leq LNT \tag{42}$$

$$\mathbb{E}\left\|\bar{e}\right\|_F^2 = T^{-2}\mathbb{E}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{t'=1}^{T} e_{it}e_{it'} = NT^{-2}\sum_{t=1}^{T}\sum_{t'=1}^{T}|\gamma_N(t,t')| \leq LNT^{-1} \tag{43}$$

Assumption C.2.(2) implies that

$$\mathbb{E}\left\|ee' - n\Gamma_u\right\|_F^2 = \sum_{t=1}^{T}\sum_{t'=1}^{T}\mathbb{E}\left(\sum_{i=1}^{N} e_{it}e_{it'} - \mathbb{E}\left(\sum_{i=1}^{N} e_{it}e_{it'}\right)\right)^2 \leq LNT^2 \tag{44}$$

Assumption C.4 implies that

$$\mathbb{E}\left\|G'e\right\|_F^2 = NT\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} G_t e_{it}\right\|_F^2\right) \leq LNT \tag{45}$$

$$\left\|G'e\right\|_F = O_p(N^{\frac{1}{2}}T^{\frac{1}{2}}); \left\|\left(0, \iota_T\bar{g}'\right)' e\right\|_F = O_p(N^{\frac{1}{2}}T^{\frac{1}{2}}) \tag{46}$$

$$\left\|\bar{G}'e\right\|_F \leq \left\|G'e\right\|_F + \left\|\left(0, \iota_T\bar{g}'\right)' e\right\|_F = O_p(N^{\frac{1}{2}}T^{\frac{1}{2}}) \tag{47}$$

$$\left\|P_{\bar{G}}e\right\|_F \leq T^{-1}\left\|\bar{G}\right\|_F\left\|\left(\bar{G}'\bar{G}/T\right)^{-1}\right\|_F\left\|\bar{G}'e\right\|_F = O_p(N^{\frac{1}{2}}) \tag{48}$$

Then it follows that

$$\left\|M_{\bar{G}}ee'M_{\bar{G}} - n\Gamma_u\right\| \leq \left\|ee' - n\Gamma_u\right\|_F + 2\left\|P_{\bar{G}}ee'\right\|_F + \left\|P_{\bar{G}}ee'P_{\bar{G}}\right\|_F$$
$$= O_p(N^{\frac{1}{2}}T) + O_p(NT^{\frac{1}{2}}) + O_p(N) \tag{49}$$

From assumption C.4 (1) we would have

$$\mathbb{E}\left\|\Gamma_u\right\|_F^2 = \mathbb{E}\sum_{t=1}^{T}\sum_{t'=1}^{T}|\gamma_N(t,t')| \leq LT \tag{50}$$

$$\left\|M_{\bar{G}}ee'M_{\bar{G}}\right\|_F \leq \left\|M_{\bar{G}}ee'M_{\bar{G}} - n\Gamma_u\right\|_F + \left\|n\Gamma_u\right\|_F = O_p(N^{\frac{1}{2}}T) + O_p(NT^{\frac{1}{2}})$$

For the term $M_{\bar{G}}u_{vz}c'_{\beta\gamma}e'M_{\bar{G}}$, we have

$$\left\|M_{\bar{G}}u_{vz}c'_{\beta\gamma}e'M_{\bar{G}}\right\|_F \leq \left\|u_{vz}c'_{\beta\gamma}e'\right\|_F + \left\|P_{\bar{G}}u_{vz}c'_{\beta\gamma}e'\right\|_F + \left\|u_{vz}c'_{\beta\gamma}e'P_{\bar{G}}\right\|_F + \left\|P_{\bar{G}}u_{vz}c'_{\beta\gamma}e'P_{\bar{G}}\right\|_F$$
$$\leq (2(K_{vz}+1))^2\left\|u_{vz}c'_{\beta\gamma}e'\right\|_F \tag{51}$$

Assumption C.5 implies that

$$\mathbb{E}\left\|ec_{\beta\gamma}\right\|_F^2 = \mathbb{E}\sum_{j=1}^{K_{vz}}\sum_{t=1}^{T}\left(\sum_{i=1}^{N}c_{\beta\gamma,ij}e_{it}\right)^2 \leq LNT \tag{52}$$

Thus

$$\left\|u_{vz}\right\| \leq KT\left\|T^{-1}u'_{vz}u_{vz}\right\|_{\max} = O_p(T^{\frac{1}{2}}) \tag{53}$$

$$\left\|u_{vz}c'_{\beta\gamma}e'\right\| \leq \left\|u_{vz}\right\|\left\|c'_{\beta\gamma}e'\right\|_F = O_p(N^{\frac{1}{2}}T) \tag{54}$$

$$\left\|M_{\bar{G}} u_{vz} c'_{\beta\gamma} e' M_{\bar{G}}\right\| \leq O_p(N^{\frac{1}{2}} T) \tag{55}$$

Therefore

$$N^{-1}T^{-1}|\lambda_j(\widehat{u}_g \widehat{u}'_g) - \lambda_j(M_{\bar{G}} u_{vz} c'_{\beta\gamma} c_{\beta\gamma} u'_{vz} M_{\bar{G}})| = o_p(1) \tag{56}$$

(1.2) To finish the proof of part (1) we only need to show the following two results:

$$(1.2.1)\quad |N^{-1}T^{-1}\lambda_j(M_{\bar{G}} u_{vz} c'_{\beta\gamma} c_{\beta\gamma} u'_{vz} M_{\bar{G}}) - T^{-1}\lambda_j(M_{\bar{G}} u_{vz} Q_{\beta\gamma} u'_{vz} M_{\bar{G}})| = o_p(1) \tag{57}$$

$$(1.2.2)\quad |T^{-1}\lambda_j(M_{\bar{G}} u_{vz} Q_{\beta\gamma} u'_{vz} M_{\bar{G}}) - \lambda_j\left((Q_{(\beta;\gamma)})^{\frac{1}{2}} \Sigma_{(v;z)} (Q_{(\beta;\gamma)})^{\frac{1}{2}}\right)| = o_p(1) \tag{58}$$

The equation 57 is a direct result of Assumption C.1 and Weyl's inequality such that

$$\begin{aligned} &|N^{-1}T^{-1}\lambda_j(M_{\bar{G}} u_{vz} c'_{\beta\gamma} c_{\beta\gamma} u'_{vz} M_{\bar{G}}) - T^{-1}\lambda_j(M_{\bar{G}} u_{vz} Q_{\beta\gamma} u'_{vz} M_{\bar{G}})| \\ \leq & T^{-1}\left\|N^{-1}c'_{\beta\gamma} c_{\beta\gamma} - Q_{\beta\gamma}\right\| \left\|M_{\bar{G}} u_{vz}\right\|^2 = o_p(1) \end{aligned} \tag{59}$$

The equation 57 is a direct result of Assumption C.6 and Weyl's inequality such that

$$\begin{aligned} &|T^{-1}\lambda_j(M_{\bar{G}} u_{vz} Q_{\beta\gamma} u'_{vz} M_{\bar{G}}) - \lambda_j\left((Q_{(\beta;\gamma)})^{\frac{1}{2}} \Sigma_{vz} (Q_{(\beta;\gamma)})^{\frac{1}{2}}\right)| \\ \leq & L\left\|T^{-1}u'_{vz} u_{vz} - \Sigma_{vz}\right\| = O_p(T^{-\frac{1}{2}}) \end{aligned} \tag{60}$$

Therefore, for $1 \leq j \leq K_{vz}$

$$|N^{-1}T^{-1}\lambda_j\left(\widehat{u}_g \widehat{u}'_g\right) - \lambda_j\left((Q_{(\beta;\gamma)})^{\frac{1}{2}} \Sigma_{(v;z)} (Q_{(\beta;\gamma)})^{\frac{1}{2}}\right)| = o_p(1) \tag{61}$$

(2) In part (2) we finish the proof of the consistency $\widehat{K}_{vz}$ by showing the following statement

$$\lim \inf_{T,N\to\infty} \mathbb{P}\left(\widehat{F}(K_{vz}+1) \leq \widehat{F}(j), j = 1, \cdots, N\right) = 1 \tag{62}$$

with $\widehat{F}(j) = N^{-1}T^{-1}\lambda_j\left(\widehat{u}_g\widehat{u}_g'\right) + j\phi(N,T)$. Notice a direct result from step (1) is that for $1 \leq j \leq K_{vz}$, we can find $l > 0$ such that $\widehat{F}(j) > l + o_p(1)$, while for $K_{vz}+1 \leq j$, $\widehat{F}(j) = o_p(1)$. Then we only need to show that for $j > K_{vz}+1$, $\widehat{F}(j) > \widehat{F}(K_{vz}+1)$ with probability approaching to one. Notice for $j \geq K_{vz}+1$,

$$\lambda_j\left(\widehat{u}_g\widehat{u}_g'\right) \leq \left\|ee'\right\| = O_p(TN^{-\frac{1}{2}} + T^{-\frac{1}{2}}N) \tag{63}$$

and this implies $\widehat{F}(j) > \widehat{F}(K_{vz}+1)$ with probability approaching to one as $(N^{\frac{1}{2}}T^{\frac{1}{2}})(j - K_{vz} - 1)\phi(N,T) > (N^{-\frac{1}{2}}T^{-\frac{1}{2}})\left(\lambda_{K_{vz}+1}\left(\widehat{u}_g\widehat{u}_g'\right) - \lambda_j\left(\widehat{u}_g\widehat{u}_g'\right)\right)$ with probability approaching to one.

□

C.2 (2)

Denote $m_{NT} = \min\{N,T\}$.

Lemma C.2. *Suppose Assumptions 2.1 - 2.2, C.1 - C.6 hold, let N,T increase then*

$$m_{NT}\frac{1}{T}\sum_{t=1}^{T}\left\|\widetilde{u}_{vz,t} - H'u_{vz,t}\right\|_F^2 = O_p(1)$$

with $H = \left(c_{\beta\gamma}'c_{\beta\gamma}/N\right)\left(u_{vz}'\widetilde{u}_{vz}/T\right)V_{NT}$ and V_{NT} being the $K_{vz}\times K_{vz}$ diagonal matrix of $\lambda_i\left(\widehat{u}_g\widehat{u}_g'/NT\right), i = 1,\cdots,K_{vz}$.

Proof. This proof resembles the proof of Theorem 1 in Bai and Ng (2002). From the normalization $\widetilde{u}_{vz}'\widetilde{u}_{vz}/T = I_{K_{vz}}$ we know

$$\|\widetilde{u}_{vz}\|_F^2 = \sum_{t=1}^{T}\|\widetilde{u}_{vz,t}\|_F^2 = O_p(T) \tag{64}$$

From the proof in section and the above equation we know

$$\|H\|_F \leq \left\|c_{\beta\gamma}'c_{\beta\gamma}/N\right\|_F\left\|u_{vz}'u_{vz}/T\right\|_F^{1/2}\left\|\widetilde{u}_{vz}'\widetilde{u}_{vz}/T\right\|_F^{1/2} = O_p(1) \tag{65}$$

From the equation (80) and (40), we know

$$\left\|\widetilde{u}_{vz,t} - H' u_{vz,t}\right\|_F^2 \leq L \sum_{i=1}^{6} a_{i,t}^2 \tag{66}$$

with

$$\begin{aligned}
a_{1,t}^2 =& T^{-2} \left\| \sum_{s=1}^{T} \widetilde{u}_{vz,s} \gamma_N(s,t) \right\|_F^2 \\
a_{2,t}^2 =& T^{-2} \left\| \sum_{s=1}^{T} \widetilde{u}_{vz,s} \zeta_{ee,st} \right\|_F^2 \\
a_{3,t}^2 =& T^{-2} \left\| \sum_{s=1}^{T} \widetilde{u}_{vz,s} \zeta_{ue,st} \right\|_F^2 \\
a_{4,t}^2 =& T^{-2} \left\| \sum_{s=1}^{T} \widetilde{u}_{vz,s} \zeta_{eu,st} \right\|_F^2 \\
a_{5,t}^2 =& \left\| \frac{1}{NT} \widetilde{u}_{vz}' P_{\bar{G}} u_g u_{g,t}' \right\|_F^2 \\
a_{6,t}^2 =& \left\| \frac{1}{NT} \widetilde{u}_{vz}' M_{\bar{G}} u_g \left(\widehat{u}_{g,t} - u_{g,t}\right) \right\|_F^2
\end{aligned} \tag{67}$$

From the equation (64) and Assumption C.2 (i),

$$\sum_{t=1}^{T} a_{1,t}^2 \leq \left(T^{-1} \sum_{s=1}^{T} \|\widetilde{u}_{vz,s}\|_F^2 \right) \left(T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} \gamma_N(s,t)^2 \right) = O_p(1) \tag{68}$$

From the equation (64) and Assumption 2.2,

$$\sum_{t=1}^{T} a_{2,t}^2 \leq \left(T^{-1} \sum_{s=1}^{T} \|\widetilde{u}_{vz,s}\|_F^2 \right) \left(T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\sum_{u=1}^{T} \zeta_{ee,su} \zeta_{ee,ut} \right)^2 \right)^{\frac{1}{2}} = O_p(T/N) \tag{69}$$

and from the equation (52)

$$\sum_{t=1}^{T} a_{3,t}^2 \leq \left(\sum_{t=1}^{T} \left\|c_{\beta\gamma}' e_t / N\right\|_F^2 \right) \left(T^{-1} \sum_{s=1}^{T} \|\widetilde{u}_{vz,s}\|_F^2 \right) \left(T^{-1} \sum_{s=1}^{T} \|u_{vz,s}\|_F^2 \right) = O_p(T/N) \tag{70}$$

and similarly $\sum_{t=1}^{T} a_{4,t}^2 \leq O_p(T/N)$.

From Assumption 2.2,

$$\left\|P_{\bar{G}} u_g\right\|_F \leq T^{-1}\left\|\bar{G}\right\|_F\left\|\left(\bar{G}'\bar{G}/T\right)^{-1}\right\|_F\left\|\bar{G}'u_g\right\|_F = O_p(N^{\frac{1}{2}}) \tag{71}$$

and thus

$$\sum_{t=1}^{T} a_{5,t}^2 \leq \left(T^{-1}\sum_{s=1}^{T}\left\|\widetilde{u}_{vz,s}\right\|_F^2\right)\left(N^{-1}\left\|P_{\bar{G}} u_g\right\|_F^2\right)\left(N^{-1}T^{-1}\sum_{t=1}^{T}\left\|u_{g,t}\right\|_F^2\right) = O_p(1) \tag{72}$$

From the equation (40) and (63)

$$\left\|M_{\bar{G}} u_g\right\|_F^2 = O_p(NT) \tag{73}$$

and since $\sqrt{T}\left(\widehat{\beta}_g - \beta_g\right) = O_p(1)$

$$\sum_{t=1}^{T}\left\|\widehat{u}_{g,t} - u_{g,t}\right\|_F^2 \leq \left\|\sqrt{T}\left(\widehat{\beta}_g - \beta_g\right)\right\|_F^2\left\|\bar{G}/\sqrt{T}\right\|_F^2 \leq O_p(N) \tag{74}$$

Therefore,

$$\sum_{t=1}^{T} a_{6,t}^2 \leq \left(T^{-1}\sum_{s=1}^{T}\left\|\widetilde{u}_{vz,s}\right\|_F^2\right)\left(N^{-1}T^{-1}\left\|M_{\bar{G}} u_g\right\|_F^2\right)\left(N^{-1}\sum_{t=1}^{T}\left\|\widehat{u}_{g,t} - u_{g,t}\right\|_F^2\right) \leq O_p(1) \tag{75}$$

It follows

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\widetilde{u}_{vz,t} - H'u_{vz,t}\right\|_F^2 = O_p(1/N) + O_p(1/T) \tag{76}$$

□

Given the discussion in the last subsection, it would be safe to assume that we know the value of K_{vz}, namely the number of the strong omitted factors are known. In this section, we show the asymptotic properties of the common component estimator. Essentially, this section is a building block for the consistent θ_g estimator as introduced in the next section.

The method of principal components method gives the following estimators

$$(\widetilde{c}_{\beta\gamma}, \widetilde{u}_{vz}) = \arg \min_{c_i, u_t \ s.t. \ \sum_{t=1}^{T} u_t u_t'/T = I_{K_{vz}}} \sum_{i,t} \left(\widehat{u}_{g,it} - c_i' u_t\right)^2 \tag{77}$$

The estimator $\widetilde{u}_{vz}$ is equal to the $\sqrt{T}$ times eigenvector associated with the K_{vz} largest eigenvalues of the matrix $\widehat{u}_g \widehat{u}_g'$, and $\widetilde{c}_{\beta\gamma} = \widetilde{u}_{vz}' \widehat{u}_g / T$ corresponds to the OLS estimator regressing $\widehat{u}_g$ over $\widetilde{u}_v z$. Specially we are interested in the common component estimator $\widetilde{c} = \widetilde{u}_{vz}(\widetilde{c}_{\beta\gamma})'$, which would serve as a proxy for the common component in u_g in our proposed estimator.

Theorem C.3. *Suppose Assumptions 2.1 - 2.2, C.1 - C.6, C.7, C.9 hold, and* $\sqrt{N}/T \to 0$,

$$\sqrt{N}\left(\widetilde{u}_{vz,t} - H' u_{vz,t}\right) = V_{NT}^{-1} \frac{1}{T} \sum_{s=1}^{T} \left(\widetilde{u}_{vz,s} u_{vz,s}'\right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_{\beta\gamma,i} e_{it} + o_p(1) \tag{78}$$

when $\liminf \sqrt{N}/T \to \tau > 0$,

$$T\left(\widetilde{u}_{vz,t} - H' u_{vz,t}\right) = O_p(1) \tag{79}$$

Proof. We make use the equation $\widetilde{u}_{vz} = \frac{1}{NT} M_{\bar{G}} u_g u_g' M_{\bar{G}} \widetilde{u}_{vz} V_{NT}^{-1}$, and $H = \left(c_{\beta\gamma}' c_{\beta\gamma}/N\right)\left(u_{vz}' \widetilde{u}_{vz}/T\right) V_{NT}^{-1}$, and V_{NT} is the $K_{vz} \times K_{vz}$ diagonal matrix of $\lambda_i\left(\widehat{u}_g \widehat{u}_g'/NT\right), i = 1, \cdots, K_{vz}$. Then

$$\begin{aligned}
&\widetilde{u}_{vz,t} - H' u_{vz,t} \\
=& \frac{1}{NT} V_{NT}^{-1} \widetilde{u}_{vz}' \left(M_{\bar{G}} u_g \widehat{u}_{g,t} - u_{vz}\left(c_{\beta\gamma}' c_{\beta\gamma}\right) u_{vz,t}\right) \\
=& \frac{1}{NT} V_{NT}^{-1} \widetilde{u}_{vz}' \left(u_g u_{g,t} - u_{vz}\left(c_{\beta\gamma}' c_{\beta\gamma}\right) u_{vz,t} - P_{\bar{G}} u_g u_{g,t} + M_{\bar{G}} u_g \left(\widehat{u}_{g,t} - u_{g,t}\right)\right) \\
=& \frac{1}{NT} V_{NT}^{-1} \widetilde{u}_{vz}' \left(\left(u_{vz} c_{\beta\gamma}' + e\right)\left(c_{\beta\gamma} u_{vz,t} + e_t\right) - u_{vz}\left(c_{\beta\gamma}' c_{\beta\gamma}\right) u_{vz,t} - P_{\bar{G}} u_g u_{g,t} + M_{\bar{G}} u_g \left(\widehat{u}_{g,t} - u_{g,t}\right)\right) \\
=& V_{NT}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T} \widetilde{u}_{vz,s} \gamma_N(s,t) + \frac{1}{T}\sum_{s=1}^{T} \widetilde{u}_{vz,s} \zeta_{ee,st} + \frac{1}{T}\sum_{s=1}^{T} \widetilde{u}_{vz,s} \zeta_{ue,st} + \frac{1}{T}\sum_{s=1}^{T} \widetilde{u}_{vz,s} \zeta_{eu,st}\right) \\
&+ V_{NT}^{-1}\left(-\frac{1}{NT} \widetilde{u}_{vz}' P_{\bar{G}} u_g u_{g,t} + \frac{1}{NT} \widetilde{u}_{vz}' M_{\bar{G}} u_g \left(\widehat{u}_{g,t} - u_{g,t}\right)\right)
\end{aligned} \tag{80}$$

with

$$\zeta_{ee,st} = e_s' e_t / N - \gamma_N(s,t) \tag{81}$$

$$\zeta_{ue,st} = u_{vz,s}' c_{\beta\gamma}' e_t / N \tag{82}$$

$$\zeta_{eu,st} = u_{vz,t}' c_{\beta\gamma}' e_s / N \tag{83}$$

Now we analyze the terms on the left hand of equation (80).

(1) Assumption C.7 and Assumption C.2 imply that

$$\frac{1}{T}\sum_{s=1}^{T} u_{vz,s}\gamma_N(s,t) = O_p(1/T) \tag{84}$$

Lemma C.2 and Assumption C.2 imply that

$$\begin{aligned}\frac{1}{T}\left\|\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s} - u_{vz,s}\right)\gamma_N(s,t)\right\|_F \leq & \frac{1}{\sqrt{T}}\left(\frac{1}{T}\sum_{s=1}^{T}\|\widetilde{u}_{vz,s} - u_{vz,s}\|_F^2\right)^{1/2}\left(\sum_{t}^{T}|\gamma_N(s,t)|^2\right)^{1/2} \\ = & O_p(1/(\sqrt{T m_{NT}}))\end{aligned} \tag{85}$$

and thus

$$\frac{1}{T}\sum_{s=1}^{T}\widetilde{u}_{vz,s}\gamma_N(s,t) = \frac{1}{T}\sum_{s=1}^{T} u_{vz,s}\gamma_N(s,t) + \frac{1}{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s} - u_{vz,s}\right)\gamma_N(s,t) \leq O_p(1/(\sqrt{T m_{NT}})) \tag{86}$$

(2)

Assumption C.9 implies that

$$\frac{1}{T}\sum_{s=1}^{T} u_{vz,s}\zeta_{ee,st} = O_p(1/\sqrt{NT}) \tag{87}$$

Assumption 2.2 implies

$$\frac{1}{T}\sum_{s=1}^{T}|\zeta_{ee,st}|^2 = \frac{1}{T}\frac{1}{N}\sum_{s=1}^{T}\left|\frac{1}{\sqrt{N}}\left(e_{is}e_{it} - \mathbb{E}(e_{is}e_{it})\right)\right|^2 = O_p(1/(N)) \tag{88}$$

Lemma C.2 and equation (88) imply that

$$\frac{1}{T}\left\|\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s}-u_{vz,s}\right)\zeta_{ee,st}\right\| \leq \left(\frac{1}{T}\sum_{s=1}^{T}\|\widetilde{u}_{vz,s}-u_{vz,s}\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}|\zeta_{ee,st}|^2\right)^{1/2}$$
$$=O_p(1/(\sqrt{Nm_{NT}})) \tag{89}$$

and thus

$$\frac{1}{T}\sum_{s=1}^{T}\widetilde{u}_{vz,s}\zeta_{ee,st} =\frac{1}{T}\sum_{s=1}^{T}u_{vz,s}\zeta_{ee,st}+\frac{1}{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s}-u_{vz,s}\right)\zeta_{ee,st} \leq O_p(1/(\sqrt{Nm_{NT}})) \tag{90}$$

(3) Assumption C.9 implies that

$$\frac{1}{T}\sum_{s=1}^{T}u_{vz,s}\zeta_{ue,st} = \left(\frac{1}{T}\sum_{s=1}^{T}u_{vz,s}u'_{vz,s}\right)\frac{1}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}c_{\beta\gamma,i}e_{ti}\right) = O_p(1/\sqrt{N}) \tag{91}$$

and

$$\frac{1}{T}\sum_{s=1}^{T}|\zeta_{ue,st}|^2 =\frac{1}{T}\sum_{s=1}^{T}\left(u'_{vz,s}c'_{\beta\gamma}e_t/N\right)^2 \leq \|u_{vz}\|_F/T\left\|c'_{\beta\gamma}e'_t/\sqrt{N}\right\|_F/N = O_p(1/N) \tag{92}$$

Lemma C.2 and equation (92) imply that

$$\frac{1}{T}\left\|\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s}-u_{vz,s}\right)\zeta_{ue,st}\right\| \leq \left(\frac{1}{T}\sum_{s=1}^{T}\|\widetilde{u}_{vz,s}-u_{vz,s}\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t}^{T}|\zeta_{ue,st}|^2\right)^{1/2}$$
$$=O_p(1/(\sqrt{Nm_{NT}})) \tag{93}$$

and thus

$$\frac{1}{T}\sum_{s=1}^{T}\widetilde{u}_{vz,s}\zeta_{ue,st} =\frac{1}{T}\sum_{s=1}^{T}u_{vz,s}\zeta_{ue,st}+\frac{1}{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s}-u_{vz,s}\right)\zeta_{ue,st} = O_p(1/\sqrt{N}) \tag{94}$$

(4) Lemma C.2 and Assumption C.9 imply that

$$\left\|\frac{1}{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s}-u_{vz,s}\right)e_s'c_{\beta\gamma}/Nu_{vz,t}\right\|_F \leq \frac{1}{\sqrt{N}}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{u}_{vz,s}-u_{vz,s}\right\|_F^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|e_s'c_{\beta\gamma}/\sqrt{N}\right\|_F^2\right)^{\frac{1}{2}}\left\|u_{vz,t}\right\|_F$$

$$=O_p(1/\sqrt{Nm_{NT}}) \tag{95}$$

Assumption C.9 implies

$$\frac{1}{\sqrt{NT}}\left(\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}u_{vz,s}e_s'c_{\beta\gamma}\right)u_{vz,t}=O_p(\frac{1}{\sqrt{NT}}) \tag{96}$$

and thus

$$\frac{1}{T}\sum_{s=1}^{T}\widetilde{u}_{vz,s}\zeta_{ue,st}=\frac{1}{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s}-u_{vz,s}\right)e_s'c_{\beta\gamma}/Nu_{vz,t}+\frac{1}{T}\sum_{s=1}^{T}u_{vz,s}e_s'c_{\beta\gamma}/Nu_{vz,t}\leq O_p(1/\sqrt{Nm_{NT}}) \tag{97}$$

(5) Similar to the derivation of equation (47)

$$\left\|\bar{G}'u_g\right\|_F=O_p(N^{\frac{1}{2}}T^{\frac{1}{2}});\quad \left\|\bar{G}'u_{vz}\right\|_F=O_p(T^{\frac{1}{2}}) \tag{98}$$

which implies

$$\frac{1}{NT}\left\|\left(\widetilde{u}_{vz}-u_{vz}\right)'\bar{G}\left(\bar{G}'\bar{G}\right)^{-1}\bar{G}'u_gu_{g,t}'\right\|_F$$

$$\leq\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\widetilde{u}_{vz,t}-u_{vz,t}\right\|_F^2\right)^{\frac{1}{2}}\frac{1}{\sqrt{T}}\left(\left\|\bar{G}/\sqrt{T}\left(\bar{G}'\bar{G}/T\right)^{-1}\right\|_F\right)\left(\frac{1}{\sqrt{NT}}\left\|\bar{G}'u_g\right\|_F\right)\left\|u_{g,t}/\sqrt{N}\right\|_F$$

$$=O_p(1/\sqrt{Tm_{NT}}) \tag{99}$$

and

$$\begin{aligned}
&\frac{1}{NT}\left\|u_{vz}'\bar{G}\left(\bar{G}'\bar{G}\right)^{-1}\bar{G}'u_g u_{g,t}'\right\|_F \\
\leq&\left(\frac{1}{\sqrt{T}}\left\|u_{vz}'\bar{G}\right\|_F\right)\frac{1}{T}\left(\left\|\left(\bar{G}'\bar{G}/T\right)^{-1}\right\|_F\right)\left(\frac{1}{\sqrt{NT}}\left\|\bar{G}'u_g\right\|_F\right)\left\|u_{g,t}/\sqrt{N}\right\|_F \\
=&O_p(1/T)
\end{aligned} \tag{100}$$

Therefore,

$$\begin{aligned}
\frac{1}{NT}\widetilde{u}_{vz}'P_{\bar{G}}u_g u_{g,t}' =&\frac{1}{NT}\left(\widetilde{u}_{vz}-u_{vz}\right)'\bar{G}\left(\bar{G}'\bar{G}\right)^{-1}\bar{G}'u_g u_{g,t}'+\frac{1}{NT}u_{vz}'\bar{G}\left(\bar{G}'\bar{G}\right)^{-1}\bar{G}'u_g u_{g,t}' \\
\leq&O_p(1/\sqrt{Tm_{NT}})
\end{aligned} \tag{101}$$

(6)

$$\frac{1}{NT}\left\|u_{vz}'u_{vz}c_{\beta\gamma}'\left(\widehat{u}_{g,t}-u_{g,t}\right)\right\|_F \leq\frac{1}{\sqrt{NT}}\left\|u_{vz}'u_{vz}/T\right\|_F\left\|\frac{\sqrt{T}c_{\beta\gamma}'\left(\widehat{u}_{g,t}-u_{g,t}\right)}{\sqrt{N}}\right\|_F = O_p(\frac{1}{\sqrt{NT}}) \tag{102}$$

$$\frac{1}{NT}\left\|u_{vz}'e\left(\widehat{u}_{g,t}-u_{g,t}\right)\right\|_F \leq \frac{1}{T}\left\|u_{vz}'e/\sqrt{NT}\right\|_F\left\|\frac{\sqrt{T}\left(\widehat{u}_{g,t}-u_{g,t}\right)}{\sqrt{N}}\right\|_F = O_p(\frac{1}{T}) \tag{103}$$

and thus

$$\frac{1}{NT}\left\|\widetilde{u}_{vz}'M_{\bar{G}}u_g\right\|_F \leq\frac{1}{NT}\left\|\widetilde{u}_{vz}'u_g\left(\widehat{u}_{g,t}-u_{g,t}\right)\right\|_F+\frac{1}{NT}\left\|\widetilde{u}_{vz}'P_{\bar{G}}u_g\left(\widehat{u}_{g,t}-u_{g,t}\right)\right\|_F = O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T}) \tag{104}$$

From the above discussion, when $\sqrt{N}/T \to 0$, only the third term in equation 80 matters in the asymptotic behavior and thus

$$\sqrt{N}\left(\widetilde{u}_{vz,t}-H'u_{vz,t}\right) = V_{NT}^{-1}\frac{1}{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s}u_{vz,s}'\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}c_{\beta\gamma,i}e_{it}+o_p(1) \tag{105}$$

when $\liminf \sqrt{N}/T \to \tau > 0$, it is straightforward that

$$T\left(\widetilde{u}_{vz,t} - H' u_{vz,t}\right) = O_p(1) \tag{106}$$

□

Theorem C.4. *Suppose Assumptions 2.1 - 2.2, C.1 - C.9 hold, and* $\sqrt{T}/N \to 0$

$$\sqrt{T}\left(\widetilde{c}_{\beta\gamma,i} - H^{-1} c_{\beta\gamma,i}\right) = H' \frac{1}{\sqrt{T}} u_{vz}' e_i + o_p(1)$$

Proof.

$$\begin{aligned}
\widetilde{c}_{\beta\gamma,i} =& \frac{1}{T}\widetilde{u}_{vz}' \widehat{u}_{g,i} \\
=& \frac{1}{T}\widetilde{u}_{vz}' \left(\left(\widetilde{u}_{vz} H^{-1} + u_{vz} - \widetilde{u}_{vz} H^{-1}\right) c_{\beta\gamma,i} + e_i\right) - \frac{1}{T}\widetilde{u}_{vz}' \left(\widehat{u}_{g,i} - u_{g,i}\right) \\
=& H^{-1} c_{\beta\gamma,i} + \frac{1}{T} H' u_{vz}' e_i + \frac{1}{T}\widetilde{u}_{vz}' \left(u_{vz} - \widetilde{u}_{vz} H^{-1}\right) c_{\beta\gamma,i} + \frac{1}{T}\left(\widetilde{u}_{vz}' - H' u_{vz}'\right) e_i - \frac{1}{T}\widetilde{u}_{vz}' \left(\widehat{u}_{g,i} - u_{g,i}\right)
\end{aligned} \tag{107}$$

We show the last three terms are at most $O_p(1/m_{NT})$.

(1) For the term $\frac{1}{T}\left(\widetilde{u}_{vz} - u_{vz} H\right)' e_i$, equation (80) implies

$$\begin{aligned}
\frac{1}{T}\left(\widetilde{u}_{vz} - u_{vz} H\right)' e_i =& V_{NT}^{-1}\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{u}_{vz,s}\gamma_N(s,t) e_{it} + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{u}_{vz,s}\zeta_{ee,st} e_{it} + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{u}_{vz,s}\zeta_{ue,st} e_{it}\right) \\
&+ V_{NT}^{-1}\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{u}_{vz,s}\zeta_{eu,st} e_{it} - \frac{1}{NT^2}\sum_{t=1}^{T}\widetilde{u}_{vz}' P_{\bar{G}} u_g u_{g,t} e_{it} + \frac{1}{NT^2}\sum_{t=1}^{T}\widetilde{u}_{vz}' M_{\bar{G}} u_g \left(\widehat{u}_{g,t} - u_{g,t}\right) e_{it}\right) \\
=& V_{NT}^{-1}\sum_{j=1}^{6} a_j
\end{aligned} \tag{108}$$

There are six terms on the left hand side of the above equation, and we analyze each of $a_j, j = 1, \cdots, 6$ to determine the order of $\frac{1}{T}\left(\widetilde{u}_{vz} - u_{vz} H\right)' e_i$.

Assumption 2.2 and Lemma 1(i) in Bai and Ng (2002) imply that

$$\begin{aligned}&\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s}-H'u_{vz,s}\right)\gamma_N(s,t)e_{it}\right\|_F\\ \le&\frac{1}{\sqrt{T}}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{u}_{vz,s}-H'u_{vz,s}\right\|_F^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma_N(s,t)^2\frac{1}{T}\sum_{t=1}^{T}e_{it}^2\right)^{\frac{1}{2}}\le O_p(\frac{1}{\sqrt{Tm_{NT}}})\end{aligned} \tag{109}$$

Assumption 2.2,C.7 and Lemma 1(i) in Bai and Ng (2002) imply that

$$\mathbb{E}\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}u_{vz,s}\gamma_N(s,t)e_{it}\right\|_F\le\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\mathbb{E}\left\|u_{vz,s}\right\|_F^2\right)^{\frac{1}{2}}\gamma_N(s,t)^2(\mathbb{E}e_{it}^2)^{\frac{1}{2}}=O(\frac{1}{T}) \tag{110}$$

Therefore,

$$a_1=\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s}-H'u_{vz,s}\right)\gamma_N(s,t)e_{it}+\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}H'u_{vz,s}\gamma_N(s,t)e_{it}\le O_p(\frac{1}{\sqrt{Tm_{NT}}}) \tag{111}$$

Assumption 2.2 implies that

$$\frac{1}{T}\sum_{t=1}^{T}\zeta_{ee,st}e_{it}=\frac{1}{\sqrt{N}}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(e_{is}e_{it}-\mathbb{E}\left(e_{is}e_{it}\right)\right)\right)e_{it}=O_p(\frac{1}{\sqrt{N}}) \tag{112}$$

Equation (112) and Lemma C.2 imply that

$$\begin{aligned}\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s}-H'u_{vz,s}\right)\zeta_{ee,st}e_{it}\right\|_F\le&\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{u}_{vz,s}-H'u_{vz,s}\right\|_F^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{s=1}^{T}\left(\frac{1}{T}\sum_{t=1}^{T}\zeta_{ee,st}e_{it}\right)^2\right)^{\frac{1}{2}}\\ \le&O_p(\frac{1}{\sqrt{Nm_{NT}}})\end{aligned} \tag{113}$$

Assumption C.9 implies that

$$\begin{aligned}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}u_{vz,s}\zeta_{ee,st}e_{it}=&\frac{1}{\sqrt{NT}}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{i=1}^{N}u_{vz,s}\left(e_{is}e_{it}-\mathbb{E}\left(e_{is}e_{it}\right)\right)\right)e_{it}\right)\\ \le&O_p(\frac{1}{\sqrt{NT}})\end{aligned} \tag{114}$$

and thus

$$a_2 = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s} - H' u_{vz,s}\right)\zeta_{ee,st}e_{it} + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} H' u_{vz,s}\zeta_{ee,st}e_{it} \leq O_p(\frac{1}{\sqrt{N m_{NT}}}) \tag{115}$$

Assumption C.9 implies that

$$\frac{1}{T}\sum_{t=1}^{T}\zeta_{ue,st}e_{it} = \frac{1}{\sqrt{N}}u'_{vz,s}\left(\frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{N}}\sum_{i=1}c_{\beta\gamma,i}e_{it}\right)e_{it}\right) = O_p(\frac{1}{\sqrt{N}}) \tag{116}$$

and thus with Theorem C.3 we know

$$\begin{aligned}\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s} - H' u_{vz,s}\right)\zeta_{ue,st}e_{it}\right\|_F \leq \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{u}_{vz,s} - H' u_{vz,s}\right\|_F^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{s=1}^{T}\left(\frac{1}{T}\sum_{t=1}^{T}\zeta_{ue,st}e_{it}\right)^2\right)^{\frac{1}{2}} \\ \leq O_p(\frac{1}{\sqrt{N m_{NT}}})\end{aligned} \tag{117}$$

Assumption C.8 and C.1 imply that

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}\|c_{\beta\gamma,i}\tau_t(i,j)\|_F \leq L\frac{1}{N}\sum_{j=1}^{N}\|\tau(i,j)\|_F = O_p(\frac{1}{N}) \tag{118}$$

Assumption C.9 implies that $\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}c_{\beta\gamma,i}\left(e_{jt}e_{it} - \tau_t(i,j)\right) = O_p(\frac{1}{\sqrt{NT}})$ and thus with equation (118) we know

$$\begin{aligned}\frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}c_{\beta\gamma,i}e_{jt}e_{it} = \frac{1}{NT}\sum_{t=1}^{T}\sum_{j=1}^{N}c_{\beta\gamma,i}\left(e_{jt}e_{it} - \tau_t(i,j) + \tau_t(i,j)\right) \\ \leq O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{N})\end{aligned} \tag{119}$$

Therefore,

$$a_3 = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s} - H' u_{vz,s}\right)\zeta_{ue,st}e_{it} + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} H' u_{vz,s}\zeta_{ue,st}e_{it} \leq O_p(\frac{1}{m_{NT}}) \tag{120}$$

Assumption C.9 implies that

$$\frac{1}{T}\sum_{t=1}^{T}\zeta_{eu,st}e_{it} = \frac{1}{\sqrt{N}}\left(\frac{1}{T}\sum_{t=1}^{T}e_{it}u_{vz,t}'\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N}c_{\beta\gamma,j}e_{js}\right)\right) = O_p(\frac{1}{\sqrt{N}}) \tag{121}$$

and thus with Theorem C.3 we know

$$\begin{aligned}\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s}-H'u_{vz,s}\right)\zeta_{eu,st}e_{it}\right\|_F \leq &\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{u}_{vz,s}-H'u_{vz,s}\right\|_F^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{s=1}^{T}\left(\frac{1}{T}\sum_{t=1}^{T}\zeta_{ue,st}e_{it}\right)^2\right)^{\frac{1}{2}} \\ \leq &O_p(\frac{1}{\sqrt{Nm_{NT}}})\end{aligned} \tag{122}$$

Assumption C.9 implies that

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}u_{vz,s}\zeta_{eu,st}e_{it} = \frac{1}{\sqrt{NT}}\left(\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}u_{vz,s}e_s'c_{\beta\gamma}\right)\left(\frac{1}{T}\sum_{t=1}^{T}u_{vz,t}e_{it}\right) = O_p(\frac{1}{\sqrt{NT}}) \tag{123}$$

Therefore,

$$a_4 = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s}-H'u_{vz,s}\right)\zeta_{eu,st}e_{it} + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}H'u_{vz,s}\zeta_{eu,st}e_{it} \leq O_p(\frac{1}{m_{NT}}) \tag{124}$$

Similar to the derivation of equation (100)

$$\left\|\frac{1}{NT^2}\sum_{t=1}^{T}\widetilde{u}_{vz}'P_{\bar{G}}u_g u_{g,t}e_{it}\right\|_F \leq \left\|\frac{1}{NT}\sum_{t=1}^{T}\widetilde{u}_{vz}'P_{\bar{G}}u_g\right\|_F \frac{\sqrt{N}}{\sqrt{T}}\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}u_{g,t}e_{it}\right\|_F = O_p(\frac{1}{T}) \tag{125}$$

$$a_5 = -\frac{1}{NT^2}\sum_{t=1}^{T}\widetilde{u}_{vz}'P_{\bar{G}}u_g u_{g,t}e_{it} \leq O_p(\frac{1}{T}) \tag{126}$$

Equation (104) implies that

$$a_6 = \frac{1}{NT^2}\sum_{t=1}^{T}\widetilde{u}_{vz}'M_{\bar{G}}u_g\left(\widehat{u}_{g,t}-u_{g,t}\right)e_{it} \leq O_p(\frac{1}{m_{NT}}) \tag{127}$$

(2) For the term $\frac{1}{T}\widetilde{u}_{vz}'\left(u_{vz}-\widetilde{u}_{vz}H^{-1}\right)c_{\beta\gamma,i}$

Similar to the analysis of the term $\frac{1}{T}\left(\widetilde{u}_{vz}-u_{vz}H\right)' e_i$, equation (80) implies

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left(\widetilde{u}_{vz}-u_{vz}H\right)' u_{vz} \\
=&V_{NT}^{-1}\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{u}_{vz,s}u_{vz,t}'\gamma_N(s,t)+\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{u}_{vz,s}u_{vz,t}'\zeta_{ee,st}+\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{u}_{vz,s}u_{vz,t}'\zeta_{ue,st}\right) \\
&+V_{NT}^{-1}\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{u}_{vz,s}u_{vz,t}'\zeta_{eu,st}-\frac{1}{NT^2}\sum_{t=1}^{T}\widetilde{u}_{vz}'P_{\bar{G}}u_g u_{g,t}u_{vz,t}'+\frac{1}{NT^2}\sum_{t=1}^{T}\widetilde{u}_{vz}'M_{\bar{G}}u_g\left(\widehat{u}_{g,t}-u_{g,t}\right)u_{vz,t}'\right) \\
&=V_{NT}^{-1}\sum_{j=1}^{6}a_j
\end{aligned} \tag{128}$$

Following the same proof as the one for a_1 in the term $\frac{1}{T}\left(\widetilde{u}_{vz}-u_{vz}H\right)' e_i$, Assumption C.6 and Lemma 1(i) in Bai and Ng (2002) imply that

$$\begin{aligned}
&\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s}-H'u_{vz,s}\right)\gamma_N(s,t)u_{vz,t}'\right\|_F \\
\leq&\frac{1}{\sqrt{T}}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{u}_{vz,s}-H'u_{vz,s}\right\|_F^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\gamma_N(s,t)^2\frac{1}{T}\sum_{t=1}^{T}u_{vz,t}'u_{vz,t}\right)^{\frac{1}{2}}\leq O_p\left(\frac{1}{\sqrt{Tm_{NT}}}\right)
\end{aligned} \tag{129}$$

and Assumption C.7 and Lemma 1(i) in Bai and Ng (2002) imply that

$$\mathbb{E}\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}u_{vz,s}\gamma_N(s,t)u_{vz,t}\right\|_F\leq\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\mathbb{E}\left\|u_{vz,s}\right\|_F^2\right)^{\frac{1}{2}}\gamma_N(s,t)^2(\mathbb{E}\left\|u_{vz,t}\right\|_F^2)^{\frac{1}{2}}=O\left(\frac{1}{T}\right) \tag{130}$$

and thus

$$a_1\leq O_p\left(\frac{1}{\sqrt{Tm_{NT}}}\right) \tag{131}$$

Assumption C.9 and Lemma C.2 imply that

$$\left\| \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\widetilde{u}_{vz,s} - H' u_{vz,s} \right) \zeta_{ee,st} u'_{vz,t} \right\|_F \leq \left(\frac{1}{T} \sum_{s=1}^{T} \left\| \widetilde{u}_{vz,s} - H' u_{vz,s} \right\|_F^2 \right)^{\frac{1}{2}} \left(\frac{1}{T} \sum_{s=1}^{T} \left\| \frac{1}{T} \sum_{t=1}^{T} \zeta_{ee,st} u'_{vz,t} \right\|_F^2 \right)^{\frac{1}{2}}$$
$$\leq O_p\left(\frac{1}{\sqrt{NTm_{NT}}}\right) \tag{132}$$

Assumption C.9 implies

$$\frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1}^{T} u_{vz,s} u'_{vz,t} \zeta_{ee,st} = \frac{1}{\sqrt{NT}} \left(\frac{1}{T} \sum_{s=1}^{T} u_{vz,s} \left(\frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(e_{it} e_{is} - \mathbb{E} e_{it} e_{is} \right) u'_{vz,t} \right) \right) = O_p\left(\frac{1}{\sqrt{NT}}\right) \tag{133}$$

and thus

$$a_2 = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\widetilde{u}_{vz,s} - H' u_{vz,s} \right) u'_{vz,t} \zeta_{ee,st} + \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} H' u_{vz,s} u'_{vz,t} \zeta_{ee,st} \leq O_p\left(\frac{1}{m_{NT}}\right) \tag{134}$$

Assumption C.9 implies that

$$\frac{1}{T} \sum_{t=1}^{T} \zeta_{ue,st} u'_{vz,t} = \frac{1}{\sqrt{NT}} u'_{vz,s} \left(\frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} c_{\beta\gamma,i} e_{it} u'_{vz,t} \right) = O_p\left(\frac{1}{\sqrt{NT}}\right) \tag{135}$$

and thus

$$\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} u_{vz,s} \zeta_{ue,st} u'_{vz,t} = \frac{1}{\sqrt{NT}} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} u_{vz,s} u'_{vz,s} \left(\frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} c_{\beta\gamma,i} e_{it} u'_{vz,t} \right) = O_p\left(\frac{1}{\sqrt{NT}}\right) \tag{136}$$

and with Lemma C.2 we know

$$\left\| \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\widetilde{u}_{vz,s} - H' u_{vz,s} \right) \zeta_{ue,st} u_{vz,t} \right\|_F \leq \left(\frac{1}{T} \sum_{s=1}^{T} \left\| \widetilde{u}_{vz,s} - H' u_{vz,s} \right\|_F^2 \right)^{\frac{1}{2}} \left(\frac{1}{T} \sum_{s=1}^{T} \left(\frac{1}{T} \sum_{t=1}^{T} \zeta_{ue,st} u_{vz,t} \right)^2 \right)^{\frac{1}{2}}$$
$$\leq O_p\left(\frac{1}{\sqrt{NTm_{NT}}}\right) \tag{137}$$

Therefore,

$$a_3 = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s} - H' u_{vz,s}\right)\zeta_{ue,st}u'_{vz,t} + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} H' u_{vz,s}\zeta_{ue,st}u'_{vz,t} \leq O_p(\frac{1}{\sqrt{NT}}) \tag{138}$$

Assumption C.9 implies that

$$\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} u_{vz,s}\zeta_{eu,st}u'_{vz,t} = \frac{1}{\sqrt{NT}}\left(\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{i=1}^{N} u_{vz,s}c'_{\beta\gamma,i}e_{is}\right)\left(\frac{1}{T}\sum_{t=1}^{T} u_{vz,t}u'_{vz,t}\right) = O_p(\frac{1}{\sqrt{NT}}) \tag{139}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\zeta_{eu,st}u_{vz,t} = \frac{1}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N} c'_{\beta\gamma,i}e_{is}\right)\left(\frac{1}{T}\sum_{t=1}^{T} u_{vz,t}u'_{vz,t}\right) = O_p(\frac{1}{\sqrt{N}}) \tag{140}$$

and thus with Lemma C.2 we know

$$\begin{aligned}\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s} - H' u_{vz,s}\right)\zeta_{ue,st}u_{vz,t}\right\|_F &\leq \left(\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{u}_{vz,s} - H' u_{vz,s}\right\|_F^2\right)^{\frac{1}{2}}\left(\frac{1}{T}\sum_{s=1}^{T}\left(\frac{1}{T}\sum_{t=1}^{T}\zeta_{ue,st}u_{vz,t}\right)^2\right)^{\frac{1}{2}} \\ &\leq O_p(\frac{1}{\sqrt{Nm_{NT}}})\end{aligned} \tag{141}$$

$$a_4 = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s} - H' u_{vz,s}\right)\zeta_{eu,st}u'_{vz,t} + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} H' u_{vz,s}\zeta_{eu,st}u'_{vz,t} \leq O_p(\frac{1}{\sqrt{Nm_{NT}}}) \tag{142}$$

Similar to the derivation of equation (126)

$$a_5 = -\frac{1}{NT^2}\sum_{t=1}^{T}\widetilde{u}'_{vz}P_{\bar{G}}u_g u_{g,t}u'_{vz,t} \leq O_p(\frac{1}{T}) \tag{143}$$

and similar to the derivation of equation (127)

$$a_6 = \frac{1}{NT^2}\sum_{t=1}^{T}\widetilde{u}'_{vz}M_{\bar{G}}u_g\left(\widehat{u}_{g,t} - u_{g,t}\right)u'_{vz,t} \leq O_p(\frac{1}{m_{NT}}) \tag{144}$$

Therefore,

$$\frac{1}{T}\left(\widetilde{u}_{vz}-u_{vz}H\right)'\widetilde{u}_{vz}=\frac{1}{T}\left(\widetilde{u}_{vz}-u_{vz}H\right)'u_{vz}H+\frac{1}{T}\left(\widetilde{u}_{vz}-u_{vz}H\right)'\left(\widetilde{u}_{vz}-u_{vz}H\right)\leq O_p(\frac{1}{m_{NT}}) \quad (145)$$

(3) For the term $\frac{1}{T}\widetilde{u}'_{vz}\left(\widehat{u}_{g,i}-u_{g,i}\right)$

$$\left\|\frac{1}{T}\left(\widetilde{u}_{vz}-u_{vz}H\right)'P_{\bar{G}}u_{g,i}\right\|_F\leq\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\widetilde{u}_{vz,s}-H'u_{vz,s}\right\|_F^2\right)^{\frac{1}{2}}\frac{1}{\sqrt{T}}\left\|P_{\bar{G}}u_{g,i}\right\|_F\leq O_p(\frac{1}{\sqrt{T}m_{NT}}) \quad (146)$$

Similar to the derivation of equation (100)

$$\left\|\frac{1}{T}u'_{vz}P_{\bar{G}}u_{g,i}\right\|_F\leq O_p(\frac{1}{T}) \quad (147)$$

and thus

$$\frac{1}{T}\widetilde{u}'_{vz}\left(\widehat{u}_{g,i}-u_{g,i}\right)=\frac{1}{T}\left(\widetilde{u}_{vz}-u_{vz}H\right)'P_{\bar{G}}u_{g,i}+\frac{1}{T}u'_{vz}P_{\bar{G}}u_{g,i} \quad (148)$$

□

Theorem C.5. *Suppose Assumptions 2.1 - 2.2, C.1 - C.9 hold, let N, T increase then*

$$\begin{aligned}m_{NT}^{\frac{1}{2}}\left(\widetilde{c}_{it}-c_{it}\right)=&\frac{m_{NT}^{\frac{1}{2}}}{\sqrt{N}}c'_{\beta\gamma,i}\left(c'_{\beta\gamma}c_{\beta\gamma}/N\right)^{-1}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}c_{\beta\gamma,j}e_{jt}\\&+\frac{m_{NT}^{\frac{1}{2}}}{\sqrt{T}}u'_{vz,t}\left(u'_{vz}u_{vz}/T\right)^{-1}\frac{1}{\sqrt{T}}u'_{vz}e_i+O_p(\frac{1}{\sqrt{m_{NT}}})\end{aligned}$$

Proof.

$$\begin{aligned}\widetilde{c}_{it}-c_{it}=&\widetilde{u}'_{vz,t}\widetilde{c}_{\beta\gamma,i}-u'_{vz,t}c_{\beta\gamma,i}=\left(\widetilde{u}_{vz,t}-H'u_{vz,t}\right)'H^{-1}c_{\beta\gamma,i}+u'_{vz,t}H\left(\widetilde{c}_{\beta\gamma,i}-H^{-1}c_{\beta\gamma,i}\right)\\&+\left(\widetilde{u}_{vz,t}-H'u_{vz,t}\right)'\left(\widetilde{c}_{\beta\gamma,i}-H^{-1}c_{\beta\gamma,i}\right)\\=&\left(\widetilde{u}_{vz,t}-H'u_{vz,t}\right)'H^{-1}c_{\beta\gamma,i}+u'_{vz,t}H\left(\widetilde{c}_{\beta\gamma,i}-H^{-1}c_{\beta\gamma,i}\right)+O_p(\frac{1}{m_{NT}})\end{aligned} \quad (149)$$

Theorem C.3 implies that

$$
\begin{aligned}
m_{NT}^{\frac{1}{2}} c_{\beta\gamma,i}' H'^{-1}\left(\widetilde{u}_{vz,t}-H' u_{vz,t}\right) = & \frac{m_{NT}^{\frac{1}{2}}}{\sqrt{N}} c_{\beta\gamma,i}' H'^{-1} V_{NT}^{-1} \frac{1}{T} \sum_{s=1}^{T}\left(\widetilde{u}_{vz,s} u_{vz,s}'\right) \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_{\beta\gamma,i} e_{it} + o_p(1) \\
= & \frac{m_{NT}^{\frac{1}{2}}}{\sqrt{N}} c_{\beta\gamma,i}'\left(c_{\beta\gamma}' c_{\beta\gamma}/N\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_{\beta\gamma,i} e_{it} + o_p(1)
\end{aligned} \tag{150}
$$

and Theorem C.4 implies that

$$
\begin{aligned}
m_{NT}^{\frac{1}{2}} u_{vz,t}' H\left(\widetilde{c}_{\beta\gamma,i}-H^{-1} c_{\beta\gamma,i}\right) = & \frac{m_{NT}^{\frac{1}{2}}}{\sqrt{T}} u_{vz,t}' H H' \frac{1}{\sqrt{T}} u_{vz}' e_i + o_p(1) \\
= & \frac{m_{NT}^{\frac{1}{2}}}{\sqrt{T}} u_{vz,t}'\left(u_{vz}' u_{vz}/T\right)^{-1} \frac{1}{\sqrt{T}} u_{vz}' e_i + o_p(1)
\end{aligned} \tag{151}
$$

where the last equality is due to

$$
H H' = \left(u_{vz}' u_{vz}/T\right)^{-1} + O_p\left(\frac{1}{m_{NT}}\right) \tag{152}
$$

and finally Therefore,

$$
\begin{aligned}
m_{NT}^{\frac{1}{2}}\left(\widetilde{c}_{it}-c_{it}\right) = & \frac{m_{NT}^{\frac{1}{2}}}{\sqrt{N}} c_{\beta\gamma,i}'\left(c_{\beta\gamma}' c_{\beta\gamma}/N\right)^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} c_{\beta\gamma,i} e_{it} \\
& + \frac{m_{NT}^{\frac{1}{2}}}{\sqrt{T}} u_{vz,t}'\left(u_{vz}' u_{vz}/T\right)^{-1} \frac{1}{\sqrt{T}} u_{vz}' e_i + O_p\left(\frac{1}{m_{NT}}\right)
\end{aligned} \tag{153}
$$

□

C.3 (3)

Proof of Theorem 4.1. Rewrite moment conditions as consider

$$
\iota_N = \widetilde{q}_{G,T} \widetilde{\theta} + \widetilde{\epsilon} \tag{154}
$$

with $\widetilde{\theta} = \theta_G + \left(0, \left(\left(-\widehat{V}_g^{-1}\left(\bar{g}-\mu_g\right), \left(\widehat{V}_g^{-1} V_g - I_K\right)\right)\theta_G\right)'\right)'$.

Equation (154) implies

$$\widetilde{\theta}^{(1)}_{g,I} - \widetilde{\theta}^{(1)}_{g} = \left(\widetilde{q}^{(1)\prime}_{G,T} P_{\widetilde{q}^{(2)}_{G,T}} \widetilde{q}^{(1)}_{G,T}\right)^{-1} \widetilde{q}^{(1)}_{G,T} P_{\widetilde{q}^{(2)}_{G,T}} \widetilde{\epsilon}^{(1)} \tag{155}$$

$$\begin{aligned}\widetilde{q}^{(1)}_{G,T} =& (c, \beta_g)\, \widehat{Q}^{(1)}_{G} + \left(\widetilde{cc}^{(1)} - cc^{(1)}\right)' \bar{G}^{(1)}/|\mathcal{T}_{(1)}| + e^{(1)\prime} \bar{G}^{(1)}/|\mathcal{T}_{(1)}| \\ =& \widetilde{X}^{(1)}_{g,c\beta_g} + \widetilde{X}^{(1)}_{g,cc} + \widetilde{X}^{(1)}_{g,e}\end{aligned}$$

$$\widetilde{x}^{(1)}_{g,c\beta_g,i} = \widehat{Q}^{(1)}_{G} \left(c_i, \beta'_{g,i}\right)' \tag{156}$$

$$\begin{aligned}\widetilde{x}^{(1)}_{g,cc,i} =& \frac{1}{|\mathcal{T}_{(1)}|} \sum_{t\in\mathcal{T}_{(1)}} \left(\widetilde{cc}_{it} - cc_{it}\right) \bar{G}_t \\ =& \frac{1}{|\mathcal{T}_{(1)}|} \sum_{t\in\mathcal{T}_{(1)}} \frac{1}{\sqrt{N}} c'_{\beta\gamma,i} \left(c'_{\beta\gamma} c_{\beta\gamma}/N\right)^{-1} \frac{1}{\sqrt{N}} \sum_{j=1}^{N} c_{\beta\gamma,j} e_{jt} \bar{G}_t \\ &+ \frac{1}{|\mathcal{T}_{(1)}|} \sum_{t\in\mathcal{T}_{(1)}} \frac{1}{\sqrt{T}} u'_{vz,t} \left(u'_{vz} u_{vz}/T\right)^{-1} \frac{1}{\sqrt{T}} u'_{vz} e_i \bar{G}_t + O_p\left(\frac{1}{m_{NT}}\right) \\ =& \frac{1}{\sqrt{N|\mathcal{T}_{(1)}|}} c'_{\beta\gamma,i} \left(c'_{\beta\gamma} c_{\beta\gamma}/N\right)^{-1} \frac{1}{\sqrt{N|\mathcal{T}_{(1)}|}} \sum_{t\in\mathcal{T}_{(1)}} \sum_{i=1}^{N} c_{\beta\gamma,i} e_{it} \bar{G}_t \\ &+ \frac{1}{\sqrt{T|\mathcal{T}_{(1)}|}} \frac{1}{\sqrt{|\mathcal{T}_{(1)}|}} \sum_{t\in\mathcal{T}_{(1)}} \bar{G}_t u'_{vz,t} \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \left(u'_{vz} u_{vz}/T\right)^{-1} u_{vz,s} e_{is} + O_p\left(\frac{1}{m_{NT}}\right) \\ \leq& O_p\left(\frac{1}{\sqrt{m_{NT}|\mathcal{T}_{(1)}|}}\right)\end{aligned} \tag{157}$$

$$\widetilde{x}^{(1)}_{g,e,i} = \frac{1}{\sqrt{|\mathcal{T}_{(1)}|}} \frac{1}{\sqrt{|\mathcal{T}_{(1)}|}} \sum_{t\in\mathcal{T}_{(1)}} e_{it} \bar{G}'_t \tag{158}$$

Suppose $|\mathcal{T}_{(1)}| = |\mathcal{T}_{(2)}| = \tau$. We next discuss the properties of the following three terms: (1) $Q_{B_g,T}\widetilde{q}^{(2)'}_{G,T}\widetilde{q}^{(2)}_{G,T}Q_{B_g,T}$; (2)$\widetilde{q}^{(1)'}_{G,T}\widetilde{q}^{(2)}_{G,T}$; (3)$\widetilde{q}^{(2)'}_{G,T}\widetilde{\epsilon}^{(1)}$.

(1)

$$\frac{\tau}{N}\sum_{i=1}^{N}\widetilde{x}^{(2)}_{g,cc,i}\widetilde{x}^{(2)'}_{g,cc,i} \leq O_p(\frac{1}{\sqrt{\tau}}) \tag{159}$$

$$\frac{1}{N}Q_{B_g,T}\widetilde{X}^{(2);}_{g,c\beta_g}\widetilde{X}^{(2)}_{g,c\beta_g}Q_{B_g,T} \to_p Q_g\eta_{c\beta_g}Q_g \tag{160}$$

Equation (31) implies

$$\frac{1}{N|\mathcal{T}_{(1)}|}\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_{(1)}} e_{it}^2\bar{G}'_t\bar{G}_t \to_p \Sigma_{\bar{G}\bar{G}\gamma_N} \tag{161}$$

and

$$\frac{1}{\sqrt{N}}\frac{1}{\sqrt{N}|\mathcal{T}_{(1)}|}\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_{(1)}}\sum_{s\neq t} e_{it}\bar{G}'_t e_{is}\bar{G}_s = O_p(\frac{1}{\sqrt{N}}) \tag{162}$$

Therefore,

$$\begin{aligned}
\frac{\tau}{N}\widetilde{X}^{(2)'}_{g,e}\widetilde{X}^{(2)}_{g,e} &= \frac{\tau}{N}\sum_{i=1}^{N}\widetilde{x}^{(2)}_{g,e,i}\widetilde{x}^{(2)'}_{g,e,i} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{|\mathcal{T}_{(2)}|}}\sum_{t\in\mathcal{T}_{(2)}} e_t\bar{G}'_t\right)\left(\frac{1}{\sqrt{|\mathcal{T}_{(2)}|}}\sum_{t\in\mathcal{T}_{(2)}} e_t\bar{G}'_t\right)' \\
=&\frac{1}{N|\mathcal{T}_{(2)}|}\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_{(2)}} e_{it}\bar{G}'_t\bar{G}_t e_{it} + \frac{1}{N|\mathcal{T}_{(2)}|}\sum_{i=1}^{N}\sum_{t\in\mathcal{T}_{(2)}}\sum_{s\neq t} e_{it}\bar{G}'_t\bar{G}_s e_{is} \to_p \Sigma_{\bar{G}\bar{G}\gamma_N}
\end{aligned} \tag{163}$$

$$
\begin{aligned}
&\frac{\sqrt{\tau}}{N}\sum_{i=1}^{N} Q_{B_g,T}\widetilde{x}^{(2)}_{g,c\beta_g,i}\widetilde{x}^{(2)\prime}_{g,cc,i} \\
=&\frac{1}{\sqrt{N|\mathcal{T}_{(2)}|}}\frac{\sqrt{\tau}}{N}\sum_{i=1}^{N}\left(\widehat{Q}^{(1)}_{G}Q_{B_g,T}\left(c_i,\beta'_{g,i}\right)'\right)\left(c'_{\beta\gamma,i}\left(c'_{\beta\gamma}c_{\beta\gamma}/N\right)^{-1}\frac{1}{\sqrt{N|\mathcal{T}_{(2)}|}}\sum_{t\in\mathcal{T}_{(2)}}\sum_{i=1}^{N}c_{\beta\gamma,i}e_{it}\bar{G}_t\right)' \\
&+\frac{1}{\sqrt{T|\mathcal{T}_{(2)}|}}\frac{\sqrt{\tau}}{N}\sum_{i=1}^{N}\left(\widehat{Q}^{(2)}_{g}Q_{B_g,T}\left(c_i,\beta'_{g,i}\right)'\right)\left(\frac{1}{\sqrt{|\mathcal{T}_{(2)}|}}\sum_{t\in\mathcal{T}_{(2)}}\bar{G}_t u'_{vz,t}\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\left(u'_{vz}u_{vz}/T\right)^{-1}u_{vz,s}e_{is}\right)' \\
&+O_p(\frac{1}{m_{NT}})\leq O_p(\frac{1}{\sqrt{m_{NT}}})
\end{aligned}
$$

$$
\frac{\sqrt{\tau}}{N}Q_{B_g,T}\sum_{i=1}^{N}\widetilde{x}^{(2)}_{g,c\beta_g,i}\widetilde{x}^{(2)\prime}_{g,e,i}=\frac{\sqrt{\tau}}{N}\sum_{i=1}^{N}\left(\widehat{Q}^{(1)}_{G}Q_{B_g,T}\left(c_i,\beta'_{g,i}\right)'\right)\left(\frac{1}{|\mathcal{T}_{(2)}|}\sum_{t\in\mathcal{T}_{(2)}}e_{it}\bar{G}_t\right)'=O_p(\frac{1}{\sqrt{N}}) \tag{164}
$$

$$
\begin{aligned}
&\frac{\tau}{N}\sum_{i=1}^{N}\widetilde{x}^{(2)}_{g,cc,i}\widetilde{x}^{(2)\prime}_{g,e,i} \\
=&\frac{\tau}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{N|\mathcal{T}_{(2)}|}}c'_{\beta\gamma,i}\left(c'_{\beta\gamma}c_{\beta\gamma}/N\right)^{-1}\frac{1}{\sqrt{N|\mathcal{T}_{(2)}|}}\sum_{t\in\mathcal{T}_{(2)}}\sum_{j=1}^{N}c_{\beta\gamma,j}e_{jt}\bar{G}_t\right)\left(\frac{1}{|\mathcal{T}_{(2)}|}\sum_{t\in\mathcal{T}_{(2)}}e_{it}\bar{G}_t\right)' \\
&+\frac{\tau}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T|\mathcal{T}_{(2)}|}}\frac{1}{\sqrt{|\mathcal{T}_{(2)}|}}\sum_{t\in\mathcal{T}_{(2)}}\bar{G}_t u'_{vz,t}\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\left(u'_{vz}u_{vz}/T\right)^{-1}u_{vz,s}e_{is}+O_p(\frac{1}{m_{NT}})\right) \\
&\times\left(\frac{1}{|\mathcal{T}_{(2)}|}\sum_{t\in\mathcal{T}_{(2)}}e_{it}\bar{G}_t\right)' \\
=&O_p(\frac{1}{\sqrt{\tau}})
\end{aligned} \tag{165}
$$

Therefore,

$$\begin{aligned}
&\frac{1}{N}Q_{B_g,T}\widetilde{q}^{(2)'}_{G,T}\widetilde{q}^{(2)}_{G,T}Q_{B_g,T} = \frac{1}{N}\sum_{i=1}^{N}Q_{B_g,T}\widetilde{x}^{(2)}_{g,i}\widetilde{x}^{(2)'}_{g,i}Q_{B_g,T}\\
=&\frac{1}{N}\sum_{i=1}^{N}Q_{B_g,T}\left(\widetilde{x}^{(2)}_{g,c\beta_g,i}+\widetilde{x}^{(2)}_{g,cc,i}+\widetilde{x}^{(2)}_{g,e,i}\right)\left(\widetilde{x}^{(2)}_{g,c\beta_g,i}+\widetilde{x}^{(2)}_{g,cc,i}+\widetilde{x}^{(2)}_{g,e,i}\right)'Q_{B_g,T}\\
=&\frac{1}{N}Q_{B_g,T}\left(\sum_{i=1}^{N}\widetilde{x}^{(2)}_{g,c\beta_g,i}\widetilde{x}^{(2)'}_{g,c\beta_g,i}\right)Q_{B_g,T}+Q_{B_g,T}/\sqrt{\tau}\left(\frac{\tau}{N}\sum_{i=1}^{N}\widetilde{x}^{(2)}_{g,cc,i}\widetilde{x}^{(2)'}_{g,cc,i}\right)Q_{B_g,T}/\sqrt{\tau}\\
&+Q_{B_g,T}/\sqrt{\tau}\left(\frac{\tau}{N}\sum_{i=1}^{N}\widetilde{x}^{(2)}_{g,e,i}\widetilde{x}^{(2)'}_{g,e,i}\right)Q_{B_g,T}/\sqrt{\tau}+O_p(\frac{1}{\sqrt{\tau}})\\
=&Q_g\eta_{c\beta_g}Q_g+W_x\Sigma_{\bar{G}\bar{G}\gamma_N}W_x+O_p(\frac{1}{\sqrt{\tau}})
\end{aligned}\tag{166}$$

with $W_x = \lim_{\tau\to\infty}Q_{B_g,T}/\sqrt{\tau}$.

(2)

$$\frac{1}{N}Q_{B_g,T}\widetilde{X}^{(1);}_{g,c\beta_g}\widetilde{X}^{(2)}_{g,c\beta_g}Q_{B_g,T}\to_p Q_g\eta_{c\beta_g}Q_g \tag{167}$$

$$\frac{\tau}{N}\sum_{i=1}^{N}\widetilde{x}^{(1)}_{g,cc,i}\widetilde{x}^{(2)'}_{g,cc,i}\leq O_p(\frac{1}{\sqrt{\tau}}) \tag{168}$$

$$\begin{aligned}
&\frac{\tau}{N}\sum_{i=1}^{N}\widetilde{x}^{(1)}_{g,e,i}\widetilde{x}^{(2)'}_{g,e,i}=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{|\mathcal{T}_{(1)}|}}\sum_{s\in\mathcal{T}_{(1)}}e_s\bar{G}'_s\right)\left(\frac{1}{\sqrt{|\mathcal{T}_{(2)}|}}\sum_{t\in\mathcal{T}_{(2)}}e_t\bar{G}'_t\right)'\\
=&\frac{1}{N\tau}\sum_{i=1}^{N}\sum_{s\in\mathcal{T}_{(1)}}\sum_{t\in\mathcal{T}_{(2)}}e_{it}\bar{G}'_t\bar{G}_s e_{is}=O_p(\frac{1}{\sqrt{N}})
\end{aligned}\tag{169}$$

$$
\begin{aligned}
&\frac{\sqrt{\tau}}{N}\sum_{i=1}^{N} Q_{B_g,T}\widetilde{x}^{(1)}_{g,c\beta_g,i}\widetilde{x}^{(2)\prime}_{g,cc,i} \\
=&\frac{1}{\sqrt{N|\mathcal{T}_{(2)}|}}\frac{\sqrt{\tau}}{N}\sum_{i=1}^{N}\left(\widehat{Q}_G^{(1)}Q_{B_g,T}\left(c_i,\beta_{g,i}'\right)'\right)\left(c_{\beta\gamma,i}'\left(c_{\beta\gamma}'c_{\beta\gamma}/N\right)^{-1}\frac{1}{\sqrt{N|\mathcal{T}_{(2)}|}}\sum_{t\in\mathcal{T}_{(2)}}\sum_{i=1}^{N}c_{\beta\gamma,i}e_{it}\bar{G}_t\right)' \\
&+\frac{1}{\sqrt{T|\mathcal{T}_{(2)}|}}\frac{\sqrt{\tau}}{N}\sum_{i=1}^{N}\left(\widehat{Q}_G^{(1)}Q_{B_g,T}\left(c_i,\beta_{g,i}'\right)'\right)\left(\frac{1}{\sqrt{|\mathcal{T}_{(2)}|}}\sum_{t\in\mathcal{T}_{(2)}}\bar{G}_t u_{vz,t}'\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\left(u_{vz}'u_{vz}/T\right)^{-1}u_{vz,s}e_{is}\right)' \\
&+O_p(\frac{1}{\sqrt{m_{NT}}}) = O_p(\frac{1}{\sqrt{m_{NT}}})
\end{aligned}
$$

$$
\frac{\sqrt{\tau}}{N}Q_{B_g,T}\sum_{i=1}^{N}\widetilde{x}^{(1)}_{g,c\beta_g,i}\widetilde{x}^{(2)\prime}_{g,e,i} = \frac{\sqrt{\tau}}{N}\sum_{i=1}^{N}\left(\widehat{Q}_G^{(1)}Q_{B_g,T}\left(c_i,\beta_{g,i}'\right)'\right)\left(\frac{1}{|\mathcal{T}_{(2)}|}\sum_{t\in\mathcal{T}_{(2)}}e_{it}\bar{G}_t\right)' = O_p(\frac{1}{\sqrt{N}}) \tag{170}
$$

$$
\begin{aligned}
&\frac{\tau}{N}\sum_{i=1}^{N}\widetilde{x}^{(1)}_{g,cc,i}\widetilde{x}^{(2)\prime}_{g,e,i} \\
=&\frac{\tau}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{N|\mathcal{T}_{(1)}|}}c_{\beta\gamma,i}'\left(c_{\beta\gamma}'c_{\beta\gamma}/N\right)^{-1}\frac{1}{\sqrt{N|\mathcal{T}_{(1)}|}}\sum_{t\in\mathcal{T}_{(1)}}\sum_{j=1}^{N}c_{\beta\gamma,j}e_{jt}\bar{G}_t\right)\left(\frac{1}{|\mathcal{T}_{(2)}|}\sum_{t\in\mathcal{T}_{(2)}}e_{it}\bar{G}_t\right)' \\
&+\frac{\tau}{N}\sum_{i=1}^{N}\left(\frac{1}{\sqrt{T|\mathcal{T}_{(1)}|}}\frac{1}{\sqrt{|\mathcal{T}_{(1)}|}}\sum_{t\in\mathcal{T}_{(1)}}\bar{G}_t u_{vz,t}'\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\left(u_{vz}'u_{vz}/T\right)^{-1}u_{vz,s}e_{is}+O_p(\frac{1}{m_{NT}})\right) \\
&\times\left(\frac{1}{|\mathcal{T}_{(2)}|}\sum_{t\in\mathcal{T}_{(2)}}e_{it}\bar{G}_t\right)' \\
=&O_p(\frac{1}{\sqrt{N}})+O_p(\frac{1}{\sqrt{\tau}})
\end{aligned}
\tag{171}
$$

Therefore,

$$\frac{1}{N}Q_{B_g,T}\widetilde{q}^{(1)\prime}_{G,T}\widetilde{q}^{(2)}_{G,T}Q_{B_g,T} = \frac{1}{N}\sum_{i=1}^{N}Q_{B_g,T}\widetilde{x}^{(1)}_{g,i}\widetilde{x}^{(2)\prime}_{g,i}Q_{B_g,T}$$
$$=\frac{1}{N}\sum_{i=1}^{N}Q_{B_g,T}\left(\widetilde{x}^{(1)}_{g,c\beta_g,i}+\widetilde{x}^{(1)}_{g,cc,i}+\widetilde{x}^{(1)}_{g,e,i}\right)\left(\widetilde{x}^{(2)}_{g,c\beta_g,i}+\widetilde{x}^{(2)}_{g,cc,i}+\widetilde{x}^{(2)}_{g,e,i}\right)'Q_{B_g,T}$$
$$=\frac{1}{N}Q_{B_g,T}\left(\sum_{i=1}^{N}\widetilde{x}^{(1)}_{g,c\beta_g,i}\widetilde{x}^{(2)\prime}_{g,c\beta_g,i}\right)Q_{B_g,T}+O_p(\frac{1}{\sqrt{\tau}}) = Q_g\eta_{c\beta_g}Q_g+O_p(\frac{1}{\sqrt{\tau}}) \tag{172}$$

(3)

$$\widetilde{\epsilon}^{(1)}_i = \left(\widetilde{x}^{(1)}_{g,e,i}+\widetilde{x}^{(1)}_{g,cc,i}\right)\widetilde{\theta} \tag{173}$$

Equations (169)-(171) imply that

$$\frac{\sqrt{\tau}}{\sqrt{N}}Q_{B_g,T}\sum_{i=1}^{N}\widetilde{x}^{(2)}_{g,i}\left(\widetilde{x}^{(1)}_{g,e,i}+\widetilde{x}^{(1)}_{g,cc,i}\right)'$$
$$=\frac{\sqrt{\tau}}{\sqrt{N}}Q_{B_g,T}\sum_{i=1}^{N}\left(\widetilde{x}^{(2)}_{g,c\beta_g,i}+\widetilde{x}^{(2)}_{g,cc,i}+\widetilde{x}^{(2)}_{g,e,i}\right)\left(\widetilde{x}^{(1)}_{g,e,i}+\widetilde{x}^{(1)}_{g,cc,i}\right)' = O_p(1) \tag{174}$$

which provided that $\|\theta_G\|_F \leq L$ then gives

$$\sqrt{\frac{\tau}{N}}Q_{B_g,T}\widetilde{q}^{(2)\prime}_{G,T}\widetilde{\epsilon}^{(1)} = O_p(1) \tag{175}$$

Finally

$$\sqrt{N\tau}Q^{-1}_{B_g,T}\left(\widetilde{\theta}^{(1)}_{g,I}-\widetilde{\theta}^{(1)}_{g}\right)$$
$$=\left(\frac{1}{N}Q_{B_g,T}\widetilde{q}^{(1)\prime}_{G,T}\widetilde{q}^{(2)}_{G,T}Q_{B_g,T}\left(\frac{1}{N}Q_{B_g,T}\widetilde{X}^{(2)\prime}_{g}\widetilde{q}^{(2)}_{G,T}Q_{B_g,T}\right)^{-1}\frac{1}{N}Q_{B_g,T}\widetilde{q}^{(2)\prime}_{G,T}\widetilde{q}^{(1)}_{G,T}Q_{B_g,T}\right)^{-1}$$
$$\times\frac{1}{N}Q_{B_g,T}\widetilde{q}^{(1)}_{G,T}\widetilde{q}^{(2)}_{G,T}Q_{B_g,T}\left(\frac{1}{N}Q_{B_g,T}\widetilde{X}^{(2)\prime}_{g}\widetilde{q}^{(2)}_{G,T}Q_{B_g,T}\right)^{-1}\sqrt{\frac{\tau}{N}}Q_{B_g,T}\widetilde{q}^{(2)\prime}_{G,T}\widetilde{\epsilon}^{(1)} = O_p(1), \tag{176}$$

which then leads to $\sqrt{NT}Q^{-1}_{B_g,T}\left(\widetilde{\theta}_G-\theta_G\right)\rightarrow O_p(1)$.

Next we propose the following $\widehat{\Sigma}_{\theta_G}$:

$$\widehat{\Sigma}_{\theta_G} = \frac{1}{2N}\sum_{i=1}^{2}\widehat{\Sigma}_{IV}^{(i)} + \frac{1}{T}\widehat{\Sigma}_{\widetilde{\theta}}$$

where $\widehat{\Sigma}_{\widetilde{\theta}}$ is a consistent estimator for the variance of $\widetilde{\theta}$ and

$$\widehat{\Sigma}_{IV}^{(1)} = \left(\widetilde{q}_{G,T}^{(1)\prime} P_{\widetilde{q}_{G,T}^{(2)}} \widetilde{q}_{G,T}^{(1)}/N\right)^{-1} \left(\sum_{i=1}^{N}\left(\left(\widetilde{q}_{G,T}^{(1)\prime} P_{\widetilde{q}_{G,T}^{(2)}}\right)_i \widehat{\widetilde{\epsilon}}_i^{(1)}\right)\left(\left(\widetilde{q}_{G,T}^{(1)\prime} P_{\widetilde{q}_{G,T}^{(2)}}\right)_i \widehat{\widetilde{\epsilon}}_i^{(1)}\right)'\right) \left(\widetilde{q}_{G,T}^{(1)\prime} P_{\widetilde{q}_{G,T}^{(2)}} \widetilde{q}_{G,T}^{(1)}/N\right)^{-1}$$
$$\widehat{\Sigma}_{IV}^{(2)} = \left(\widetilde{q}_{G,T}^{(2)\prime} P_{\widetilde{q}_{G,T}^{(1)}} \widetilde{q}_{G,T}^{(2)}/N\right)^{-1} \left(\sum_{i=1}^{N}\left(\left(\widetilde{q}_{G,T}^{(2)\prime} P_{\widetilde{q}_{G,T}^{(1)}}\right)_i \widehat{\widetilde{\epsilon}}_i^{(2)}\right)\left(\left(\widetilde{q}_{G,T}^{(2)\prime} P_{\widetilde{q}_{G,T}^{(1)}}\right)_i \widehat{\widetilde{\epsilon}}_i^{(2)}\right)'\right) \left(\widetilde{q}_{G,T}^{(2)\prime} P_{\widetilde{q}_{G,T}^{(1)}} \widetilde{q}_{G,T}^{(2)}/N\right)^{-1}$$

The validity of our proposed covariance estimator relies on some additional regular assumptions:

Assumption C.10. *We assume the following holds:*

(1)

$$\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}\sqrt{\tau}\widetilde{e}_{G,i}^{(j)} \\ \sqrt{\tau}\widetilde{e}_{G,i}^{(j)}\widetilde{e}_{G,i}^{(j^*)} \\ \widetilde{u}_{G,i}^{(j)} \\ \widetilde{u}_{e,i}^{(j)}\end{pmatrix} = \frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}\xi_{1,i,T} \\ \xi_{2,jj^*,i,T} \\ \eta_{1,i,T} \\ \eta_{2,i,T}\end{pmatrix} \to_d \begin{pmatrix}\xi_1 \\ \xi_{2,jj^*} \\ \eta_1 \\ \eta_2\end{pmatrix}$$

with $\widetilde{e}_{G,i}^{(j)} = \frac{1}{|\mathcal{T}_{(j)}|}\sum_{t\in\mathcal{T}_{(j)}} e_{it}\bar{G}_t'$, $\widetilde{u}_G^{(j)} = \frac{1}{|\mathcal{T}_{(j)}|}\sum_{t\in\mathcal{T}_{(j)}} u_{vz,it}\bar{G}_t'$, $\widetilde{u}_e^{(j)} = \frac{1}{|\mathcal{T}_{(j)}|}\sum_{t\in\mathcal{T}_{(j)}} u_{vz,it}e_t$.

(2) ξ_1, ξ_{2,jj^*}*, are independent from* η_1, η_2*, and* $\frac{1}{N}\sum_{i=1}^{N}\xi_{i,T}\xi_{i,T}' \to_p \Sigma_\xi$ *with* $\xi_i = ()$

Assumption C.10 can be relaxed if we assume for example $\sqrt{T}/N \to 0$ and then certain sampling errors would be negligible. Now we briefly discuss the validity of our proposed covariance estimator. From the above discussions, we know

$$\left(\frac{1}{N}Q_{B_g,T}\widetilde{q}_{G,T}^{(1)\prime}\widetilde{q}_{G,T}^{(2)}Q_{B_g,T}\left(\frac{1}{N}Q_{B_g,T}\widetilde{q}_{G,T}^{(2)\prime}\widetilde{q}_{G,T}^{(2)}Q_{B_g,T}\right)^{-1}\frac{1}{N}Q_{B_g,T}\widetilde{q}_{G,T}^{(2)\prime}\widetilde{q}_{G,T}^{(1)}Q_{B_g,T}\right)^{-1}$$
$$\times\frac{1}{N}Q_{B_g,T}\widetilde{q}_{G,T}^{(1)}\widetilde{q}_{G,T}^{(2)}Q_{B_g,T}\left(\frac{1}{N}Q_{B_g,T}\widetilde{q}_{G,T}^{(2)\prime}\widetilde{q}_{G,T}^{(2)}Q_{B_g,T}\right)^{-1} \to_p \Theta_{(1)} \tag{177}$$

with $\Theta_{(1)}$ a deterministic positive definite matrix. Then we only need to look at the term $\sqrt{\frac{\tau}{N}}Q_{B_g,T}\widetilde{q}_{G,T}^{(2)\prime}\widetilde{\epsilon}^{(1)}$.

Equations (169)-(171) imply that

Similar to the previous discussion, we have

$$\widetilde{x}_{g,c\beta_g,i}^{(2)} = \widehat{Q}_g^{(2)}\left(c_i, \beta_{g,i}'\right)', \quad \widetilde{x}_{g,e,i}^{(2)} = \frac{1}{\sqrt{|\mathcal{T}_{(2)}|}}\frac{1}{\sqrt{|\mathcal{T}_{(2)}|}}\sum_{t\in\mathcal{T}_{(2)}} e_{it}\bar{G}_t', \tag{178}$$

and equation (149) implies that

$$\begin{aligned}
\widetilde{c}_{it} - c_{it} &= \left(\widetilde{u}_{vz,t} - H'u_{vz,t}\right)' H^{-1}c_{\beta\gamma,i} + u_{vz,t}'H\left(\widetilde{c}_{\beta\gamma,i} - H^{-1}c_{\beta\gamma,i}\right) \\
&\quad + \left(\frac{1}{\sqrt{N}}V_{NT}^{-1}\frac{1}{T}\sum_{s=1}^{T}\left(\widetilde{u}_{vz,s}u_{vz,s}'\right)\frac{1}{\sqrt{N}}\sum_{i=1}^{N}c_{\beta\gamma,i}e_{it}\right)'\left(\frac{1}{\sqrt{T}}H'\frac{1}{\sqrt{T}}u_{vz}'e_i\right) + o_p(\frac{1}{m_{NT}}) \\
&= \left(\widetilde{u}_{vz,t} - H'u_{vz,t}\right)' H^{-1}c_{\beta\gamma,i} + u_{vz,t}'H\left(\widetilde{c}_{\beta\gamma,i} - H^{-1}c_{\beta\gamma,i}\right) \\
&\quad + \left(\frac{1}{T}u_{vz}'e_i\right)'\left(\frac{1}{\sqrt{N}}HH'\left(c_{\beta\gamma}'c_{\beta\gamma}/N\right)^{-1}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}c_{\beta\gamma,j}e_{jt}\right) + o_p(\frac{1}{m_{NT}}) \\
&= \left(\widetilde{u}_{vz,t} - H'u_{vz,t}\right)' H^{-1}c_{\beta\gamma,i} + u_{vz,t}'H\left(\widetilde{c}_{\beta\gamma,i} - H^{-1}c_{\beta\gamma,i}\right) \\
&\quad + \left(\frac{1}{T}u_{vz}'e_i\right)'\left(\frac{1}{\sqrt{N}}\left(u_{vz}'u_{vz}/T\right)^{-1}\left(c_{\beta\gamma}'c_{\beta\gamma}/N\right)^{-1}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}c_{\beta\gamma,j}e_{jt}\right) + o_p(\frac{1}{m_{NT}}) \\
&= \frac{1}{\sqrt{N}}c_{\beta\gamma,i}'\left(c_{\beta\gamma}'c_{\beta\gamma}/N\right)^{-1}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}c_{\beta\gamma,j}e_{jt} + \frac{1}{\sqrt{T}}u_{vz,t}'\left(u_{vz}'u_{vz}/T\right)^{-1}\frac{1}{\sqrt{T}}u_{vz}'e_i \\
&\quad + \left(\frac{1}{T}u_{vz}'e_i\right)'\left(\frac{1}{\sqrt{N}}\left(u_{vz}'u_{vz}/T\right)^{-1}\left(c_{\beta\gamma}'c_{\beta\gamma}/N\right)^{-1}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}c_{\beta\gamma,j}e_{jt}\right) + o_p(\frac{1}{m_{NT}}),
\end{aligned} \tag{179}$$

and thus

$$
\begin{aligned}
\widetilde{x}_{g,cc,i}^{(2)} =& \frac{1}{|\mathcal{T}_{(2)}|}\sum_{t\in\mathcal{T}_{(2)}}\frac{1}{\sqrt{N}}c'_{\beta\gamma,i}\left(c'_{\beta\gamma}c_{\beta\gamma}/N\right)^{-1}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}c_{\beta\gamma,j}e_{jt}\bar{G}_t \\
&+\frac{1}{|\mathcal{T}_{(2)}|}\sum_{t\in\mathcal{T}_{(2)}}\frac{1}{\sqrt{T}}u'_{vz,t}\left(u'_{vz}u_{vz}/T\right)^{-1}\frac{1}{\sqrt{T}}u'_{vz}e_i\bar{G}_t \\
&+\frac{1}{|\mathcal{T}_{(2)}|}\sum_{t\in\mathcal{T}_{(2)}}\left(\frac{1}{T}u'_{vz}e_i\right)'\frac{1}{\sqrt{N}}\left(u'_{vz}u_{vz}/T\right)^{-1}\left(c'_{\beta\gamma}c_{\beta\gamma}/N\right)^{-1}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}c_{\beta\gamma,j}e_{jt}\bar{G}_t+o_p(\frac{1}{m_{NT}}) \\
=&\frac{1}{\sqrt{N|\mathcal{T}_{(2)}|}}\left(\frac{1}{\sqrt{N|\mathcal{T}_{(2)}|}}\sum_{j=1}^{N}\sum_{t\in\mathcal{T}_{(2)}}c'_{\beta\gamma,i}\left(c'_{\beta\gamma}c_{\beta\gamma}/N\right)^{-1}c_{\beta\gamma,j}e_{jt}\bar{G}_t\right) \\
&+\frac{1}{\sqrt{T|\mathcal{T}_{(2)}|}}\left(\frac{1}{\sqrt{|\mathcal{T}_{(2)}|}}\sum_{t\in\mathcal{T}_{(2)}}\bar{G}_t u'_{vz,t}\right)\left(u'_{vz}u_{vz}/T\right)^{-1}\left(\frac{1}{\sqrt{T}}u'_{vz}e_i\right) \\
&+\frac{1}{\sqrt{|\mathcal{T}_{(2)}|}}\left(\frac{1}{T}u'_{vz}e_i\right)'\frac{1}{\sqrt{N}}\left(u'_{vz}u_{vz}/T\right)^{-1}\left(c'_{\beta\gamma}c_{\beta\gamma}/N\right)^{-1}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}c_{\beta\gamma,j}\left(\frac{1}{\sqrt{|\mathcal{T}_{(2)}|}}\sum_{t\in\mathcal{T}_{(2)}}e_{jt}\bar{G}_t\right) \\
&+o_p(\frac{1}{m_{NT}})=O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T})+O_p(\frac{1}{\sqrt{NT}})+o_p(\frac{1}{m_{NT}}) \qquad (180)
\end{aligned}
$$

Combined with previous discussions

$$\sqrt{\frac{\mathcal{T}}{N}}Q_{B_g,T}\widetilde{q}_{G,T}^{(2)\prime}\widetilde{\epsilon}^{(1)}\to_d H(\eta)\xi$$

with $H(.)$ a deterministic function. The limiting distribution is a mixed Gaussian distribution, and the rests are implied then by Assumption C.10.

□

Proof of 4.1.2. By construction of the HJN statistic, $\sqrt{T}q_{g,T,R}\left(\widetilde{\theta}_G-\theta_G\right)$ would be of the order $O_p(\frac{1}{N^{1/2}})$ when $N/T\to c$, and thus these sampling errors would be negligible. Thus the asymptotic distribution of the HJN statistic is determined by the distribution of the sample pricing errors $e_{T,R}(\theta_G)$. The consistency of $\widetilde{S}=\frac{1}{T}\sum_{t=1}^{T}e_{g,t,R}(\widetilde{\theta}_G)e_{g,t,R}(\widetilde{\theta}_G)'$ is implied by Theorem 4.1 and Assumptions 2.1 - 2.3. □