Jump Factor Models in Large Cross-Sections*

Jia Li,† Viktor Todorov,‡ and George Tauchen§

August 15, 2018

Abstract

We develop tests for deciding whether a large cross-section of asset prices obey an exact factor structure at the times of factor jumps. Such jump dependence is implied by standard linear factor models. Our inference is based on a panel of asset returns with asymptotically increasing cross-sectional dimension and sampling frequency, and essentially no restriction on the relative magnitude of these two dimensions of the panel. The test is formed from the high-frequency returns at the times when the risk factors are detected to have a jump. The test statistic is a cross-sectional average of a measure of discrepancy in the estimated jump factor loadings of the assets at consecutive jump times. Under the null hypothesis the discrepancy in the factor loadings is due to a measurement error, which shrinks with the increase of the sampling frequency, while under an alternative of a noisy jump factor model this discrepancy contains also non-vanishing firm-specific shocks. The limit behavior of the test under the null hypothesis is non-standard and reflects the strong-dependence in the cross-section of returns as well as their heteroskedasticity which is left unspecified. We further develop inference for assessing the magnitude of firm-specific risk in asset prices at the factor jump events. Empirical application to S&P 100 stocks provides evidence for exact one-factor structure at times of big market-wide jump events.

Keywords: factor model, panel, high-frequency data, jumps, semimartingale, specification test, stochastic volatility.

JEL classification: C51, C52, G12.

*Li's work was partially funded by NSF grant SES-1326819. Todorov's work was partially funded by NSF grant SES-1530748. We would like to thank a co-editor and three anonymous referees as well as seminar participants and discussants at various conferences for their thoughtful comments, which helped greatly improve the paper.

†Department of Economics, Duke University, Durham, NC 27708; e-mail: jl410@duke.edu.

‡Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208; e-mail: v-todorov@northwestern.edu.

§Department of Economics, Duke University, Durham, NC 27708; e-mail: george.tauchen@duke.edu.

1 Introduction

Asset prices often "jump" simultaneously in response to important market-wide events such as macroeconomic announcements, political news and natural disasters. Linear factor models which are widely used in asset pricing have strong implications regarding this co-jump structure of assets in the cross-section. To fix ideas, consider a continuous-time factor model given by

$$dY_t = \alpha_t dt + \mathbf{B} dF_t + d\widetilde{Y}_t, \tag{1.1}$$

where Y_t is a $N \times 1$ vector of asset prices, F_t is a $r \times 1$ vector of factors capturing systematic risk, α_t is a drift term due to the compensation for risk and time demanded by investors for holding the stocks, $\mathbf{B}$ is a $N \times r$ matrix of factor loadings, and $\widetilde{Y}_t$ is a $N \times 1$ vector of idiosyncratic risks. Idiosyncratic risk is formally defined as being orthogonal to the systematic one, that is, having zero quadratic covariation with it. Regarding the jumps, this orthogonality condition means $\Delta F_t \Delta \widetilde{Y}_t = 0$ for every t, where for a generic process Z we denote with $\Delta Z_t = Z_t - Z_{t-}$ its jump at time t.[1] Therefore, we have

$$\Delta Y_\tau = \mathbf{B} \Delta F_\tau, \qquad \text{for } \tau \text{ such that } \Delta F_\tau \neq 0. \tag{1.2}$$

That is, at the times of the factor jumps, the whole cross-section of asset prices follows an exact factor model with no idiosyncratic risk, i.e., $\widetilde{Y}$ is not present in (1.2).[2]

Moreover, if the jumps in the components of the vector F are normalized such that they are orthogonal to each other, i.e., have zero quadratic covariation, then we also have

$$\Delta Y_\tau = b \Delta Z_\tau, \qquad \text{for } \tau \in \mathcal{T} \equiv \{t \in [0,T] : \Delta Z_t \neq 0\}, \tag{1.3}$$

where Z_t is one of the elements of the vector F_t and b is the corresponding vector of factor loadings in the matrix $\mathbf{B}$.[3] That is, we have an exact one-factor structure of the jumps in Y at the jump times of the univariate factor Z_t. This one-factor structure of jumps in Y holds true for the jump times of each of the factors in F, but with different factor loadings in general.

[1] In many applications, orthogonality of systematic and idiosyncratic risk is defined in a much stronger sense by imposing independence between F and $\widetilde{Y}$.

[2] Put differently, if τ_1, τ_2,...,τ_k are the jump times of the systematic factors in the observation interval $[0,T]$ (these times are in general random) and if without loss of generality we assume $k > r$ and $N > r$, then the matrix $[\Delta Y_{\tau_1}, \Delta Y_{\tau_2}, ..., \Delta Y_{\tau_k}]$ is of reduced-rank r and the same holds true for the quadratic variation of Y at the factor jump times $\sum_{i=1}^k \Delta Y_{\tau_i} \Delta Y_{\tau_i}^\top$. Only when we aggregate the quadratic variation of *all* jump risk in Y over the interval $[0,T]$, given by $\sum_{t\in[0,T]} \Delta Y_t \Delta Y_t^\top$, we restore the full rank of N. This is because $\sum_{t\in[0,T]} \Delta Y_t \Delta Y_t^\top$ includes also the quadratic variation due to the jumps in the idiosyncratic component $\widetilde{Y}$ (which happen outside the set $\mathcal{T}$ of factor jump times) which is a diagonal matrix.

[3] Note that b is a $N \times 1$ vector that contains the slope coefficients (i.e., betas) for the N assets. There is no restriction on the relationship among individual assets' betas.

Our goal in this paper is to test whether a hypothesis of exact factor structure for the systematic jumps in (1.3) is true.[4] The test is based on a large cross-section of assets sampled at high-frequency on a time interval of fixed length. In its most basic form, which will be generalized in our formal analysis, our test discriminates (1.3) from its "noisy" counterpart given by

$$\Delta Y_\tau = b\Delta Z_\tau + \bar{\chi}_\tau, \qquad \text{for } \tau \in \mathcal{T}, \tag{1.4}$$

where $\bar{\chi}_\tau$ is a random $N \times 1$ vector that captures deviations from the exact factor model at the systematic jump events; we refer to (1.4) as the noisy linear factor model henceforth. This model can be equivalently written in the form of a random jump beta model as

$$\beta_\tau \equiv \frac{\Delta Y_\tau}{\Delta Z_\tau} = b + \tilde{\chi}_\tau, \tag{1.5}$$

where $\tilde{\chi}_\tau \equiv \bar{\chi}_\tau / \Delta Z_\tau$.[5] There are many reasons for the presence of $\tilde{\chi}_\tau$ in the jump factor loadings (i.e., betas). One of them is predictable time-variations in these loadings. Indeed, factor loadings in asset pricing are often modeled as functions of assets' characteristics and/or macro state variables; see, for example, Connor, Hagmann, and Linton (2012) and Gagliardini, Ossola, and Scaillet (2016b) and the theoretical analysis of Hansen and Richard (1987). Typically, such time-variation in the factor loadings happens at low frequencies and by performing our tests on intervals of short time span we will minimize the possibility for such a violation of the null hypothesis in (1.3). Another reason for $\bar{\chi}_\tau$ in (1.4) is the presence of (locally unpredictable) firm-specific shocks in Y even at the times of systematic jumps. In this case, we have $\mathbb{E}_{\tau-}[\bar{\chi}_\tau \Delta Z_\tau] = 0$, i.e., $\bar{\chi}_\tau$ is uncorrelated with the jump ΔZ_τ conditional on the information prior to the jump time τ, and (1.3) gets replaced by the weaker linear projection condition

$$\mathbb{E}_{\tau-}\left[\Delta Y_\tau \Delta Z_\tau\right] = b\mathbb{E}_{\tau-}\left[\Delta Z_\tau^2\right]. \tag{1.6}$$

Obviously, (1.3) implies (1.6) but the reverse is generally not true. The test we propose here will be able to discriminate (1.3) from alternatives under which we have firm-specific shocks at the systematic jump events and only (1.6) holds.

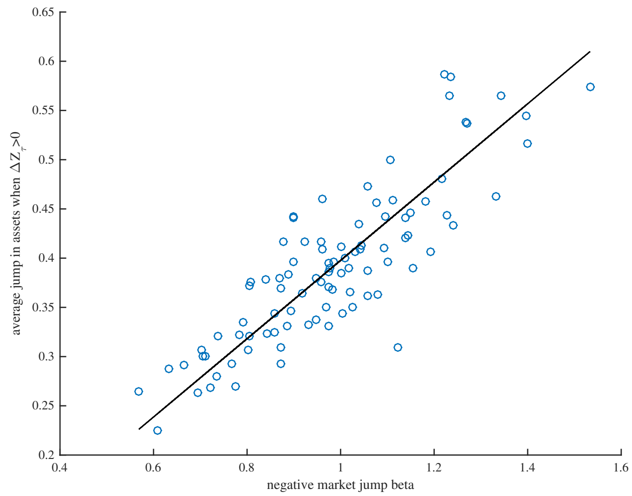
From a practical point of view, separating the null from such an alternative hypothesis is important as this has strong implications regarding the inference for the matrix of jump factor loadings $\mathbf{B}$. Indeed, if the null hypothesis holds, we can use a fixed (small) number of systematic jump times to identify the (constant) jump factor loadings. On the other hand, if only (1.6) holds,

[4] When the factors in the vector F do not have orthogonal jump components, then for $\tau \in \mathcal{T}$, we will have in general an exact multifactor model. The analysis of the paper can be extended to cover this more general setup but at the cost of significantly more involved derivations and notation. Given the empirical evidence presented in Section 6, however, such an extension seems to be of little practical relevance.

[5] The random jump betas may be viewed as "random coefficients" like in classical panel data analysis. However, we note that the jump betas only concern the jump returns, which form a small subset of all high-frequency returns.

then we need to use a long span of data to identify the permanent component of $\mathbf{B}$, which would be further complicated by the possible time-variation in the factor loadings at low frequencies. The separation of the null from the alternative hypothesis is also important for practical risk management decisions and more generally can help shed light on the sources of priced jump risk in the cross-section of asset prices.

Figure 1: Stock Jump Returns versus Jump Beta



Note: The vertical axis is the time series average two-minute stock return at times when a positive jump in the S&P 500 index futures is detected; the horizontal axis is the average jump beta measured from two-minute asset returns at times when a negative jump in the index futures is detected. The straight line is a linear fit implied by the exact linear jump factor model in (1.3) and the recovered jump betas. The panel comprises S&P 100 stocks, 2007-2015, and the market jump returns are selected according the thresholding procedure described in Section 5.

To motivate empirically our theoretical analysis, we show in Figure 1 the cross-sectional relationship between stock returns versus their market jump betas for S&P 100 stocks: the two-minute average stock returns are computed at the times when a positive market jump is detected (i.e., $\Delta Z_\tau > 0$), and the jump betas are calculated from the two-minute asset returns at the times when a negative market jump is detected (i.e., $\Delta Z_\tau < 0$). We use negative jump betas because the estimation error is relatively lower for negative jumps, while we measure average stock returns at times of positive market jumps to maintain complete separation of the data sets for computing the variables on the x and y axes of the figure. This cross-sectional relationship appears quite tight, even though the market betas were computed from only 44 two-minute returns when negative market jumps were detected in our sample — hence the question: Is this fit consistent with the exact linear jump factor model (1.3)?

To address this question, we develop a formal test for deciding whether such hypothesis is true and we further propose a measure for the magnitude of the firm-specific risk at factor jump events. Our asymptotic theory employs a joint asymptotic setting in which both the cross-sectional dimension of the panel and the sampling frequency increase simultaneously, while keeping the time span of the sample fixed. To compute the test statistic, we first use high-frequency returns to nonparametrically estimate the individual assets' betas at the jump times of Z; the test statistic is then formed as the cross-sectional average of the temporal variations in these beta estimates. Under the null hypothesis, changes in the beta estimates across jump times are due to high-frequency estimation error that shrinks to zero asymptotically, but is non-degenerate under the alternative. We characterize theoretically the asymptotic behavior of the test statistic and propose an easy-to-implement algorithm for computing its critical values.

The asymptotic theory underlying our test is non-standard for several reasons. Firstly, we allow individual assets' returns to be *strongly* dependent in the cross-section, through their loadings on the common diffusive factors. Importantly, these systematic shocks are not "averaged out" under the cross-sectional aggregation (only the loadings on them are), but they remain to have non-degenerate distributions in the limit. As a result, unlike conventional econometric settings, the limiting behavior of the aggregated systematic shocks is not obtained from a central limit theory for weakly dependent data; instead, it is implied by the local Gaussianity of diffusive processes. Secondly, the limit distribution of our test statistic depends on the spot covariance process of the *latent* diffusive factors and individual assets' time-varying loadings on them within local windows around jump times. Therefore, our feasible inference on the panel of jumps actually involves estimating characteristics of the diffusive factor component as a by-product. We provide novel theoretical results in this direction, which are further used to construct critical values of our test. Thirdly, the limit behavior of our test statistic is of "mixing" type. More precisely, the limit distribution can be realized on an extension of the original probability space and changes (in general) depending on the realizations of various sources of randomness on the original space, e.g., it depends on the realized value of the systematic risk factors. Finally, we impose essentially no assumption on the relative growth rates of the cross-section dimension with respect to the sampling frequency. This accommodates situations in which either the cross-section or the sampling frequency is much higher than the other dimension of our panel and is very convenient empirically.

Going one step further, we develop an estimator for assessing the magnitude of the firm-specific risk at the times of systematic jump events. In particular, we measure the cross-sectional mean of the squared difference $(\tilde{\chi}_{j,\tau} - \tilde{\chi}_{j,\eta})^2$ for two jump times τ and η, where $\tilde{\chi}_{j,\tau}$ denotes the jth element of $\tilde{\chi}_\tau$. This measure provides an estimate of the temporal variation in jump betas for a "representative" asset. We show that the asymptotic distribution of this estimator is doubly-mixed Gaussian. The limit distribution captures two sources of sampling errors. One is the cross-sectional sampling variability in $\tilde{\chi}_{j,\tau}$ and its magnitude depends on the cross-sectional dimension. The other source of sampling error is due to the systematic diffusive component underlying individual asset returns and its size is governed by the length of the sampling interval. We provide confidence intervals and justify their asymptotic validity.

We examine the performance of our inference procedures on simulated data from a model that captures salient features of typical financial data sets. In an empirical application to high-frequency data on stocks in the S&P 100 index covering the period 2007–2015, we test the exact jump factor model in (1.3) with the sole jump factor being the market jump. Our results provide strong empirical support for the model with no statistically significant role for firm-specific shocks in the assets at the market-wide extreme events.[6] These findings are consistent with the empirical results in Savor and Wilson (2014) regarding the validity of CAPM around macroeconomic announcement days as well as earlier empirical evidence, see e.g., Longin and Solnik (2001) and Ang and Chen (2002), for increased asset correlation during extreme market events.

The theoretical results of the current paper are related to several strands of literature. First, the nonparametric separation of jumps from diffusive volatility using high-frequency data was initiated in work by Barndorff-Nielsen and Shephard (2004, 2006) and Mancini (2001, 2009).

Second, in Li, Todorov, and Tauchen (2017), we developed univariate methods for estimating and testing the validity of the exact jump factor model for a single asset, while being silent about the co-jump behavior of the cross-section of assets. We further extended these results to a cross-section of *fixed* size in Li, Todorov, Tauchen, and Lin (2017). The asymptotics in these papers is of in-fill type for high-frequency data. By sharp contrast, here we consider a *joint* asymptotic setup in which the number of assets also grows asymptotically so as to accommodate the large cross-section. Relative to prior work, a key theoretical challenge in the current paper is the strong dependence among a large cross-section of asset returns resulting from the latent factor structure of the diffusive price components. We address this problem by developing factor-analytical tools in the spirit of Stock and Watson (2002) and Bai (2003, 2009) (see also the recent work of Gagliardini, Ossola, and Scaillet (2016a)), but in the non-standard (infill, non-ergodic and non-stationary) setting for a large panel of high-frequency data. This theoretical innovation is absent from Li, Todorov, and Tauchen (2017) and Li, Todorov, Tauchen, and Lin (2017). The novel econometric setup of the current paper thus leads to a very different inference procedure and asymptotic theory than prior work, reflected in particular in the distinct roles of the systematic and idiosyncratic diffusive shocks in the limiting behavior of the test. In addition, unlike our prior work, the current setup allows us to study how the whole cross-section of assets reacts to economy-wide events, and further make inference for the firm-specific risk at the systematic factor jump times in the case when the null

[6] We note, however, that our evidence is for the "big" jumps only as separating the "small" jumps from the diffusive component of returns is statistically much harder.

hypothesis of an exact jump factor model is not satisfied.

Third, Pelger (2015a,b) develops methods for determining the number of systematic jump factors and further proposes inference procedures for recovering the latent systematic jump factors from a large cross-section of high-frequency return data. The inference in Pelger (2015a,b) is based on the quadratic variation of the jumps of the assets over the observation interval. The latter includes the contribution of the idiosyncratic jump risks and, hence, it cannot be used to separate between the null in (1.3) and alternatives that satisfy (1.6).

Fourth, our results are also related to the panel data literature but with very distinctive features. Unlike conventional panel regressions, the identification strategy underlying our inference for jump betas can be viewed as "identification-by-discontinuity."[7] An important consequence is that we can use high-frequency data at a *fixed* number of jump events to estimate firm-specific jump betas nonparametrically. The "error terms" generated from this nonparametric measurement consist of diffusive returns in the vicinity of jumps, which shrink with the sampling interval, and the asymptotics of our statistics is mainly driven by their local Gaussianity (which is a generic property of Itô processes). Our in-fill asymptotic theory works under very weak conditions allowing for essentially unrestricted non-stationarity and heteroskedasticity for the underlying stochastic processes in a non-ergodic setting. The theory in the current paper is thus very different from conventional panel data settings where weakly dependent disturbance terms in the regression equation are "averaged out" with a Gaussian limit distribution, which in turn drives the asymptotics underlying the econometric inference.

Finally, since we allow the diffusive "error terms" to have a latent factor structure, our results are closely related to the growing literature in panel data analysis with interactive fixed effects; see, for example, Pesaran (2006) and Bai (2009). Besides the very different settings mentioned above, we stress an important unique feature in our asymptotic theory relative to this literature. Unlike Bai (2009) where the asymptotic distribution of the estimator is only driven by the aggregated idiosyncratic shocks,[8] here in our setting systematic shocks in the error term also have a non-degenerate contribution to the limit distribution. Intuitively, the systematic shocks are clearly not "averaged out" in the cross-section and, because we consider a fixed number of jump events (as they are rare), they also "survive" the time-series aggregation. Nevertheless, we show that feasible inference is still possible by relying on the local Gaussianity of the underlying Itô process, formally represented as a stable convergence in law under the in-fill asymptotics. In addition, we allow the latent diffusive factors to have general stochastic volatility and the stocks' loadings on

[7] Identification-by-discontinuity is also widely used in microeconometric applications with regression discontinuity designs (RDD); see, for example, Lee and Lemieux (2010) for a review. However, we use jumps to identify the equilibrium relationship between asset prices, instead of a causal effect as is typically done in the RDD.

[8] See Proposition A.3 of Bai (2009), which is used in the proofs of the main theorems in that paper.

them to be (nonparametrically) time-varying; indeed, we allow these processes to jump, which leads to an additional source of "non-diversifiable" sampling variability due to the indeterminacy of exact systematic jump times within a discrete observation interval. Another useful feature of our theory is that it does not require knowing exactly the number of latent diffusive factors; this feature is clearly desirable in practice and has been recently studied by Moon and Weidner (2015) in standard linear panel regressions with interactive fixed effects using different techniques.

The rest of this paper is organized as follows. In Section 2 we introduce the econometric setting. Our main theoretical results are given in Section 3 where we propose a pairwise test for the exact jump factor model and derive feasible limit theory for it. In this section we also develop inference procedures for assessing the magnitude of the firm-specific risk in assets at the times of the systematic jump events. Section 4 proposes various extensions to our main theory. Section 5 contains results from a Monte Carlo study and Section 6 presents our empirical application, with additional results reported in a Supplemental Appendix. Section 7 concludes. Further theoretical results and the proofs are given in Section 8.

2 The Econometric Setup and Assumptions

2.1 The Panel of Jump Betas

We start with introducing formally the asset price dynamics on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. The dynamics for the vector of asset prices Y_t, to which our analysis applies, generalizes the linear factor model in (1.1) and is given by:

$$dY_{j,t} = \alpha_{j,t}dt + \lambda_{j,t}^{\top}df_t + dJ_{Y,j,t} + d\epsilon_{j,t}, \quad 1 \leq j \leq N, \tag{2.1}$$

where as in the introduction $\alpha_{j,t}$ is a drift term in the asset price. Here, we denote with f_t a r-dimensional diffusive factor process and with $\lambda_{j,t}$ the time-varying loading on it, and $J_{Y,j,t}$ denotes the jump component of $Y_{j,t}$. Finally, $\epsilon_{j,t}$ is a one-dimensional diffusive idiosyncratic component that is orthogonal to the systematic factor f_t (in the sense that it has zero quadratic covariation with f_t; see Assumption 3 below).

The main object of interest of the current paper is the factor structure of the jump component $J_{Y,j,t}$. The diffusive component $\int_0^t \lambda_{j,s}^{\top} df_s + \epsilon_{j,t}$, on the other hand, plays the role of a "disturbance" as in classical regression settings.[9] We stress that these "disturbance" terms in the asset prices are allowed to be strongly dependent in the cross-section through their loadings on the common

[9] The diffusive component of returns can have different factor structure and factor loadings from that of the jump component, and this can have asset pricing implications, see e.g., Bollerslev, Li, and Todorov (2016) for the case when the factor is the return on the market portfolio. We leave for future work the development of tests for deciding whether diffusive and jump factor loadings are the same and efficient inference techniques under such a scenario.

factors. This has important implications for the asymptotic theory and is a non-trivial departure from classical panel data applications in econometrics.

Turning to the jump component $J_{Y,j,t}$, we will be interested in its behavior only at the jump times of an one-dimensional reference asset (systematic factor) which we denote with Z. In our application, Z will be the market portfolio and hence our attention will be on the co-jump behavior of the cross-section of assets during market-wide events, located at times of "big" market jumps. Economically speaking, such jumps are mostly due to important public news arrival, such as macroeconomic announcements, major political events and natural disasters. In addition to co-jumping with Z, we will also allow Y to jump at different times. As the model in the introduction, this can allow for asset prices to have idiosyncratic jumps as in Merton (1976) (in the sense that these jumps are asset-specific and arrive at different times) as well as to have exposure to systematic jump factors which jump at different times than Z.

The dynamics of the reference asset Z is given by

$$dZ_t = \lambda_{Z,t}^\top df_t + dJ_{Z,t}, \tag{2.2}$$

with $J_{Z,t}$ denoting its jump component. When Z is the market portfolio, it will typically be one of the factors itself as in the market factor model and its extensions. Hence, in this case, one can set the first element of f to be the continuous part of Z and $\lambda_{Z,t}$ will be a vector with the first entry being 1 and the rest being zero.

We collect the jump times of Z in the random set $\mathcal{T} \equiv \{t \in [0,T] : |\Delta Z_t| \neq 0\}$. The relationship between the asset price jumps and those of the one-dimensional Z is succinctly summarized by the jump betas of individual assets with respect to Z. At each jump time $\tau \in \mathcal{T}$, the *spot jump beta* of asset j is simply defined by

$$\beta_{j,\tau} \equiv \frac{\Delta Y_{j,\tau}}{\Delta Z_\tau}.$$

We note that these betas are defined in a nonparametric fashion and in general they are random quantities. However, when a factor model as noted in the introduction is in force, the above spot jump betas will be constant. That is, we will have

$$\beta_{j,\tau} = \beta_j \quad \Longleftrightarrow \quad \Delta Y_{j,\tau} = \beta_j \Delta Z_\tau \text{ for all } \tau \in \mathcal{T}.$$

We note that for the above to be true, we do not require that Z is the only jump factor that affects the asset prices in the vector Y. This is because the above exact one-factor relationship *only* concerns the set of jump times $\mathcal{T}$. Jumps of Y at times outside the set $\mathcal{T}$ do not need to obey this exact factor structure.

In this paper, we are interested in testing whether the above constant beta restriction holds *jointly* for a large cross-section of assets (N will be asymptotically increasing). This is clearly a very strong hypothesis, especially when $\mathcal{T}$ involves many jump times. Hence, we start with the shortest possible event horizon by focusing on pairs of consecutive jump times of the systematic factor Z. That is, we examine, for the successful jump times $\eta, \tau \in \mathcal{T}$ with $\eta < \tau$,

$$\beta_{j,s} = \beta_j, \quad s \in \{\eta, \tau\} \quad \text{all } 1 \leq j \leq N. \tag{2.3}$$

The extension to the more general case with multiple jump times is discussed in Section 4.1.

The hypothesis of the constant jump beta in (2.3) can be nested within a more general alternative of the random jump beta model given by

$$\beta_{j,s} = \beta_j + \tilde{\chi}_{j,s}, \quad 1 \leq j \leq N. \tag{2.4}$$

The constant beta model (2.3) is a special case of (2.4) with $\tilde{\chi}_{j,s} = 0$ identically. The random term $\tilde{\chi}_{j,s}$ captures the possible violation of an exact linear one-factor model for the jumps $\{\Delta Y_{j,s}\}_{s\in\{\eta,\tau\}}$. In particular, if there are firm-specific shocks at the times when Z jumps, then this risk will be reflected in $\tilde{\chi}_{j,s}$. Our test will be able to discriminate this alternative from the exact jump factor model in (2.3).

We stress the importance of separating the null from the alternative hypothesis on a practical level. If the null hypothesis is true, then one can in principle estimate consistently the jump beta from a single jump event or, more generally from a fixed number of jump events. Intuitively, by zooming into the short window around the big jump of Z, we obtain high signal-to-noise measurement of the shocks to the underlying efficient prices. This gives rise to a type of "identification-by-discontinuity." On the other hand, if $\tilde{\chi}_{j,s}$ in (2.4) is non-degenerate, then we cannot estimate consistently the permanent component β_j from a fixed number of jump events. This is because $\tilde{\chi}_{j,s}$ is $O_p(1)$ and, hence, one would need to appeal to long-span asymptotics in this case in order to "average out" $\tilde{\chi}_{j,s}$ (this is exactly as in the first-step of the classical Fama-MacBeth regression that is widely used in cross-sectional asset pricing).

By focusing on two jump times, the structure of (2.4) resembles that of the "large N small T" setting in microeconometric panel data analysis. In particular, the time-invariant coefficient β_j plays the role of a fixed effect (or the permanent component) and $\tilde{\chi}_{j,s}$ plays the role of a random shock (or the transitory component) in the spot jump betas.

The challenge in our analysis is that the jump betas $(\beta_{j,\eta}, \beta_{j,\tau})$ are not directly observed from discrete return data. We will use data sampled at high frequencies to recover them nonparametrically. The econometric problem at hand may thus be classified as one with large cross-sectional dimension, short and fixed time span, and with nonparametrically generated dependent variables (i.e., the high-frequency beta estimates). In addition, due to the factor structure in the diffusive returns, the estimation errors for these betas are *strongly* cross-sectionally dependent with nonstandard asymptotic distribution, which further sets our study apart from prior work in microeconometric panels.

We now describe our test statistic. If the jump betas were observed, we could use the following (infeasible) statistic to detect deviations from the constant beta restriction:

$$V_N^{\star} \equiv \frac{1}{N} \sum_{j=1}^{N} L\left(|\beta_{j,\tau} - \beta_{j,\eta}|\right)$$

where $L(\cdot)$ is a loss function that we use to gauge the temporal variation in the jump betas. In particular, if the innovation terms $\tilde{\chi}_{j,\tau} - \tilde{\chi}_{j,\eta}$ are non-degenerate, we expect $V_N^{\star}$ to converge to the cross-sectional mean of the loss $L\left(|\beta_{j,\tau} - \beta_{j,\eta}|\right)$, which summarizes the temporal variation in beta for any "average" stock.

In this paper, we consider loss functions that are "quadratic near zero." More precisely, we maintain the following assumption for the loss function.

Assumption 1. *The function $L : \mathbb{R} \mapsto [0, \infty)$ satisfies the following: (i) for some fixed $\bar{x} \in [0, \infty]$, $L(x) = x^2$ for all $x \in [0, \bar{x})$; (ii) $L(x)$ is non-decreasing in $|x|$; (iii) $L(\cdot)$ is Lipschitz on bounded sets.*

Assumption 1 ensures that $L(\cdot)$ behaves for our testing purposes like the quadratic loss function under the null hypothesis. On the other hand, the finite-sample power of the test naturally depends on the global shape of the loss function. In this way, Assumption 1 provides some flexibility in "directing" the power of the test. For example, one may consider a "Huber-like" loss function $L(x) = \min\{|x|, x^2\}$ that is less sensitive to outliers than the quadratic one (Huber (2004)).

Our feasible test statistic is constructed as a sample analogue of $V_N^{\star}$ by replacing the spot betas with their high-frequency estimates. We suppose that the processes Y_j and Z are observed at discrete times $\{i\Delta_n : 0 \leq i \leq [T/\Delta_n]\}$, where Δ_n is the sampling interval. For the limit theory that we develop below, we consider an asymptotic setting in which the sample span T is fixed, but both $\Delta_n \to 0$ and $N \to \infty$ jointly. Since the number of assets N also depends on the asymptotic stage n, we write N_n to emphasize this dependency. Below, we index all estimators by n and all limits are for $n \to \infty$.

For each $\tau \in \mathcal{T}$, we denote by $i(n, \tau)$ the unique integer i such that $\tau \in ((i-1)\Delta_n, i\Delta_n]$. We then collect these indices using $\mathcal{I}_n = \{i(n, \tau) : \tau \in \mathcal{T}\}$, which is finite almost surely. This set can be consistently recovered by

$$\widehat{\mathcal{I}}_n \equiv \{i : |\Delta_i^n Z| > u_n\},$$

where the truncation threshold satisfies $u_n \asymp \Delta_n^{\varpi}$ for some $\varpi \in (0, 1/2)$.[10] In light of this jump detection result, we can assume that $i(n, \tau)$ is observed for each $\tau \in \mathcal{T}$ without loss of generality.

[10] It can be shown that $\widehat{\mathcal{I}}_n = \mathcal{I}_n$ with probability approaching one; see Proposition 1 in Li, Todorov, and Tauchen (2017).

The spot jump beta estimator associated with each jump time τ is then given by,

$$\hat{\beta}_{n,j,\tau} = \frac{\Delta^n_{i(n,\tau)} Y_j}{\Delta^n_{i(n,\tau)} Z}, \quad 1 \le j \le N_n.$$

The corresponding feasible test statistic can be constructed naturally as

$$\hat{V}^*_n \equiv \frac{1}{N_n} \sum_{j=1}^{N_n} L\left(\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}\right). \tag{2.5}$$

Under the null hypothesis of (2.3) and in the absence of idiosyncratic jumps in Y_j, the leading term of $\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}$ is due to the diffusive component in $Y_j - \beta_j Z$ in the high-frequency increments containing the two jumps. Therefore, this difference is of order $O_p(\sqrt{\Delta_n})$ under the null, and thus $\hat{V}^*_n$ should be $O_p(\Delta_n)$. On the other hand, under the alternative hypothesis, $\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}$ will contain also the difference $\tilde{\chi}_{j,\tau} - \tilde{\chi}_{j,\eta}$ which is $O_p(1)$ and hence so is $\hat{V}^*_n$. This explains on an intuitive level how $\hat{V}^*_n$ can discriminate between the exact and noisy linear jump factor model. Below, we develop formal statistical tests based on this intuition.

The situation becomes more complicated when individual assets also contain idiosyncratic jumps. Although the idiosyncratic jumps of each asset do not occur exactly at systematic jump times, there is a small probability that these jumps occur in the same high-frequency sampling interval containing the jump in Z. This effect would be asymptotically negligible if the number of assets were fixed, but this will generally not be the case in the current setting in which the number of assets goes to infinity. This issue can be addressed via winsorization. Formally, we consider a sequence $q^w_n \to 0$ which specifies the proportion of data to be winsorized. We denote by $\bar{B}_{n,\eta,\tau}$ the $1 - q^w_n$ quantile of the variables $(|\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}|)_{1 \le j \le N_n}$. The winsorized test statistic is then

$$\hat{V}_n = \frac{1}{N_n} \sum_{j=1}^{N_n} L\left(|\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}| \wedge \bar{B}_{n,\eta,\tau}\right). \tag{2.6}$$

Below, we focus on the asymptotic properties of the above winsorized statistic, while noting that the unwinsorized statistic can be used equivalently if the jump arrivals in Y outside of the set $\mathcal{T}$ are driven by a finite counting measure.[11] We maintain the following condition for the winsorization quantile q^w_n.

Assumption 2. $q^w_n \asymp \Delta^\kappa_n$ *for some constant* $\kappa \in (0, 1)$.

2.2 Regularity Conditions

We proceed with some regularity conditions that will be used throughout. We start with some standard conditions regarding the pathwise regularities of the underlying processes.

[11] That is, when the total number of idiosyncratic jump times in all stocks is finite almost surely. Technical details are available upon request.

Assumption 3. *The processes $(Y_j)_{1\leq j\leq N_n}$ and Z are given by (2.1) and (2.2) such that the following conditions hold.*

(i) The jump processes J_Z and $(J_{Y,j})_{1\leq j\leq N_n}$ have the form

$$
\begin{aligned}
J_{Z,t} &= \int_0^t \int_E \delta_Z(s,u)\mu\left(ds,du\right), &(2.7)\\
J_{Y,j,t} &= \int_0^t \int_E \delta_{Y,j}(s,u)\mu\left(ds,du\right) + \int_0^t \int_E \tilde{\delta}_{Y,j}(s,u)\tilde{\mu}_j\left(ds,du\right), &(2.8)
\end{aligned}
$$

where μ is a Poisson random measure on $\mathbb{R}_+\times E$ with compensator $\nu(ds,du) = ds\otimes v'(du)$ for some finite measure v' on a Polish space E; $(\tilde{\mu}_j)_{1\leq j\leq N_n}$ are Poisson random measures that satisfy the same conditions as μ; and the jump size functions δ_Z, $(\delta_{Y,j})_{j\geq 1}$ and $(\tilde{\delta}_{Y,j})_{j\geq 1}$, which are mappings $\Omega\times\mathbb{R}_+\times E\mapsto\mathbb{R}$, are predictable. Moreover, the jump processes $\tilde{J}_{Y,j,s}\equiv\int_0^t\int_E\tilde{\delta}_{Y,j}(s,u)\tilde{\mu}_j\left(ds,du\right)$ are uniformly locally bounded and satisfy $\Delta\tilde{J}_{Y,j,s}\Delta J_{Z,s}=0$ almost surely for $s\in[0,T]$.

(ii) The diffusive factor process f is a r-dimensional continuous Itô semimartingale of the form:

$$
f_t = \int_0^t b_{f,s}ds + \int_0^t \sigma_{f,s}dW_s,
$$

where the processes b_f and σ_f are locally bounded and W is a r-dimensional Brownian motion. Moreover, the spot covariance process $\Sigma_f = \sigma_f\sigma_f^\top$ is non-singular almost surely.

(iii) For each j, ϵ_j is a one-dimensional continuous local martingale given by

$$
\epsilon_{j,t} = \int_0^t \tilde{\sigma}_{j,s}d\widetilde{W}_{j,s},
$$

where $(\widetilde{W}_j)_{1\leq j\leq N_n}$ are one-dimensional Brownian motions that are orthogonal to W and the processes $(\tilde{\sigma}_j)_{1\leq j\leq N_n}$ are locally uniformly bounded.[12]

(iv) The processes α_j, λ_j and λ_Z are locally bounded and the spot jump betas $(\beta_{j,\tau})_{\tau\in\mathcal{T}}$ are bounded, uniformly for $1\leq j\leq N_n$.

(v) For an increasing sequence $(T_m)_{m\geq 1}$ of stopping times that goes to infinity and a sequence $(K_m)_{m\geq 1}$ of constants,

$$
\mathbb{E}\left[\sup_{s,t\in[0,T\wedge T_m]}\left(\left|\tilde{\sigma}_{j,s}^2-\tilde{\sigma}_{j,t}^2\right|^2+\left|\lambda_{j,s}-\lambda_{j,t}\right|^2+\left|\lambda_{Z,s}-\lambda_{Z,t}\right|^2\right)\right]\leq K_m\left|s-t\right|,
$$

uniformly in $1\leq j\leq N_n$.

Assumption 3 imposes a set of regularity conditions that are commonly used in the analysis of high-frequency data. Conditions (i)–(iii) mainly require that the price processes are Itô semimartingales. We allow for leverage effect, price-volatility co-jumps, as well as co-jumps between the

[12]That is, for a sequence $(T_m)_{m\geq 1}$ of stopping times that increases to infinity, the processes $\tilde{\sigma}_j$, $1\leq j\leq N_n$, are uniformly bounded on $[0,T\wedge T_m]$.

factor volatility and idiosyncratic volatility. We further allow for idiosyncratic jump risk in Y and multiple systematic jump factors (in addition to Z) in the jump component $\tilde{J}_{Y,j,t}$. Condition (iv) requires (local) boundedness for the factor loadings and the jump betas, which is reasonable from an empirical point of view. Condition (v) requires that the idiosyncratic volatility and the factor loading processes are (1/2)-Hölder continuous under the L_2-norm and locally in time. Note that these processes are allowed to have jumps in their sample paths with arbitrary activity. This condition holds for many stochastic processes such as Itô semimartingales and long-memory processes driven by the fractional Brownian motion.

We note that Assumption 3 imposes a finite activity restriction on the jump measures μ and $\tilde{\mu}_j$. Therefore, our analysis here applies for the "big" jumps in asset price. Allowing for jumps of infinite activity is nontrivial and we leave such an extension for future work.

In order to obtain well-defined asymptotic limits, we shall assume that the factor loadings and the idiosyncratic variances are "moderately heterogeneous" in the sense that they are well-behaved under cross-sectional aggregation; see Assumption 4 below. In the analysis that follows, it is convenient to introduce factor loadings for the residual process that are defined as

$$\tilde{\lambda}_{j,\tau\pm} = \lambda_{j,\tau\pm} - \beta_{j,\tau}\lambda_{Z,\tau\pm}, \quad \tau \in \mathcal{T}. \tag{2.9}$$

In particular, $\tilde{\lambda}_{j,\tau-}$ and $\tilde{\lambda}_{j,\tau+}$ are the factor loadings on f for the residual process $Y_j - \beta_{j,\tau}Z$ before and after the jump time τ, respectively. For notational simplicity, below, we abuse our notation slightly by writing $\tilde{\lambda}_{j,q}$ in place of $\tilde{\lambda}_{j,\tau-}$ (resp. $\tilde{\lambda}_{j,\tau+}$) with $q = \tau-$ (resp. $q = \tau+$) indicating the pre-jump (resp. post-jump) window; the same convention also applies to other variables.

Assumption 4. *For $p, q \in \{\tau-, \tau+, \eta-, \eta+\}$, the following hold:*

(i) $N_n^{-1}\sum_{j=1}^{N_n}\tilde{\lambda}_{j,p}\tilde{\lambda}_{j,q}^\top \xrightarrow{\mathbb{P}} M_\Lambda(p,q)$ for some $\mathcal{F}$-measurable $r \times r$ random matrix $M_\Lambda(p,q)$;

(ii) $N_n^{-1}\sum_{j=1}^{N_n}\tilde{\sigma}_{j,q}^2 \xrightarrow{\mathbb{P}} M_\epsilon(q)$ for some $\mathcal{F}$-measurable random variable $M_\epsilon(q)$.

(iii) The factor loadings $(\tilde{\lambda}_{j,q})_{1\leq j\leq N_n}$ are independent of the idiosyncratic diffusive components $(\epsilon_j)_{1\leq j\leq N_n}$ conditional on the jump times in $\mathcal{T}$.

Conditions (i,ii) in Assumption 4 concern the convergence of cross-sectional sample averages. These conditions can be verified with an appeal to the law of large numbers provided that the loadings $\tilde{\lambda}_j$ and the idiosyncratic variances $\tilde{\sigma}_j^2$ are conditionally weakly dependent in the cross-section. We note that these variables can still be unconditionally strongly dependent, in which case the limiting variables $M_\Lambda(p,q)$ and $M_\epsilon(q)$ are generally non-degenerate random variables. We allow this complication in our econometric inference, which requires characterizing asymptotic distributions in terms of stable convergence in law. Condition (iii) appears to be mild; indeed, this assumption holds automatically if the factor loadings are non-random, which is typically assumed in the factor analysis literature.

Assumption 4 allows us to characterize precisely the aggregate behavior of the residual diffusive disturbances. Indeed, the cross-sectional average of the spot variances of the residual returns (i.e., $dY_{j,t} - \beta_{j,t}dZ_t$) before/after each jump time τ can be written as

$$\frac{1}{N_n}\sum_{j=1}^{N_n}\left(\tilde{\lambda}_{j,\tau\pm}^{\top}\Sigma_{f,\tau\pm}\tilde{\lambda}_{j,\tau\pm} + \tilde{\sigma}_{j,\tau\pm}^2\right) = \text{Trace}\left[\Sigma_{f,\tau\pm}\left(\frac{1}{N_n}\sum_{j=1}^{N_n}\tilde{\lambda}_{j,\tau\pm}\tilde{\lambda}_{j,\tau\pm}^{\top}\right)\right] + \frac{1}{N_n}\sum_{j=1}^{N_n}\tilde{\sigma}_{j,\tau\pm}^2,$$

which converges in probability (under Assumption 4) to

$$M_{\text{total}}(\tau\pm) \equiv \text{Trace}\left[\Sigma_{f,\tau\pm}M_{\Lambda}(\tau\pm,\tau\pm)\right] + M_{\epsilon}(\tau\pm). \tag{2.10}$$

The two components on the right-hand side of (2.10) are due to the diffusive factor component in the residual process and the diffusive idiosyncratic components in Y, respectively.

Finally, we maintain a very mild condition for the relative growth rates of Δ_n and N_n.

Assumption 5. *N_n grows to infinity at most polynomially in Δ_n^{-1}, that is, $N_n = O(\Delta_n^{-k})$ for some (arbitrary but fixed) constant $k > 0$.*

Assumption 5 is very weak because it does not require any specific rate at which N_n grows to infinity relative to Δ_n. This assumption is used to show that the cross-section of diffusive increments around the jump time fall uniformly in any fixed (small) neighborhood around zero, so that the loss function $L(\cdot)$ behaves like the quadratic function when acting on these returns. This assumption is not needed if the loss function is (globally) quadratic.

3 Testing the Exact Linear Jump Factor Model

This section contains the core of our econometric analysis. We derive a feasible test on the basis of the statistic $\hat{V}_n$ to test the null hypothesis of the exact linear jump factor model at two jump times of the systematic factor Z. We further derive a measure for the magnitude of the firm-specific shocks in the jump betas when the null might be violated.

3.1 Asymptotic Properties of the Test Statistic

We start with characterizing the asymptotic behavior of our test statistic. We first show the convergence in probability of the test statistic under the general random jump beta model (2.4), for which we need the following condition.

Assumption 6. *The innovations in jump betas $\{\chi_{j,\eta,\tau} \equiv \tilde{\chi}_{j,\tau} - \tilde{\chi}_{j,\eta} : 1 \leq j \leq N_n\}$ are $\mathcal{F}_{\eta-}$-conditionally independent on the cross-section.*

Assumption 6 mainly requires that the innovations in the jump beta (i.e., $\chi_{j,\eta,\tau}$) are cross-sectionally independent conditional on the information set before the jumps occur. We can further accommodate in the analysis, situations in which $\chi_{j,\eta,\tau}$ are cross-sectionally dependent. Such a situation can happen, if for example, there are extra factors at the jump times of Z (in addition to the jump size of Z). In this case, under some weak regularity conditions for the factor loadings on the omitted factor(s), the asymptotic behavior of our test statistic $L(\chi_{j,\eta,\tau})$ will be similar to that under Assumption 6 in terms of asymptotic order of magnitude. For ease of exposition, we do not consider this situation here.

Proposition 1 below describes the first-order asymptotic behavior of the test statistic.

Proposition 1. *Under Assumptions 1, 2, 3 and 6,* $\hat{V}_n = N_n^{-1}\sum_{j=1}^{N_n}\mathbb{E}[L(\chi_{j,\eta,\tau})|\mathcal{F}_{\eta-}] + o_p(1)$. *In particular, under the constant beta restriction (2.3),* $\hat{V}_n = o_p(1)$.

Proposition 1 shows that $\hat{V}_n$ converges to the cross-sectional average of the conditional expected loss that is due to the temporal variation in jump betas. This convergence result readily accommodates cross-sectional heterogeneity in the data. Of course, if the loss $L(\chi_{j,\eta,\tau})$ has identical conditional mean over the cross-section, the limit of $\hat{V}_n$ is simply $\mathbb{E}[L(\chi_{j,\eta,\tau})|\mathcal{F}_{\eta-}]$. This proposition suggests that the test statistic $\hat{V}_n$ is able to detect on-average deviations from the exact jump factor model.

In view of Proposition 1, we can state the hypotheses to be tested precisely as follows. The testing problem is to decide in which of the two events the observed sample path falls:

$$\begin{aligned}\Omega_0 &\equiv \{\beta_{j,\eta} = \beta_{j,\tau} = \beta_j \text{ for all } j\},\\ \Omega_a &\equiv \left\{\liminf_{n\to\infty}\frac{1}{N_n}\sum_{j=1}^{N_n}\mathbb{E}[L(\chi_{j,\eta,\tau})|\mathcal{F}_{\eta-}] > 0\right\},\end{aligned}$$

where Ω_0 and Ω_a play the role of the null and alternative hypotheses, respectively. Stating the hypotheses in terms of random events is typical in the infill asymptotic setting because the "population" quantities are (random) sample paths; see Jacod and Protter (2012) and Aït-Sahalia and Jacod (2014) for many similar examples. In addition, formulating the alternative hypothesis in terms of the limit inferior is also common in testing problems involving heterogeneous data (see, e.g., Giacomini and White (2006)).

In order to construct a proper test, we need to characterize the asymptotic distribution of the test statistic under the null hypothesis. Theorem 1, below, presents the result. We need some additional notation to represent the limit distribution. To this end, we consider random variables $(\kappa_\tau, \zeta_{\tau-}, \zeta_{\tau+})_{\tau\in\mathcal{T}}$ that, conditional on $\mathcal{F}$, are mutually independent with the following marginal distributions: κ_τ is uniformly distributed on $[0,1]$ and $\zeta_{\tau\pm}$ are r-dimensional standard normal.

With each $s \in \mathcal{T}$, we associate the following weights

$$w_{s-} \equiv \frac{\sqrt{\kappa_s}}{\Delta Z_s}, \quad w_{s+} \equiv \frac{\sqrt{1-\kappa_s}}{\Delta Z_s}. \tag{3.1}$$

Following our notational convention in Assumption 4, we write the weights in (3.1) as w_q indexed by $q \in \{s-, s+\}$; similarly, we denote for $p, q \in \{\eta-, \eta+, \tau-, \tau+\}$,

$$M_C(p,q) = \Sigma_{f,p}^{1/2} M_\Lambda(p,q) \Sigma_{f,q}^{1/2}. \tag{3.2}$$

Finally, we set

$$\begin{cases} \mathcal{A}(s) \equiv \sum_{p,q \in \{s-,s+\}} w_p w_q \zeta_p^\top M_C(p,q)\, \zeta_q + \sum_{q \in \{s-,s+\}} w_q^2 M_\epsilon(q), & s \in \{\eta, \tau\}, \\ \mathcal{B}(\eta,\tau) \equiv \sum_{p \in \{\tau-,\tau+\}} \sum_{q \in \{\eta-,\eta+\}} w_p w_q \zeta_p^\top M_C(p,q)\, \zeta_q. \end{cases} \tag{3.3}$$

Theorem 1. *Suppose that Assumptions 1–5 hold. In restriction to Ω_0, the sequence $\Delta_n^{-1}\hat{V}_n$ of variables converges $\mathcal{F}$-stably in law towards $\mathcal{L}(\eta,\tau)$ given by*

$$\mathcal{L}(\eta,\tau) \equiv \mathcal{A}(\eta) + \mathcal{A}(\tau) - 2\mathcal{B}(\eta,\tau). \tag{3.4}$$

COMMENTS. (i) The limiting variable $\mathcal{L}(\eta,\tau)$ given in (3.4) captures two types of sampling variability. The first type is generated by the diffusive factor process and it takes the form of quadratic functions of the Gaussian variables $\zeta_{\eta\pm}$ and $\zeta_{\tau\pm}$. The $\mathcal{F}$-conditional distributions of these terms are non-degenerate. Intuitively, the variable $\zeta_{\tau-}$ (resp. $\zeta_{\tau+}$) represents the asymptotic distribution of the normalized diffusive factor returns before (resp. after) the jump time τ, that is,

$$\frac{\Sigma_{f,\tau-}^{-1/2}\left(f_\tau - f_{(i(n,\tau)-1)\Delta_n}\right)}{\sqrt{\tau - (i(n,\tau)-1)\Delta_n}} \quad \left(\text{resp. } \frac{\Sigma_{f,\tau}^{-1/2}\left(f_{i(n,\tau)\Delta_n} - f_\tau\right)}{\sqrt{i(n,\tau)\Delta_n - \tau}}\right).$$

Importantly, the factor diffusive returns represent systematic risk and, hence, they have a nontrivial conditional distribution even after the cross-sectional aggregation. This is in sharp contrast to conventional econometric settings, because the (mixed) normality of the limiting variable here is not obtained from the aggregation of weakly dependent variables via a central limit theorem, but is implied from the local Gaussianity of the continuous Itô process f.

(ii) The second type of sampling variability reflected in $\mathcal{L}(\eta,\tau)$ (more specifically the second term $\sum_{q\in\{s-,s+\}} w_q^2 M_\epsilon(q)$ in the definition of $\mathcal{A}(s)$) is attributed to the idiosyncratic diffusive components of Y. Although the idiosyncratic Brownian shocks $d\widetilde{W}_j$ are "averaged out" in the cross-sectional averaging, this term can still have a non-degenerate $\mathcal{F}$-conditional distribution because of the variables $(\kappa_\eta, \kappa_\tau)$ that appear in the weights w_q for $q = \{\eta-, \eta+, \tau-, \tau+\}$. Intuitively, κ_τ represents the relative location of the exact jump time τ within its observation interval. This source of sampling variability is, again, systematic (as it is due to Z), so it "survives" the cross-sectional aggregation. In the case when average idiosyncratic spot variance does not co-jump with the price (i.e., $M_\epsilon(s-) = M_\epsilon(s+)$, for $s = \eta, \tau$), the $\mathcal{F}$-conditional distribution of this term becomes degenerate.

(iii) Jumps in the vector Y that happen outside the set $\mathcal{T}$ of jump times of Z play no role in the limit result for our test statistic in Theorem 1. This is because jump arrivals are "sparsely" scattered in time. Hence, with probability approaching one, each high-frequency interval containing one of the jump times in the set $\mathcal{T}$ will not contain jump times outside of $\mathcal{T}$ for "most" components in Y; the number of exceptions is small and their effect is "regularized" by the winsorization.

We note that the test statistic $\hat{V}_n$ is non-negative and, hence, the limiting variable $\mathcal{L}(\eta, \tau)$ concentrates on the positive line. We can gauge the relative contribution of the components in (3.4) by computing their $\mathcal{F}$-conditional means. It is easy to see that $\mathbb{E}[\mathcal{B}(\eta, \tau)|\mathcal{F}] = 0$ and, for $s \in \{\eta, \tau\}$, we have

$$\mathbb{E}[\mathcal{A}(s)|\mathcal{F}] = \frac{M_{\text{total}}(s-) + M_{\text{total}}(s+)}{2\Delta Z_s^2},$$

where $M_{\text{total}}(s\pm)$ is the (large-sample) cross-sectional average of individual stocks' residual spot variances after/before the jump time s; recall (2.10). Consequently, the limit distribution $\mathcal{L}(\eta, \tau)$ has a strictly positive conditional mean depending on the relative magnitude of the residual spot variances with respect to the jump size. In the current setting, it is therefore natural to define the signal-to-noise ratio for the post- and pre-jump window as

$$SNR_{s\pm} \equiv \frac{\Delta Z_s^2}{M_{\text{total}}(s\pm)}. \tag{3.5}$$

The $\mathcal{F}$-conditional mean of the limiting variable can then be succinctly written as

$$\mathbb{E}[\mathcal{L}(\eta, \tau)|\mathcal{F}] = \sum_{s \in \{\eta, \tau\}} \left(\frac{1}{2 \cdot SNR_{s-}} + \frac{1}{2 \cdot SNR_{s+}} \right). \tag{3.6}$$

From this expression, we see that under the null hypothesis, the test statistic tends to be centered "more positively" when the signal-to-noise ratios around the jump times are lower. These are exactly the times when it will be more difficult to separate the null hypothesis from the alternative.

In order to implement the test, we need to consistently estimate the critical values defined as the $\mathcal{F}$-conditional quantiles of the limiting variable $\mathcal{L}(\eta, \tau)$. This is a nontrivial task because the $\mathcal{F}$-conditional distribution of $\mathcal{L}(\eta, \tau)$ is highly nonstandard. Most importantly, it depends on the unknown random quantities $M_C(\cdot, \cdot)$ and $M_\epsilon(\cdot)$, which in turn involve the spot covariance of the *latent* diffusive factor process f, the factor loadings of a large cross-section of assets and the average idiosyncratic spot variances, before and after each jump time. The next subsection is devoted to addressing this problem.

3.2 Critical Values

For a matrix A, we denote $\|A\| = \mathrm{Trace}[A^\top A]^{1/2}$. To proceed, we suppose that some preliminary estimator $\tilde{\beta}_{n,j,\tau}$ of the jump beta is available and it satisfies the following assumption.

Assumption 7. *The following conditions hold for some variables* $(\beta^*_{j,\tau})_{1\leq j\leq N_n,\tau\in\mathcal{T}}$*:*

(i) $N_n^{-1}\sum_{j=1}^{N_n}|\tilde{\beta}_{n,j,\tau}-\beta^*_{j,\tau}|^2 = o_p(1)$*;*

(ii) for all $\tau\in\mathcal{T}$*,* $\tilde{\beta}_{n,j,\tau}$ *and* $\beta^*_{j,\tau}$ *are* $\mathcal{F}$*-measurable and bounded with probability approaching one;*

(iii) $\beta^*_{j,\tau}=\beta_j$ *in restriction to* Ω_0*.*

One possible choice for the preliminary beta estimators is to set $\tilde{\beta}_{n,j,\tau}$ to be the spot beta estimator $\hat{\beta}_{n,j,\tau}$. In this case, $\beta^*_{j,\tau}=\beta_{j,\tau}$ over the entire sample space. Condition (i) requires that these preliminary estimates are consistent in a cross-sectional average sense, which is easy to verify; see Lemma 3(b) for the formal result. More generally, the variables $\beta^*_{j,\tau}$ in Assumption 7 are interpreted as pseudo-true parameters in the sense that they coincide with the true parameter under the null hypothesis but are allowed to differ from the latter under the alternative. For example, we can take $\tilde{\beta}_{n,j,\eta}=\tilde{\beta}_{n,j,\tau}=(\hat{\beta}_{n,j,\eta}+\hat{\beta}_{n,j,\tau})/2$ which corresponds to $\beta^*_{j,\eta}=\beta^*_{j,\tau}=(\beta_{j,\eta}+\beta_{j,\tau})/2$. This type of averaging typically produces less noisy preliminary estimates under the null, which tends to have better size control in finite samples.[13]

We now describe how to calculate the critical values. The first step is to extract information concerning the spot covariance of the factors (i.e., Σ_f) and the factor loadings (i.e., λ_j and λ_Z) from the diffusive returns in local windows before and after the jump times. We consider an integer sequence k_n of local window sizes such that

$$k_n\to\infty \quad \text{and} \quad k_n\Delta_n\to 0.$$

For each detected jump time τ, we denote $i\,(n,\tau-)\equiv i\,(n,\tau)-k_n-1$ and $i\,(n,\tau+)\equiv i\,(n,\tau)$. We use $X_n^*(\tau\pm)$ to denote the $N_n\times k_n$ matrix whose (j,l) element is given by

$$X_n^*(\tau\pm)_{j,l}\equiv\frac{\Delta^n_{i(n,\tau\pm)+l}Y_j-\beta^*_{j,\tau}\Delta^n_{i(n,\tau\pm)+l}Z}{\sqrt{\Delta_n}}. \tag{3.7}$$

These matrices collect the (scaled) residual diffusive returns of the assets in the local windows.

We note that $X_n^*(\tau\pm)$ have an approximate factor structure

$$X_n^*\left(\tau\pm\right)_{j,l}\approx\left(\lambda_{j,\tau\pm}-\beta^*_{j,\tau}\lambda_{Z,\tau\pm}\right)^\top\frac{\Delta^n_{i(n,\tau\pm)+l}f}{\sqrt{\Delta_n}}+\frac{\Delta^n_{i(n,\tau\pm)+l}\epsilon_j}{\sqrt{\Delta_n}}. \tag{3.8}$$

[13] We can similarly use an average of jump betas over a larger number of systematic jump events, provided we are willing to assume the null hypothesis of the exact jump factor model holds true for this larger number of events.

This is only an approximation because the (time-varying) factor loadings (λ_j, λ_Z) before and after each jump time τ are only approximately $(\lambda_{j,\tau-}, \lambda_{Z,\tau-})$ and $(\lambda_{j,\tau}, \lambda_{Z,\tau})$, respectively.[14] It is convenient to rewrite (3.8) in matrix form as

$$X_n^*(\tau\pm) \approx \Lambda_n^*(\tau\pm) F_n(\tau\pm)^\top + \mathcal{E}_n(\tau\pm),$$

where we denote

$$\begin{cases} F_n(\tau\pm) \equiv \Delta_n^{-1/2} \left(\Delta_{i(n,\tau\pm)+1}^n f, \ldots, \Delta_{i(n,\tau\pm)+k_n}^n f \right)^\top, \\ \Lambda_n^*(\tau\pm) \equiv \left(\lambda_{1,\tau\pm} - \beta_{1,\tau}^* \lambda_{Z,\tau\pm}, \ldots, \lambda_{N_n,\tau\pm} - \beta_{N_n,\tau}^* \lambda_{Z,\tau\pm} \right)^\top, \\ \mathcal{E}_n(\tau\pm) \equiv \Delta_n^{-1/2} \left[\Delta_{i(n,\tau\pm)+l}^n \epsilon_j \right]_{1\leq j\leq N_n, 1\leq l\leq k_n}. \end{cases} \tag{3.9}$$

Assumption 8 extends Assumption 4 in order to accommodate the more general situation with pseudo-true parameters.

Assumption 8. *(i) For $p, q \in \{\eta-, \eta+, \tau-, \tau+\}$, $N_n^{-1} \Lambda_n^*(p)^\top \Lambda_n^*(q) \xrightarrow{\mathbb{P}} M_\Lambda^*(p,q)$ for some $\mathcal{F}$-measurable $r \times r$ random matrix $M_\Lambda^*(p,q)$.*

(ii) For each $q \in \{\eta-, \eta+, \tau-, \tau+\}$, the eigenvalues of $M_\Lambda^*(q,q)$ are distinct almost surely.

(iii) The factor loadings in $\Lambda_n^*(\tau\pm)$ are independent of the idiosyncratic diffusive components $(\epsilon_j)_{1\leq j\leq N_n}$.

Since $\beta_{j,\tau}^* = \beta_j$ for each $\tau \in \mathcal{T}$ under the null hypothesis, the matrix $M_\Lambda^*(p,q)$ coincides with $M_\Lambda(p,q)$ as well (recall Assumption 4). With this notation, we further complement the definition in (3.2) by setting

$$M_C^*(p,q) \equiv \Sigma_{f,p}^{1/2} M_\Lambda^*(p,q) \Sigma_{f,q}^{1/2}. \tag{3.10}$$

Then, the key to conducting feasible inference is to consistently estimate $M_C^*(p,q)$. To do this, we assume that the dimension r of the diffusive factors is bounded by a known constant $\bar{r}$. This assumption is much weaker than knowing the exact value of r. Since the number of the diffusive factors is not the main object of interest in our analysis of jumps, we aim to be agnostic on its exact value. Instead, we only assume that a bound $\bar{r}$ is known, for which the vast empirical literature provides reliable guidance. We shall show theoretically that knowing the upper bound is enough for constructing valid critical values for our test statistic.[15]

[14] The approximation error is also contributed by the drift and the idiosyncratic jump terms in Y_j. All these approximation errors are accounted for in our analysis.

[15] The number of diffusive factors may be consistently estimated by adapting the general strategy of Bai and Ng (2002). Like our setting, such methods also assume the availability of an upper bound for the number of factors (see the recent work of Gagliardini, Ossola, and Scaillet (2016a) which proposes a diagnostic criterion that does not need such a bound), subject to which a criterion function is minimized for estimating the number of factors. The resulting estimate may differ when using different methods (often associated with different penalty functions for model complexity). We avoid this potential ambiguity by only relying on the upper bound.

We now introduce our estimator for $M_C^*(p,q)$. We denote the sample analogue of $X_n^*(q)$ by $\hat{X}_n(q)$, which is constructed as

$$\hat{X}_n(\tau\pm)_{j,l} \equiv \left(\frac{\Delta^n_{i(n,\tau\pm)+l} Y_j \wedge u_n \vee (-u_n) - \tilde{\beta}_{n,j,\tau} \Delta^n_{i(n,\tau\pm)+l} Z}{\sqrt{\Delta_n}} \right), \tag{3.11}$$

where we recall that u_n is the sequence of threshold used for winsorizing jumps. We set $\hat{F}_n(q)$ as the $k_n \times \bar{r}$ matrix that consists of the eigenvectors associated with the $\bar{r}$ largest eigenvalues of $\hat{X}_n(q)^\top \hat{X}_n(q)$ under the normalization $\hat{F}_n(q)^\top \hat{F}_n(q)/k_n = I_{\bar{r}}$, where $I_{\bar{r}}$ denotes the $\bar{r}$-dimensional identity matrix. We then set

$$\hat{\Lambda}_n(q) \equiv \hat{X}_n(q) \hat{F}_n(q)/k_n. \tag{3.12}$$

Due to the normalization of the eigenvector matrix $\hat{F}_n(q)$, it no longer "carries" information concerning the spot covariance $\Sigma_{f,q}$ of the factor process f. Instead, this information is now embedded in the estimator $\hat{\Lambda}_n(q)$, which also contains information about the loading matrix $\Lambda_n^*(q)$. With this intuition in mind, we shall show that $M_C^*(p,q)$ can be estimated (in a proper sense) using

$$\widehat{M}_{C,n}(p,q) \equiv \frac{1}{N_n} \hat{\Lambda}_n(p)^\top \hat{\Lambda}_n(q), \quad p,q \in \{\eta-, \eta+, \tau-, \tau+\}. \tag{3.13}$$

It is interesting to note that the above estimator is the cross-variation of the extracted diffusive factor loadings across (possibly) different local windows, on which our feasible inference is based. Subsequently, the average idiosyncratic variance $M_\epsilon(q)$ can be estimated by

$$\widehat{M}_{\epsilon,n}(q) \equiv \frac{1}{N_n k_n} \|\hat{X}_n(q)\|^2 - \mathrm{Trace}[\widehat{M}_{C,n}(q,q)]. \tag{3.14}$$

In Theorem 2 below, we describe the asymptotic properties of the estimators $\hat{\Lambda}_n$, $\widehat{M}_{C,n}$ and $\widehat{M}_{\epsilon,n}$. We present explicitly the intermediate result for $\hat{\Lambda}_n$ in order to streamline the intuition underlying our constructions. Below, we partition $\hat{\Lambda}_n(q) = [\hat{\Lambda}_n^*(q) \vdots \hat{\Lambda}_n^0(q)]$ where the two blocks contain r and $\bar{r} - r$ columns, respectively. Similarly, we partition $\hat{F}_n(q) = [\hat{F}_n^*(q) \vdots \hat{F}_n^0(q)]$. We denote the sign function by $\mathrm{sign}(x) = 1_{\{x \geq 0\}} - 1_{\{x<0\}}$ and apply it component-wise on matrices.

Theorem 2. *Suppose that Assumptions 3, 4, 7 and 8 hold. For $p,q \in \{\eta-, \eta+, \tau-, \tau+\}$, we have the following:*

(a) $N_n^{-1} \|\hat{\Lambda}_n^*(q) - \Lambda_n^*(q) \Sigma_{f,q}^{1/2} H_q S_n^*(q)\|^2 = o_p(1)$ *where*

$$S_n^*(q) = \mathrm{diag}\left(\mathrm{sign}\left(\hat{F}_n^*(q)^\top F_n(q) (F_n(q)^\top F_n(q)/k_n)^{-1/2} H_q \right) \right),$$

and H_q *is the ordered orthogonal eigenvector matrix of* $M_C^*(q,q)$;[16]

(b) $N_n^{-1} \|\hat{\Lambda}_n^0(q)\|^2 = o_p(1)$;

[16] That is, $M_C^*(q,q) = H_q D H_q^\top$, where D is the diagonal matrix that collects the eigenvalues of $M_C^*(q,q)$ in descending order.

(c) the $\bar{r} \times \bar{r}$ matrix $\widehat{M}_{C,n}(p,q)$ satisfies

$$\widehat{M}_{C,n}(p,q) = \begin{pmatrix} S_n^*(p)H_p^\top M_C^*(p,q)H_q S_n^*(q) & 0 \\ 0 & 0 \end{pmatrix} + o_p(1),$$

where we note that the upper-left block $S_n^*(p)H_p^\top M_C^*(p,q)H_q S_n^*(q)$ is $r \times r$ dimensional;

(d) $\widehat{M}_{\epsilon,n}(q) \xrightarrow{\mathbb{P}} M_\epsilon(q)$.

COMMENTS. (i) Part (a) shows that, in a cross-sectional average sense, $\hat{\Lambda}_n^*(q)$ approximates $\Lambda_n(q)\Sigma_{f,q}^{1/2}$, up to the transformation $H_q S_n^*(q)$. Despite the presence of the latter "unobserved nuisance," this result is sufficient for our purpose of estimating the critical values. This is because the law of an $\mathcal{F}$-conditional standard normal vector is invariant to the transformation associated with the orthogonal matrix $H_q S_n^*(q)$.[17]

(ii) Part (b) shows that the "null" columns collected by $\hat{\Lambda}_n^0(q)$ are "asymptotically small." Hence, estimating the factor structure conservatively by (potentially over-) extracting factors leads to no effect asymptotically.

(iii) Parts (c,d) follow from parts (a,b) and establish a type of consistent estimation result for $\widehat{M}_{C,n}$ and $\widehat{M}_{\epsilon,n}$. Again, we note that the result of part (c) is sufficient for our feasible inference described below in spite of the appearance of the transformation $H_q S_n^*(q)$. We also note that when $p \neq q$, $\widehat{M}_{C,n}(p,q)$ estimates the cross-variation of factor loadings between two distinct blocks of data. For example, $\widehat{M}_{C,n}(\eta-,\tau+)$ concerns data blocks before the η-jump and after the τ-jump. This type of estimation is needed because we allow the factor loadings to change across jump times and actually we also allow them to co-jump with Z. This is why we need to characterize precisely the behavior of $\hat{\Lambda}_n^*(q)$ in the form of part (a). Such analysis is not common in conventional factor analysis (cf. Stock and Watson (2002), Bai (2003)).[18]

We remind the reader that our analysis is not about the diffusive factors per se. Rather, Theorem 2 is an intermediate (but important) step for making inference regarding the cross-sectional behavior of jump betas, which is the focus of the current paper. We are now ready to introduce the critical value $cv_{n,\alpha}$ for our test statistic $\hat{V}_n$ at some nominal level $\alpha \in (0,1)$. Algorithm 1 below provides the details and is quite intuitive as it follows closely from Theorem 1. The asymptotic validity of this algorithm is provided in Theorem 3 that follows.

Algorithm 1. Step 1. Simulate $(\tilde{\kappa}_s, \tilde{\zeta}_{s-}, \tilde{\zeta}_{s+})_{s\in\{\eta,\tau\}}$ independently such that $\tilde{\kappa}_s \sim$Uniform$[0,1]$ and $\tilde{\zeta}_{s\pm} \sim \mathcal{N}(0, I_{\bar{r}})$.

[17] In Bai (2003) and Pelger (2015a), the transformation matrix is only known to be invertible. Such a result would be sufficient for the consistency of the column space spanned the estimator, but is not enough for the inference problem considered here.

[18] See also Pelger (2015a,b) and Aït-Sahalia and Xiu (2017) for recent development in the high-frequency setting.

Step 2. Set $\tilde{w}_{n,s-} = \sqrt{\tilde{\kappa}_s}/\Delta^n_{i(n,s)} Z$ and $\tilde{w}_{n,s+} = \sqrt{1-\tilde{\kappa}_s}/\Delta^n_{i(n,s)} Z$ for $s \in \{\eta, \tau\}$. Then compute

$$\begin{cases} \tilde{A}_n(s) = \sum_{p,q \in \{s-,s+\}} \tilde{w}_{n,p}\tilde{w}_{n,q}\tilde{\zeta}_p^\top \widehat{M}_{C,n}(p,q)\,\tilde{\zeta}_q + \sum_{q\in\{s-,s+\}} \tilde{w}^2_{n,q}\widehat{M}_{\epsilon,n}(q), \quad s \in \{\eta,\tau\}, \\ \tilde{B}_n(\eta,\tau) = \sum_{p\in\{\tau-,\tau+\}}\sum_{q\in\{\eta-,\eta+\}} \tilde{w}_{n,p}\tilde{w}_{n,q}\tilde{\zeta}_p^\top \widehat{M}_{C,n}(p,q)\,\tilde{\zeta}_q, \\ \tilde{L}_n(\eta,\tau) = \tilde{A}_n(\eta) + \tilde{A}_n(\tau) - 2\tilde{B}_n(\eta,\tau). \end{cases}$$

Step 3. Repeat step 1 and step 2 for a large number of times and report the critical value $cv_{n,\alpha}$ as the $1-\alpha$ quantile of $\tilde{L}_n(\eta,\tau)$ in the Monte Carlo sample. □

Theorem 3. *The following statements hold under Assumptions 1–8.*

(a) The sequence $cv_{n,\alpha}$ of critical values is $O_p(1)$. Moreover, in restriction to Ω_0, $cv_{n,\alpha} \xrightarrow{\mathbb{P}} cv_\alpha$ where cv_α is the $\mathcal{F}$-conditional $1-\alpha$ quantile of $\mathcal{L}(\eta,\tau)$.

(b) The test associated with the critical region $\{\Delta_n^{-1}\hat{V}_n > cv_{n,\alpha}\}$ has asymptotic level α under the null hypothesis and has asymptotic power one under the alternative hypothesis, that is

$$\mathbb{P}\left(\Delta_n^{-1}\hat{V}_n > cv_{n,\alpha}|\Omega_0\right) \to \alpha, \quad \mathbb{P}\left(\Delta_n^{-1}\hat{V}_n > cv_{n,\alpha}|\Omega_a\right) \to 1.$$

Comment. Part (a) of Theorem 3 shows that the critical value $cv_{n,\alpha}$ consistently estimates the corresponding $\mathcal{F}$-conditional quantile of the limit distribution $\mathcal{L}(\eta,\tau)$ of the test statistic, and it remains stochastically bounded (i.e., tight) under the alternative. As a direct consequence, part (b) shows that the proposed test has correct size control under the null hypothesis, and is consistent under the alternative hypothesis.

3.3 Measuring Temporal Variation in Jump Betas

We have so far developed tests for the exact jump factor model in the cross-section. The test detects positive cross-sectional average losses resulting from innovations in the jump beta, i.e., $\chi_{j,\eta,\tau} = \tilde{\chi}_{j,\tau} - \tilde{\chi}_{j,\eta}$. From an estimation point of view, this loss also serves naturally as a measure for the variation of jump beta for a "representative" asset. Since this measure is of economic interest in its own, we further develop in this subsection econometric tools for making inference about it. We focus on the quadratic loss function, in which case the test statistic is given by

$$\hat{V}_n = \frac{1}{N_n}\sum_{j=1}^{N_n}(|\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}| \wedge \bar{B}_{n,\eta,\tau})^2.$$

As shown in equation (3.6), in the boundary case with $\chi_{j,\eta,\tau} = 0$ (i.e., when jump betas are time-invariant), the "raw" estimator $\hat{V}_n$ is not asymptotically centered, but has a positive bias that is high in situations with low signal-to-noise ratio. While this bias term is accounted for in the critical values of the test described in Theorem 3, it is of course more conventional from an estimation point of view to use a centered estimator in practice. To this end, we consider a bias-corrected estimator given by

$$\hat{v}_n \equiv \frac{1}{N_n}\sum_{j=1}^{N_n}(|\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}| \wedge \bar{B}_{n,\eta,\tau})^2 - \Delta_n \hat{b}_n(\eta,\tau), \tag{3.15}$$

where the correction term $\hat{b}_n(\eta,\tau)$ is defined as

$$\hat{b}_n(\eta,\tau) \equiv \frac{1}{N_n k_n}\sum_{s\in\{\eta,\tau\}}\left(\frac{\|\hat{X}_n(s-)\|^2 + \|\hat{X}_n(s+)\|^2}{2\left(\Delta^n_{i(n,s)}Z\right)^2}\right). \tag{3.16}$$

We note that $\hat{b}_n(\eta,\tau)$ is constructed as the sample analogue of the asymptotic bias described in (3.6).[19] In fact, as an intermediate step in the proof of Theorem 2, we have shown that

$$\frac{1}{N_n k_n}\|\hat{X}_n(s\pm)\|^2 \xrightarrow{\mathbb{P}} M^*_{\text{total}}(s\pm) \equiv \text{Trace}\left[\Sigma_{f,s\pm} M^*_\Lambda(s\pm, s\pm)\right] + M_\epsilon(s\pm).$$

We note that $M^*_{\text{total}} = M_{\text{total}}$ holds under the null hypothesis and, hence,

$$\hat{b}_n(\eta,\tau) \xrightarrow{\mathbb{P}} \mathbb{E}\left[\mathcal{L}(\eta,\tau)\,|\mathcal{F}\right]. \tag{3.17}$$

The estimator $\hat{v}_n$ is consistent for the cross-sectional average (conditional) second moment of $\chi_{j,\eta,\tau}$, denoted with $v(\eta,\tau)$. We further derive feasible CLT for it which allows constructing confidence intervals. The details are provided in the appendix.

4 Extensions

In this section, we briefly discuss two extensions of the main theoretical results developed in Section 3 above. Section 4.1 proposes a test for the joint hypothesis of constant jump beta involving multiple jump times. Section 4.2 describes how to use a mixed-scale approach to address the issue of gradual response to systematic jump events of individual stock prices, which is relevant for empirical studies about jumps of less liquid assets.

4.1 Joint Tests at Multiple Jump Times

The test described in Theorem 3 involves only a pair of jump times. The corresponding pairwise test can be easily extended to a joint analysis involving multiple jump times.[20] Among many possibilities, we describe here a specific implementation in which we simply compute the consecutive

[19] The correction term $\Delta_n\hat{b}_n(\eta,\tau)$ in (3.15) is of order $O_p(\Delta_n)$. This rate is consistent with that in Theorem 1, because this bias term is asymptotically relevant only when the jump betas are constant. That said, this correction should also be effective in finite-samples when innovations in jump betas are "local to zero," though a complete analysis under drifting sequences of data generating processes is beyond the scope of the current paper.

[20] We consider a Kolmogorov-type test. Another approach is to use multiple-testing techniques for controlling the false discovery rate; see, for example, Bajgrowicz, Scaillet, and Treccani (2010) and Bajgrowicz, Scaillet, and Treccani (2016) for this type of applications on jumps.

pairwise test statistics and then take their average. The theoretical justification for this procedure can be adapted straightforwardly from the main theory in Section 3; we thus omit the formal statements and instead provide only the algorithm for conducting the test.[21]

To fix ideas, we consider $P+1$ jump times $\{\tau_1, \ldots, \tau_{P+1}\}$ of Z. The pairwise test statistic for the pth pair of consecutive jump times $\{\tau_p, \tau_{p+1}\}$, $1 \le p \le P$, is denoted by

$$\hat{V}_n(\tau_p, \tau_{p+1}) \equiv \frac{1}{N_n} \sum_{j=1}^{N_n} L(|\hat{\beta}_{n,j,\tau_p+1} - \hat{\beta}_{n,j,\tau_p}| \wedge \bar{B}_{n,\tau_p,\tau_{p+1}}).$$

The null asymptotic distribution of $\Delta_n^{-1}\hat{V}_n(\tau_p, \tau_{p+1})$ is $\mathcal{L}(\tau_p, \tau_{p+1})$ as shown in Theorem 1. By a simple extension of that theorem, we can show that these convergences hold jointly, that is,

$$\left(\Delta_n^{-1}\hat{V}_n(\tau_p, \tau_{p+1})\right)_{1\le p\le P} \xrightarrow{\mathcal{L}\text{-}s} (\mathcal{L}(\tau_p, \tau_{p+1}))_{1\le p\le P}, \tag{4.18}$$

where $\xrightarrow{\mathcal{L}\text{-}s}$ denotes stable convergence in law. Importantly, the limiting variables are represented using the same variables $(\kappa_{\tau_p}, \zeta_{\tau_p-}, \zeta_{\tau_p+})_{1\le p\le P+1}$ that are described in Section 3.1; in particular, $\mathcal{L}(\tau_{p-1}, \tau_p)$ and $\mathcal{L}(\tau_p, \tau_{p+1})$ are $\mathcal{F}$-conditionally dependent because they both involve $\left(\kappa_{\tau_p}, \zeta_{\tau_p-}, \zeta_{\tau_p+}\right)$.

One possibility to implement a joint test is to "combine" the pairwise test statistics by taking their average, $P^{-1}\sum_{p=1}^{P}\hat{V}_n(\tau_p, \tau_{p+1})$. One drawback of this (joint) test statistic, however, is that it is not scale-invariant. In particular, it will tend to put a lot of weight to the pairwise tests which happen in very volatile periods and much less weight to those during calm periods within the testing time window.

For this reason, we consider instead a scaled statistic. In view of (3.6), we use the conditional mean of the limit distribution $\mathcal{L}(\tau_p, \tau_{p+1})$ as the scaling factor of the pairwise test, which can be estimated by $\hat{b}_n(\tau_p, \tau_{p+1})$ as shown in (3.17). The resulting joint test statistic becomes

$$\widehat{JT}_n \equiv \frac{1}{P}\sum_{p=1}^{P} \frac{\hat{V}_n(\tau_p, \tau_{p+1})}{\hat{b}_n(\tau_p, \tau_{p+1})}.$$

By the joint convergence (4.18), we see that under the joint null hypothesis, $\Delta_n^{-1}\widehat{JT}_n$ converges stably in law to

$$\frac{1}{P}\sum_{p=1}^{P} \frac{\mathcal{L}(\tau_p, \tau_{p+1})}{\mathbb{E}\left[\mathcal{L}(\tau_p, \tau_{p+1})|\mathcal{F}\right]}.$$

Under the alternative, $\Delta_n^{-1}\widehat{JT}_n$ diverges to $+\infty$ in probability. We reject the joint null hypothesis at significance level $\alpha \in (0,1)$ if $\Delta_n^{-1}\widehat{JT}_n > cv_{n,\alpha}^{JT}$, where the critical value $cv_{n,\alpha}^{JT}$ is computed using the following algorithm.

[21]Technical details are available upon request.

Algorithm 3. Step 1. Simulate independent $(\tilde{\kappa}_s, \tilde{\zeta}_{s\pm})_{s\in\{\tau_1,\ldots,\tau_{P+1}\}}$ such that $\tilde{\kappa}_s \sim$Uniform$[0,1]$ and $\tilde{\zeta}_{s\pm} \sim \mathcal{N}(0, I_{\bar{r}})$.
Step 2. Compute $\tilde{L}_n(\tau_p, \tau_{p+1})$ as in Algorithm 1 and set $\tilde{L}_n^{JT} = P^{-1}\sum_{p=1}^{P} \tilde{L}_n(\tau_p, \tau_{p+1})/\hat{b}_n(\tau_p, \tau_{p+1})$.
Step 3. Repeat step 1 and step 2 for a large number of times and report the critical value $cv_{n,\alpha}^{JT}$ as the $1-\alpha$ quantile of $\tilde{L}_n^{JT}$ in the Monte Carlo sample. □

We note that for the joint test we can take $\tilde{\beta}_{n,j,\tau_p} = P^{-1}\sum_{p=1}^{P} \hat{\beta}_{n,j,\tau_p}$ in the construction of $\hat{X}_n(\tau\pm)$ (see (3.11)) which in turn is used for computing $\widehat{M}_{C,n}(p,q)$. This is because the hypothesis now is that the jump beta remains constant across the $P+1$ jump events.

4.2 Gradual Jumps and the Mixed-Scale Approach

In practice, the analysis of the co-jump behavior among assets is often complicated by the so-called "gradual jump" phenomenon (Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009)). That is, jumps in the efficient price (e.g., due to a surprising macroeconomic announcement) may take a longer time to be fully incorporated in the observed prices of less liquid stocks than is the case for the highly liquid market index. Clearly, the resulting asynchronicity between jumps can lead to (downward) biases in the jump beta estimates and thus distort our testing results. In the jump regression setting, Li, Todorov, Tauchen, and Chen (2017) proposed a mixed-scale approach to address this issue. The idea is to precisely detect jump times in the market portfolio at the fine time scale Δ_n, and then to analyze the relationship between individual assets' price jumps with market jumps at a coarse time scale $k\Delta_n$ so as to mitigate the aforementioned asynchronicity. Li, Todorov, Tauchen, and Chen (2017) documented the empirical effectiveness of this simple approach. In this subsection, we introduce the mixed-scale strategy into our current problem of testing and estimating the noisy linear factor model. The theory underlying the mixed-scale extension is straightforward and, hence, we only focus on the implementation but we leave out the technical details for brevity.[22]

The mixed-scale version of our econometric procedure goes as follows. As in Section 2.1, we detect the jumps in the systematic factor Z at the (fine) scale Δ_n. However, we estimate the jump betas at a (possibly) coarse scale $k\Delta_n$ for some $k \geq 1$. A larger k makes the procedure more robust against the presence of gradual jumps, but at the cost of reducing the signal-to-noise ratio for making inference about the jumps.[23] We denote the coarsely sampled returns of Y_j by $\Delta_{i,k}^n Y_j = Y_{j,(i+k-1)\Delta_n} - Y_{j,(i-1)\Delta_n}$ and define $\Delta_{i,k}^n Z$ similarly. The corresponding jump beta

[22] Technical details are available upon request.
[23] Empirically, we find that taking the coarse scale to be 3 minutes is often sufficiently conservative. For robustness, it is also advisable to report results for different levels of mixed-scales which we do in our empirical section.

estimates are given by

$$\hat{\beta}^{(k)}_{n,j,\tau} \equiv \frac{\Delta^n_{i(n,\tau),k} Y_j}{\Delta^n_{i(n,\tau),k} Z},$$

where we emphasize the degree k of mixed-scale in our notation. Analogous to (2.5), we can define the mixed-scale test statistic $\hat{V}^{(k)}_n$ using these mixed-scale beta estimates.

The null asymptotic distribution of $\hat{V}^{(k)}_n$ is similar to that of $\hat{V}_n$, except that it uses different weights in the before/after jump windows as we now describe. We first generalize the definitions in (3.1) and (3.3) by setting,

$$\begin{cases} w^{(k)}_{s-} \equiv \dfrac{\sqrt{\kappa_s}}{\Delta Z_s}, \quad w^{(k)}_{s+} \equiv \dfrac{\sqrt{k-\kappa_s}}{\Delta Z_s}, \quad s \in \mathcal{T}, \\ \mathcal{A}^{(k)}(s) \equiv \displaystyle\sum_{p,q\in\{s-,s+\}} w^{(k)}_p w^{(k)}_q \zeta_p^\top M_C(p,q)\,\zeta_q + \sum_{q\in\{s-,s+\}} (w^{(k)}_q)^2 M_\epsilon(q), \\ \mathcal{B}^{(k)}(\eta,\tau) \equiv \displaystyle\sum_{p\in\{\tau-,\tau+\}} \sum_{q\in\{\eta-,\eta+\}} w^{(k)}_p w^{(k)}_q \zeta_p^\top M_C(p,q)\,\zeta_q. \end{cases} \tag{4.19}$$

By a straightforward adaptation of Theorem 1, we can show that under the null hypothesis, the normalized mixed-scale test statistic $\Delta_n^{-1}\hat{V}^{(k)}_n$ converges $\mathcal{F}$-stably in law towards

$$\mathcal{L}^{(k)}(\eta,\tau) \equiv \mathcal{A}^{(k)}(\eta) + \mathcal{A}^{(k)}(\tau) - 2\mathcal{B}^{(k)}(\eta,\tau).$$

In the general case with mixed-scale k, the $\mathcal{F}$-conditional mean of this limiting variable is now

$$\mathbb{E}[\mathcal{L}^{(k)}(\eta,\tau)|\mathcal{F}] = \sum_{s\in\{\eta,\tau\}} \left(\frac{1}{2}\frac{1}{SNR_{s-}} + \left(k - \frac{1}{2}\right)\frac{1}{SNR_{s+}} \right), \tag{4.20}$$

with the signal-to-noise ratio SNR defined as in (3.5). From (4.20), we see that the conditional mean of mixed-scaled test statistic depends more on the signal-to-noise ratio in the post-jump than the pre-jump window.

The critical values for the mixed-scaled test can be calculated by modifying the procedure in Section 3.2 as follows. First, we extend the definition of residual return matrices $\hat{X}_n(\tau\pm)$ to the k-mixed-scale version $\hat{X}^{(k)}_n(\tau\pm)$ by replacing $i(n,\tau+)$ with $i\,(n,\tau+) + k - 1$ in equation (3.11). Note that this actually only affects the definition of $\hat{X}^{(k)}_n(\tau+)$ while leaving $\hat{X}^{(k)}_n(\tau-) = \hat{X}_n(\tau-)$. The estimators $\widehat{M}_{C,n}$ and $\widehat{M}_{\epsilon,n}$ are then modified correspondingly. Second, we change the weights $\tilde{w}_{n,s\pm}$ in Algorithm 1 as follows

$$\tilde{w}^{(k)}_{n,s-} = \frac{\sqrt{\tilde{\kappa}_s}}{\Delta^n_{i(n,s),k} Z}, \quad \tilde{w}^{(k)}_{s+} = \frac{\sqrt{k-\tilde{\kappa}_s}}{\Delta^n_{i(n,s),k} Z}.$$

With these modifications, we can follow the same steps as in Algorithm 1 to compute the critical values for the mixed-scale test statistic $\hat{V}^{(k)}_n$.

Next, we describe how to adapt the estimator $\hat{v}_n$ for the variance of beta innovations (recall (3.15)) to the mixed-scale setting. Besides using the mixed-scale jump beta estimates, we also need to adjust the bias-correction term in view of (4.20). The mixed-scale generalization of $\hat{v}_n$ is given by

$$\begin{cases} \hat{v}_n^{(k)} \equiv \frac{1}{N_n} \sum_{j=1}^{N_n} \left(\hat{\beta}_{n,j,\tau}^{(k)} - \hat{\beta}_{n,j,\eta}^{(k)} \right)^2 - \Delta_n \hat{b}_n^{(k)}, \quad \text{where} \\ \hat{b}_n^{(k)} \equiv \frac{1}{N_n k_n} \sum_{s \in \{\eta,\tau\}} \left(\frac{1}{2} \frac{\|\hat{X}_n^{(k)}(s-)\|^2}{(\Delta_{i(n,s),k}^n Z)^2} + \left(k - \frac{1}{2}\right) \frac{\|\hat{X}_n^{(k)}(s+)\|^2}{(\Delta_{i(n,s),k}^n Z)^2} \right). \end{cases} \tag{4.21}$$

Similar to $\hat{v}_n$, the estimator $\hat{v}_n^{(k)}$ is consistent for the cross-sectional average (conditional) second-moment of $\chi_{j,\eta,\tau}$ and in the appendix we further provide a feasible CLT for it.

5 Monte Carlo Study

We now examine the finite-sample performance of the proposed test in simulation settings that are calibrated to match some key features of the data that is used in our empirical analysis. Below we present the results of our main simulation setup while in the Supplementary Appendix we provide additional results from a variety of extensions of this setup.

5.1 Simulation Settings

We start with describing the Monte Carlo setup. We consider $N = 100$ and $\Delta_n = 1/400$ which are very close to the corresponding numbers for the data set used in the empirical application. The unit of time is one day, so this sampling frequency corresponds roughly to sampling every minute. The log returns of the assets are generated from the following model:

$$\begin{cases} dY_{j,t} = \lambda_j^\top df_t + \tilde{\sigma}_{j,t} d\widetilde{W}_{j,t} + \beta_{j,t}\varphi_t dN_t, \\ dZ_{j,t} = \lambda_Z^\top df_t + \varphi_t dN_t, \end{cases} \tag{5.1}$$

where the process N_t is a counting process for the jumps. The diffusive systematic factor process f is a three-factor model given by

$$df_t = \Sigma_{f,t}^{1/2} dW_t, \quad \Sigma_{f,t} = \text{diag}(\sigma_{f,1,t}^2, \sigma_{f,2,t}^2, \sigma_{f,3,t}^2),$$

and $W_t = (W_{1,t}, W_{2,t}, W_{3,t})^\top$ is a three-dimensional standard Brownian motion. The log-variance process of factor $k \in \{1, 2, 3\}$ is generated from an Ornstein–Uhlenbeck process of the from

$$\begin{aligned} d\log(\sigma_{f,k,t}^2) &= 0.3 \left(\log(\bar{\sigma}_{f,k}^2) - \log(\sigma_{f,k,t}^2)\right) dt \\ &\quad + 0.3 \left(\rho \left(dW_{1,t} + dW_{2,t} + dW_{3,t}\right) + \sqrt{1-\rho^2} dB_{k,t} \right) + \tilde{\varphi}_{k,t} dN_t, \end{aligned}$$

where $\rho = -0.6$ captures the so-called leverage effect between the volatility and the price process and $(B_{k,t})_{1 \le k \le 3}$ are standard Brownian motions that are independent of W. We set

$$(\bar{\sigma}_{f,1}^2, \bar{\sigma}_{f,2}^2, \bar{\sigma}_{f,3}^2) = (0.92, 0.21, 0.07),$$

and draw the log-volatility jump sizes $\tilde{\varphi}_{k,t}$ independently from a Uniform$[0, 0.2]$ distribution. We draw the price jump size φ_t independently from

$$\varphi_t \sim \text{Uniform}[-\underline{\varphi} - 0.4, -\underline{\varphi}] \cup [\underline{\varphi}, \underline{\varphi} + 0.4].$$

We set the lower bound $\underline{\varphi} = 0.4$, which is about 8 times of the local diffusive standard deviation of Z for $\Delta_n = 1/400$. This is close to what we observe in the empirical application. The frequency of jump arrivals is set to one jump per month.

In each simulation, we draw the diffusive factor loadings λ_j independently from

$$\lambda_{j,1} \sim \text{Uniform}[0.3, 1.7], \quad \lambda_{j,2}, \lambda_{j,3} \sim \text{Uniform}[-1, 1],$$

and the factor loading of Z is fixed at $\lambda_Z = (1, 0, 0)^\top$.

Turning next to the idiosyncratic diffusive component of Y, we set

$$\tilde{\sigma}_{j,t}^2 = \tilde{\gamma}_j(\sigma_{f,1,t}^2 + \sigma_{f,2,t}^2 + \sigma_{f,3,t}^2). \tag{5.2}$$

In this setting, the idiosyncratic variances comove with the factor variances, which is empirically realistic. The "variance betas" $\tilde{\gamma}_j$ are drawn independently from a Uniform$[0.8, 1.2]$ distribution. We note that the idiosyncratic diffusive variance in our simulation is quite sizable; indeed, it is equal to the sum of all diffusive factor variances on average, so it is greater, in particular, than the variance of the leading factor Z. This simulation setting thus corresponds to a low signal-to-noise scenario for estimating the spot jump betas and, hence, presents a nontrivial challenge to the proposed test.

The above specification of Z and the diffusive part of Y is designed to mimic key features of the data set that we use in the empirical application. In particular, the volatility dynamics of Z, the frequencies of jumps and their magnitude in the Monte Carlo match approximately those of the market portfolio which we use in the next section as the reference process Z. Further, the diffusive factor structure and the magnitude of the idiosyncratic variance in Y match those of the empirical data set (recall that these quantities determine the limiting distribution of our test statistic).[24]

Finally, our specification of the jump betas is as follows. At the first jump time τ_1, the jump betas are drawn independently from

$$\beta_{j,\tau_1} \sim \text{Uniform}\,[0.7, 1.3]\,.$$

Under the null hypothesis, we set $\beta_{j,\tau_p} = \beta_{j,\tau_1}$ for $p > 1$, and under the alternative we set:

$$\beta_{j,\tau_p} = \beta_{j,\tau_{p-1}} + \chi_{j,\tau_{p-1},\tau_p}, \quad \chi_{j,\tau_{p-1},\tau_p} \sim \mathcal{N}\left(0, v_{\tau_{p-1},\tau_p}\right), \quad v_{\tau_{p-1},\tau_p} = \begin{cases} 0.1 & \text{Alternative 1,} \\ 0.2 & \text{Alternative 2.} \end{cases}$$

[24]More specifically, the eigenvalues of the matrix $M_C(\tau\pm, \tau\pm)$ decay on average as those in the real data and the ratio $M_\epsilon(\tau\pm)/\text{Trace}[M_C(\tau\pm, \tau\pm)]$ is similar to that in the data as well.

Note that Alternative 2 implies larger deviation from the null hypothesis than Alternative 1.

Below, we examine the test specified in Theorem 3 and refer to it as the baseline test. We also consider the mixed-scale versions implemented at scales $k = 2$ and $k = 3$ as described in Section 4.2. We remind the reader that the baseline test is a special case of the mixed-scale test with $k = 1$. In the empirical analysis, the mixed-scale approach is more robust to the presence of gradual jumps, but at the cost of lowering the signal-to-noise ratio; we aim to examine this trade-off in the simulation exercise. Finally, we also assess the performance of the joint test of Section 4.1. We use nine consecutive jumps for the construction of this test which is similar to the average number of jumps used in our empirical implementation. The joint test is expected to have more power relative to the pairwise test and we will study this in our Monte Carlo experiment.

The simulation results are based on 1,000 sample paths in total and the tuning parameters of the test are as follows. We set the loss function $L(x) = \min\{|x|, x^2\}$, which will also be used in our empirical analysis below. With the true number of diffusive factors being 3, we consider $\bar{r} \in \{1, 3, 5\}$ in our Monte Carlo experiments. The benchmark case is $\bar{r} = 3$, which coincides with the true number of factors. The case with $\bar{r} = 1$ is generally *not* justified by our asymptotic theory, but it allows us to understand the consequence of "undershooting" the number of diffusive factors. The case $\bar{r} = 5$ is justified by our theory and shows the effect of conservatively including "non-factors" in finite-samples. The local window parameter k_n is taken from $\{25, 30\}$. The algorithm for computing critical values is implemented using 1,000 simulations.

Finally, we set the truncation u_n in an adaptive way. In particular, for determining the jumps in Z, u_n is equal to $7 \times \Delta_n^{0.49} \times \sqrt{RV \wedge BV}$, where RV and BV denote the daily realized variance and so-called bipower variation (Barndorff-Nielsen and Shephard (2004)), respectively.[25] The former is a measure of total daily quadratic variation and the latter is a measure of the continuous part of the quadratic variation. The above high threshold minimizes the probability of erroneously classifying a diffusive increment as one containing a jump. For the truncation of the elements in Y in the calculation of $\widehat{X}_n(\tau\pm)$ in (3.11), we use twice the threshold for Z in order to account for the additional idiosyncratic risk in Y.

5.2 Simulation Results for the Pairwise Test

We start with the case with $k_n = 30$. Table 1 presents the Monte Carlo rejection rates for testing $\beta_{j,\tau} = \beta_{j,\eta}$, $1 \leq j \leq N$. We report rejection rates at nominal levels 10%, 5% and 1%. Panel A shows results for the case when $\bar{r}$ coincides with the number of diffusive factors (i.e., $\bar{r} = 3$). Under the null hypothesis, the finite-sample rejection rates for our baseline test are very close to

[25] We use a relatively stringent jump detection rule so as to avoid mis-classifying diffusive returns as jumps. This is particularly important in our current testing context, because the asymptotic theory depends crucially on using only the jump returns. The adverse effect of mis-classifying diffusive returns as jumps is illustrated in a simulation in the Supplemental Appendix.

the nominal levels. We also see similarly good performance at mixed-scale $k = 2$. These results are in line with our asymptotic theory. The test at mixed-scale $k = 3$ appears to be somewhat undersized under null hypothesis, reflecting the cost of using a larger mixed-scale.

Turning to the power analysis, we first observe that the baseline test rejects essentially with probability one under Alternative 1, which is consistent with the asymptotic theory. The mixed-scale tests are more conservative but still exhibit good power properties. The somewhat lower rejection rate at $k = 3$ is expected, because the jump signal gets weaker relative to the disturbance from the diffusive part, rendering sharp inference concerning the jumps more difficult. As the data generating process further deviates away from the null hypothesis (i.e., Alternative 2), the rejection rates of all tests become very close to one, even for the more conservative $k = 3$ mixed-scale test.

We now examine the effect of misspecifying the number of diffusive factors. Panel B of Table 1 shows the rejection rates when we "undershoot" the number of diffusive factors by setting $\bar{r} = 1$. We remind the reader that our asymptotic theory does *not* justify the validity of this implementation. Under the null hypothesis, the baseline test slightly overrejects at the 5% and 1% levels, but the overrejection appears to be more severe at the 10% level. For the mixed-scale tests with $k = 2$ and 3, the overrejection is offset by the underrejection seen in Panel A; such "offsetting" of course needs to be taken with a grain of salt. By contrast, Panel C of Table 1 shows that "overshooting" the number of diffusive factors does not lead to overrejection under the null. Indeed, the baseline test with $k = 1$ has rejection rates that are very close to the nominal levels, and the mixed-scale tests underreject only slightly in comparison with the same tests implemented using $\bar{r} = 3$. Under the alternative, results in Panel C are quite similar to those in Panel A, suggesting that the test have desirable power properties even if we set $\bar{r}$ conservatively.

Next, we investigate the sensitivity of the rejection rates with respect to the local window parameter k_n. To do so, we repeat the exercise in Table 1 except that we replace $k_n = 30$ with $k_n = 25$. The rejection rates are presented in Table 2, which are generally very similar to those in Table 1. These results suggest that reducing moderately the window size does not have much of an effect on the test. We hence focus on the case $k_n = 30$ in our numerical work below.

In summary, we find that the proposed test controls size well under the null hypothesis and has high rejection rates under the alternatives. The performance of the baseline test is closely in line with our asymptotic theory. The mixed-scale approach with $k = 2$ performs very similarly as the baseline test. When we increase the mixed-scale to $k = 3$, the test tends to be slightly undersized but it still has nontrivial power under the alternative. Finally, we find that the test is robust with respect to the choice of $\bar{r}$ and the local window k_n. Overall, we have seen that the finite-sample performance of the proposed test is satisfactory in empirically realistic settings.

Table 1: Rejection Rates of Pairwise Constant Beta Tests

Level	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Case $\bar{r} = 3$ (correct number of diffusive factors)									
	Null			Alternative 1			Alternative 2		
$k = 1$	9.7	3.8	0.5	99.9	99.9	98.5	100	100	99.9
$k = 2$	9.4	4.2	0.6	98.9	97.0	84.4	99.7	99.1	96.2
$k = 3$	7.9	2.6	0.7	93.8	87.6	64.5	99.7	99.1	96.2
Panel B: Case $\bar{r} = 1$ (underestimated number of diffusive factors)									
	Null			Alternative 1			Alternative 2		
$k = 1$	15.2	7.1	1.5	100	100	99.6	100	99.9	99.8
$k = 2$	10.8	5.4	0.7	98.6	96.9	87.7	99.7	99.2	96.5
$k = 3$	8.6	3.5	0.6	93.7	87.6	65.0	98.2	96.6	86.0
Panel C: Case $\bar{r} = 5$ (overestimated number of diffusive factors)									
	Null			Alternative 1			Alternative 2		
$k = 1$	10.5	5.5	0.9	100	100	99.4	100	100	99.8
$k = 2$	7.1	3.0	0.3	99.0	97.0	84.7	99.5	99.1	96.0
$k = 3$	6.6	2.9	0.2	93.3	86.6	60.2	97.8	95.1	83.5

Note: The table presents rejection rates of pairwise tests for constant beta in the setting with $k_n = 30$. Critical values are based on Algorithm 1 with 1,000 simulations.

Table 2: Rejection Rates of Pairwise Constant Beta Tests

	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Case $\bar{r} = 3$ (correct number of diffusive factors)									
	Null			Alternative 1			Alternative 2		
$k = 1$	11.1	5.0	0.9	100	100	99.0	100	100	99.8
$k = 2$	7.3	3.3	0.4	98.3	96.4	84.8	99.8	99.3	96.3
$k = 3$	7.5	3.3	0.5	90.5	83.3	56.1	98.3	96.1	84.0
Panel B: Case $\bar{r} = 1$ (underestimated number of diffusive factors)									
	Null			Alternative 1			Alternative 2		
$k = 1$	14.6	6.8	1.4	100	100	99.4	100	100	99.9
$k = 2$	12.3	4.5	0.9	98.4	97.4	85.9	99.7	99.4	96.7
$k = 3$	9.9	4.6	0.5	94.4	86.7	63.9	98.2	96.1	87.0
Panel C: Case $\bar{r} = 5$ (overestimated number of diffusive factors)									
	Null			Alternative 1			Alternative 2		
$k = 1$	9.4	4.5	0.5	99.9	99.9	98.3	99.9	99.9	99.9
$k = 2$	9.2	3.4	0.1	98.7	95.8	84.0	99.7	99.3	95.3
$k = 3$	6.8	2.9	0.4	93.4	84.3	59.0	97.3	94.7	82.4

Note: The table presents rejection rates of pairwise tests for constant beta in the setting with $k_n = 25$. Critical values are based on Algorithm 1 with 1,000 simulations.

Table 3: Rejection Rates of Joint Constant Beta Tests

Level	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: Case $\bar{r} = 3$ (correct number of diffusive factors)									
	Null			Alternative 1			Alternative 2		
$k = 1$	10.0	4.1	0.8	100	100	100	100	100	100
$k = 2$	8.3	4.3	0.5	100	100	99.5	100	100	100
$k = 3$	4.3	1.0	0.0	99.3	97.7	85.6	99.9	99.6	97.5
Panel B: Case $\bar{r} = 1$ (underestimated number of diffusive factors)									
	Null			Alternative 1			Alternative 2		
$k = 1$	12.6	6.1	1.3	100	100	100	100	100	100
$k = 2$	9.2	3.2	0.3	100	100	99.7	100	100	99.8
$k = 3$	6.2	3.0	0.4	99.4	97.6	84.0	99.8	99.3	94.6
Panel C: Case $\bar{r} = 5$ (overestimated number of diffusive factors)									
	Null			Alternative 1			Alternative 2		
$k = 1$	11.0	5.0	1.2	100	100	100	100	100	100
$k = 2$	6.5	2.8	0.3	100	100	98.9	100	100	100
$k = 3$	4.5	1.5	0.1	99.4	97.1	85.0	99.7	99.1	94.3

Note: The table presents rejection rates of joint tests for constant beta in the setting with $k_n = 30$. The number of jump times included in the test is 9. Critical values are based on Algorithm 3 with 1,000 simulations.

5.3 Simulation Results for the Joint Test

We finish this section with presenting simulation results for the performance of the joint test in Table 3. For brevity, we only consider the case with $k_n = 30$. Compared to the pairwise test, the under-rejection under the null hypothesis now becomes more significant for the mixed-scale $k = 3$. Intuitively, the small finite-sample biases in the pairwise test play a more prominent rule in the joint test as pooling the pairwise tests across the different jump times reduces the sampling variability without affecting these biases. On the other hand, and as expected, we now have uniformly significantly more power to reject the two alternatives.

6 Empirical Application

We now apply the econometric techniques developed above to a data set from the U.S. equity market. Our interest in this analysis will be the factor structure and presence of firm-specific risk in the cross-section of asset returns during market-wide jump events.

6.1 Data

We start with describing the data. Our sample consists of one-minute returns for a panel of 93 stocks that were in the S&P 100 index for the entire period 2007–2015.[26] The data is from the Trade and Quote (TAQ) database. Standard cleaning procedures were used in assembling the data. Our proxy for the market (which is our systematic factor Z) is the S&P 500 E-Mini futures Index on the nearest contract with appropriate rollovers. The futures data were obtained from Tick Data. Of course, nothing in our theory requires Z to be the market factor, and empirical extensions could consider alternative sets of interesting event times such as macro and/or monetary announcement times, volatility jumps, as well as jumps in the various S&P 500 industry sub-components, among others.

The detection of jumps in Z as well as the truncation needed for the construction of $\widehat{X}_n(\tau\pm)$ in (3.11) is done exactly as in the Monte Carlo, with the only difference being the scaling of the threshold by a time-of-day function in order to account for the well-known diurnal intraday pattern in volatility. We find 91 putative jump moves in Z in our sample. We further filter these moves based on the following criteria: remove detections that occur within 5 minutes of each other and remove detections of size less than 0.10 (i.e., 10 basis points); the latter occurred only in the very late part of the sample with exceptionally low volatility. These steps yield 83 (one-minute) jump returns over the 2007–2015 period.[27] Evidently, the 83 jump returns are very large moves in the futures index relative to its local volatility. Our analysis below will be based on these 83 market jump events in the sample.

6.2 Test Outcomes

We first perform tests for the constant jump beta hypothesis to each of the 82 consecutive pairs of market jump events in our sample. As in the simulations, we use the Huber-type loss function $L(x) = \min\{|x|, x^2\}$ and take the local window parameter to be $k_n = 30$. This exercise consists of testing 82 hypotheses independently, and we summarize the results by reporting the proportion of tests that reject the null hypothesis at the 5% and 1% nominal levels when averaged over the whole sample. These separate pairwise tests are for the null hypothesis of no variation in adjacent jump betas. To further increase power, we also implement the joint test of Section 4.1 for the

[26] There are 107 ticker symbols that were in the S&P 100 Index for at least four years over the same period. Our inference techniques do not require a balanced panel, but we retained the 93 stocks mainly for transparency and ease of exposition. In initial work, we did the pairwise tests for all 107 stocks with essentially no difference in outcomes relative to those reported here.

[27] Similar to Bajgrowicz, Scaillet, and Treccani (2016), we do not find much evidence for clustering of these very big jumps although the statistical power for detection of such clustering is probably not very high given the rare nature of these jumps.

jumps in each of the calendar years in our sample and report the average of the p-values across the nine years in the data set. As shown in our simulation experiment, the joint test is expected to have more power to detect violations of the null hypothesis as those considered in the Monte Carlo. We choose the time span of one calendar year for the joint test in order to avoid violations of the null hypothesis of the exact jump factor model that are due to predictable changes in the factor loadings. Indeed, the latter are typically associated with changes in the firms' characteristics and they are not expected to change much over periods of one year.[28] Therefore, violations of the null hypothesis in our setting will be more likely due to the presence of unpredictable firm-specific shocks at the market-wide jump events.

For implementing the tests, we choose $\bar{r} = 3$ as motivated by the popular Fama-French model (Fama and French (1993)) and consider $\bar{r} = 5$ as a more conservative robustness check. Since the results are not sensitive to the choice of $\bar{r}$, below we focus the discussion on the case $\bar{r} = 3$. Finally, we present results with and without winsorization; the winsorization is implemented at the 0.05 level (i.e., $q_n^w = 0.05$). Recall from Section 2 that the winsorization is needed in order to guard against the possible occurrence of idiosyncratic jumps in the high-frequency increment at which a systematic jump event has been detected. In finite samples, the winsorization also provides robustness against possible presence of outliers in the data.

We turn now to a discussion of the test results. From Panel A of Table 4, the baseline test at the one-minute frequency (i.e., $k = 1$) rejects the constant beta hypothesis for a majority of consecutive jump pairs: the rejection rate is 0.94 at the 5% level and 0.76 at the 1% level. The average p-value of the joint test is very low, suggesting a strong rejection of the joint hypothesis. However, as recognized in Li, Todorov, Tauchen, and Chen (2017), the constant beta hypothesis at a fine scale will be rejected if there is a gradual response in individual assets to a market jump event due to less liquidity in the trading process of some of the stocks in the cross-section at such frequency. The mixed-scale version of our test described in Section 4.1 mitigates the effect of such microstructure issues, and we consider it therefore as the more reliable method. By "zooming out" and measuring returns at the two-minute sampling interval (i.e., $k = 2$), we find that the rejection rate drops substantially: for example, only 23% of the tests reject at the 1% level, which is far below the rejection rate of the baseline test with $k = 1$. In addition, the averaged p-value of the joint test is slightly below the conventional significance level of 5%, suggesting only borderline rejection for the joint hypothesis. When we further zoom out by increasing the mixed-scale parameter to $k = 3$, the majority of pairwise tests no longer reject the constant beta hypothesis. Furthermore, the p-value for the joint test is far above the conventional significance level.

As seen from Panel B of Table 4, after performing winsorization, the rejection rates of the tests

[28]Our pairwise test is most "conservative" in this regard as the time between the market jumps is typically around one month.

drop uniformly across all considered settings. However, qualitatively the results are the same as in the case of no winsorization.

Overall, the significant drop in the rejection rate of both the pairwise and joint tests, when going from $k = 1$ to $k = 3$ observed in our data, is inconsistent with the behavior of the tests under a typical alternative hypothesis of a noisy jump factor model as illustrated in our simulation experiments. Instead, such behavior is consistent with a null hypothesis of exact jump factor model which is "clouded" at the fine time scale of one-minute by liquidity-related microstructure issues.[29]

We can further contrast this result for the jumps with the factor structure of the diffusive returns before and after the systematic jump events. The cross-sectional average contribution of idiosyncratic diffusive risk in the total return variation of the residual component $X_n^*(\tau\pm)_{j,l}$ in (3.7) (computed as the time series of the ratio $\widehat{M}_{\epsilon,n}/(\mathrm{Trace}[\widehat{M}_{C,n}(q,q)] + \widehat{M}_{\epsilon,n})$ for $\bar{r} = 3$) is a nontrivial 67%. This decomposition of systematic and idiosyncratic variance in $X_n^*(\tau\pm)_{j,l}$ reveals also that the market cannot solely account for the cross-sectional dependence in asset returns before and after the systematic jump events which is in sharp contrast to our empirical evidence in support of the exact jump factor model at market jump times.

6.3 The Magnitude of Firm-specific Shocks at Market-wide Jump Events

We next estimate the variance $v(\eta, \tau)$ of the innovations in jump betas using the estimator $\hat{v}_n$ (see (3.15)), as well as the mixed-scaled version $\hat{v}_n^{(k)}$ (see (4.21)). Although the estimand $v(\eta, \tau)$ is non-negative, the bias-corrected estimator $\hat{v}_n^{(k)}$ occasionally takes negative values in finite-samples because of the subtraction of the positive bias-correction term. Therefore, we report results for a sign-regularized version of $\hat{v}_n^{(k)}$ given by[30]

$$\hat{v}_n^{(k)+} = \max\{\hat{v}_n^{(k)}, 0\}. \tag{6.1}$$

To evaluate economically the relative magnitude of these estimates, we also report the following measure

$$\text{Normalized } \hat{v}_n^{(k)+} \equiv \hat{v}_n^{(k)+} \left(\frac{\left(\Delta^n_{i(n,\tau),k} Z\right)^2 + \left(\Delta^n_{i(n,\eta),k} Z\right)^2}{2} \right) \Big/ D_n, \tag{6.2}$$

where

$$\begin{aligned} D_n \quad \equiv \quad & \frac{1}{N_n} \sum_{j=1}^{N_n} (\Delta^n_{i(n,\tau)} Y_j)^2 + \frac{1}{N_n} \sum_{j=1}^{N_n} (\Delta^n_{i(n,\eta)} Y_j)^2 + \frac{1}{2} \sum_{s \in \{\tau,\eta\}} \frac{1}{k_n N_n} \sum_{j=1}^{N_n} \sum_{l=1}^{k_n} \left(\Delta^n_{i(n,s-)+l} Y_j\right)^2 \\ & + \left(k - \frac{1}{2}\right) \sum_{s \in \{\tau,\eta\}} \frac{1}{k_n N_n} \sum_{j=1}^{N_n} \sum_{l=1}^{k_n} \left(\Delta^n_{i(n,s+)+k-1+l} Y_j\right)^2. \end{aligned}$$

[29] Of course, such a behavior of the test can be also consistent with very short-lived volatility bursts which are confused for jumps, see e.g., Bajgrowicz, Scaillet, and Treccani (2016).

[30] In the Supplementary Appendix, we investigate the effect from the bias-correction as well as from imposing the non-negativity constraint.

Table 4: Test Results for U.S. Equity Data

Mixed Scale	Rejection Rate 5% Level	Rejection Rate 1% Level	Joint Test Mean P-value	Rejection Rate 5% Level	Rejection Rate 1% Level	Joint Test Mean P-value
	Panel A: Full Sample ($q_n^w = 0$)					
	$\bar{r} = 3$			$\bar{r} = 5$		
$k = 1$	0.94	0.76	0.00	0.91	0.72	0.00
$k = 2$	0.38	0.23	0.04	0.35	0.21	0.05
$k = 3$	0.18	0.00	0.37	0.17	0.00	0.37
	Panel B: Winsorized Sample ($q_n^w = 0.05$)					
	$\bar{r} = 3$			$\bar{r} = 5$		
$k = 1$	0.89	0.66	0.00	0.89	0.63	0.00
$k = 2$	0.32	0.21	0.08	0.32	0.21	0.08
$k = 3$	0.13	0.00	0.44	0.10	0.00	0.45

Note: P-values are computed using Algorithm 1 based on 1,000 simulations. The joint test is done by grouping jumps in periods of calendar years.

The ratio in (6.2) measures an average stock's price variation due to time-varying jump beta (i.e., the numerator) relative to the total price variation during market jump times (i.e., D_n).

Panel A of Table 5 displays the empirical quantiles (across jumps and stocks) of the estimates $\hat{v}_n^{(k)+}$ and the normalized $\hat{v}_n^{(k)+}$ for all 82 pairs of consecutive jumps; the quantiles are invariant to the monotonic transformation used in the regularization (6.1). Panel A1 suggests that the median value of $\hat{v}_n^{(k)+}$ is about 0.17 and this estimate varies little across various levels of mixed-scales. To gauge their relative magnitudes, we observe in Panel A2 that the median value of the normalized measure at mixed-scale $k = 1, 2$ and 3 are 7.70%, 6.80% and 5.38%, respectively. This means, for example, for jump returns measured at the 2-minute frequency, return variance due to the temporal variation of the jump betas explains only about 6.80% of the total jump return variation. This rather small percent is consistent with our testing result that the majority of pairwise tests do not reject the null hypotheses at the conventional significance levels (for $k = 2$).

In Panel B of Table 5 we report the same statistics but after performing winsorization in the cross-section. This winsorization is to guard against distortions resulting from the idiosyncratic jump risk. Indeed, the estimators of variance can be very sensitive to outliers and the estimates in Panel A may be overly influenced by a few stocks, which makes them of less interest as measures of jump beta variations for the majority of stocks.[31] Clearly, the estimates after winsorization are much smaller than those based on the full sample, suggesting that the estimates in Panel A have likely exaggerated the variations in jump betas for a typical stock because of a few influential observations.

The estimates reported in Panel B1 of Table 5 indicate that between consecutive large rare jumps the typical betas differ randomly with a variance of about 0.02, based on the estimate for $k = 3$. For perspective on this estimate, the corresponding normalized measure (6.2) indicates that such variation of beta explains only 0.99% of the total return variation of the median stock in our sample at the systematic jump event. By contrast, the relative contribution of idiosyncratic risk in the total diffusive return variation in the local windows before and after the systematic jump events (based on $\bar{r} = 3$) is a rather nontrivial 48%.

Taken together, the results in Table 5 document a rather small role for firm-specific shocks in asset returns at the times of market-wide jump events. This evidence is in line with the support for the exact jump factor model reported earlier.

[31] After all, as formalized in Assumption 9, we can only hope to make inference about $v(\eta, \tau)$ if the stocks are relatively homogeneous.

Table 5: Idiosyncratic Risk at Market-wide Jump Events

Mixed Scale	Percentiles 10%	25%	50%	75%	90%	Percentiles 10%	25%	50%	75%	90%
					Panel A: Full Sample ($q_n^w = 0$)					
	(A1) $\hat{v}_n^{(k)+}$					*(A2) Normalized* $\hat{v}_n^{(k)+}$ *(%)*				
$k = 1$	0.09	0.12	0.17	0.23	0.30	3.92	5.11	7.70	11.26	14.91
$k = 2$	0.04	0.09	0.19	0.34	0.70	1.31	3.64	6.80	11.74	22.43
$k = 3$	0.00	0.03	0.16	0.30	0.85	0.00	0.91	5.38	11.58	24.04
					Panel B: Winsorized Sample ($q_n^w = 0.05$)					
	(B1) $\hat{v}_n^{(k)+}$					*(B2) Normalized* $\hat{v}_n^{(k)+}$ *(%)*				
$k = 1$	0.06	0.09	0.13	0.19	0.22	2.51	3.61	5.84	8.37	10.86
$k = 2$	0.00	0.04	0.11	0.18	0.37	0.00	1.42	4.33	7.79	11.97
$k = 3$	0.00	0.00	0.02	0.15	0.35	0.00	0.00	0.99	6.11	10.64

7 Conclusion

In this paper we develop a formal test for deciding whether an exact factor model holds for asset returns in a large cross-section at the jump times of discretely-observed systematic risk factors. The inference is based on a panel of high-frequency asset returns in which the cross-sectional dimension and the sampling frequency increase simultaneously but with no restriction on their relative asymptotic order. The test is based on comparing temporal variation of estimates for the factor loadings from detected consecutive jump times of the risk factor. This difference in factor loading estimates shrinks asymptotically when the null hypothesis is true and is $O_p(1)$ otherwise. The limit distribution of the test is non-standard and depends on systematic and idiosyncratic diffusive risks around the jump times of the factors in distinct ways. We further develop an estimator and inference tools that allow the econometrician to formally assess the magnitude of firm-specific shocks in assets at the times of systematic factor jump events. Empirical application to stocks in the S&P 100 index provides support for an exact market jump model over the period 2007-2015. This stands in sharp contrast to our evidence for the presence of multiple systematic risk factors in the asset returns prior to and following the systematic jump events.

References

Aït-Sahalia, Y., and J. Jacod (2014): *High-Frequency Financial Econometrics*. Princeton University Press.

Aït-Sahalia, Y., and D. Xiu (2017): “Principal Component Analysis of High Frequency Data,” *Journal of the American Statistical Association, Forthcoming.*

Andrews, D. W. K. (2005): “Cross-Section Regression with Common Shocks,” *Econometrica*, 73(5), 1551–1585.

Ang, A., and J. Chen (2002): “Asymmetric Correlations of Equity Portfolios,” *Journal of Financial Economics*, 63, 443–494.

Bai, J. (2003): “Inferential Theory for Factor Models of Large Dimensions,” *Econometrica*, 71(1), 135–171.

——— (2009): “Panel Data Models with Interactive Fixed Effects,” *Econometrica*, 77(4), 1229–1279.

Bai, J., and S. Ng (2002): “Determining the Number of Factors in Approximate Factor Models,” *Econometrica*, 70(1), 191–221.

Bajgrowicz, P., O. Scaillet, and A. Treccani (2010): “False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas,” *Journal of Finance*, 65, 179–216.

——— (2016): “Jumps in High-Frequency Data: Spurious Detections, Dynamics and News,” *Management Science*, 62, 2198–2217.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008): “Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise,” *Econometrica*, 76, 1481–1536.

——— (2009): “Realized Kernels in Practice: Trades and Quotes,” *The Econometrics Journal*, 12(3), C1–C32.

Barndorff-Nielsen, O. E., and N. Shephard (2004): “Power and Bipower Variation with Stochastic Volatility and Jumps,” *Journal of Financial Econometrics*, 2, 1–37.

——— (2006): “Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation,” *Journal of Financial Econometrics*, 4, 1–30.

Bollerslev, T., S. Li, and V. Todorov (2016): “Roughing Up Beta: Continuous versus Discontinuous Betas, and the Cross-Section of Expected Stock Returns,” *Journal of Financial Economics*, 120, 464–490.

Connor, G., M. Hagmann, and O. Linton (2012): “Efficient Semiparametric Estimation of the Fama-French Model and Extensions,” *Econometrica*, 80(2), 713–754.

Fama, E., and K. French (1993): “Commnon Risk Factors in the Returns on Stocks and Bonds,” *Journal of Financial Economics*, 33.

Gagliardini, P., E. Ossola, and O. Scaillet (2016a): “A Diagnostic Criterion for Approximate Factor Structure,” Discussion paper, Swiss Finance Institute.

——— (2016b): “Time-Varying Risk Premium in Large Cross-Sectional Equity Data Sets,” *Econometrica*, 84(3), 985–1046.

Giacomini, R., and H. White (2006): “Tests of Conditional Predictive Ability,” *Econometrica*, 74(6), 1545–1578.

Hansen, L. P., and S. F. Richard (1987): “The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models,” *Econometrica*, 55(3), 587 – 613.

Huber, P. (2004): *Robust Statistics*. John Wiley & Sons.

Jacod, J., and P. Protter (2012): *Discretization of Processes*. Springer.

Lee, D. S., and T. Lemieux (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48(2), 281–355.

Li, J., V. Todorov, and G. Tauchen (2017): "Jump Regressions," *Econometrica*, 85, 173–195.

Li, J., V. Todorov, G. Tauchen, and R. Chen (2017): "Mixed-scale Jump Regressions with Bootstrap Inference," *Journal of Econometrics, Forthcoming.*

Li, J., V. Todorov, G. Tauchen, and H. Lin (2017): "Rank Tests at Jump Events," *Journal of Business and Economic Statistics, Forthcoming.*

Longin, F., and B. Solnik (2001): "Extreme Correlation of International Equity Markets," *Journal of Finance*, 56, 649–676.

Mancini, C. (2001): "Disentangling the Jumps of the Diffusion in a Geometric Jumping Brownian Motion," *Giornale dell'Istituto Italiano degli Attuari*, LXIV, 19–47.

Mancini, C. (2009): "Non-parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps," *Scandinavian Journal of Statistics*, 36, 270–296.

Merton, R. C. (1976): "Option Pricing When Underlying Stock Returns Are Discontinuous," *Journal of Financial Economics*, 3, 125 – 144.

Moon, H. R., and M. Weidner (2015): "Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects," *Econometrica*, 83(4), 1543–1579.

Pelger, M. (2015a): "Large-Dimensional Factor Modeling based on High-Frequency Observations," Discussion paper, Stanford University.

——— (2015b): "Understanding Systematic Risk: A High-Frequency Approach," Discussion paper, Stanford University.

Pesaran, M. H. (2006): "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure," *Econometrica*, 74(4), 967–1012.

Savor, P., and M. Wilson (2014): "Asset Pricing: A Tale of Two Days," *Journal of Financial Economics*, 113, 171–201.

Stock, J. H., and M. W. Watson (2002): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97(460), 1167–1179.

8 Appendix

8.1 Additional Results for Measuring the Temporal Variation in Jump Betas

Our goal in this section is to derive the asymptotic properties of the estimator $\hat{v}_n$ and develop a feasible inference procedure for it. Unlike the limit theory developed for the test, understanding the asymptotic behavior of $\hat{v}_n$ under the more general random jump beta model requires a joint characterization of the (cross-sectional) sampling variability resulting from the innovation terms $\chi_{j,\eta,\tau}$ and the (time-series) high-frequency estimation error in the jump beta estimates. Technically speaking, one needs to properly combine a type of conditional convergence in law for the former with the stable convergence for the latter. To this end, we need some additional assumptions on the dependence structure concerning the underlying variables.

Assumption 9. *Let $\mathcal{C}$ be a sub σ-field of $\mathcal{F}$ and $\mathcal{C}_t$ be the smallest filtration that contains $\mathcal{F}_t$ and $\mathcal{C}$. The following conditions hold:*

(i) Z_t and f_t are $\mathcal{C}$-measurable for all $t \in [0, T]$;

(ii) conditional on $\mathcal{C}_{\eta-}$, the variables $(\chi_{j,\eta,\tau})_{1\leq j\leq N_n}$ are independent with zero mean and bounded $4 + \iota$ moments for some fixed (small) constant $\iota > 0$;

(iii) conditional on $\mathcal{C}_{\eta-}$, the variables $(\chi_{j,\eta,\tau})_{1\leq j\leq N_n}$ are independent of the diffusive components of $(Y_j)_{1\leq j\leq N_n}$, the jump components $(\tilde{J}_{Y,j})_{1\leq j\leq N_n}$ and the jump betas $(\beta_{j,\eta})_{1\leq j\leq N_n}$ at time η.

(iv) $N_n^{-1/2} \sum_{j=1}^{N_n} (\mathbb{E}[\chi^2_{j,\eta,\tau}|\mathcal{C}_{\eta-}] - v(\eta,\tau)) = o_p(1)$ and $N_n^{-1} \sum_{j=1}^{N_n} (\mathbb{E}[\chi^2_{j,\eta,\tau}|\mathcal{C}_{\eta-}] - v(\eta,\tau))^2 = o_p(1)$.

(v) $N_n^{-1} \sum_{j=1}^{N_n} Var[\chi^2_{j,\eta,\tau}|\mathcal{C}_{\eta-}] \xrightarrow{\mathbb{P}} \Sigma_S$.

The information set $\mathcal{C}$ mentioned in Assumption 9 can be considered as a "common-shock" σ-field following the terminology of Andrews (2005). Intuitively, $\mathcal{C}$ contains systematic information through which various variables are strongly dependent in the cross-section. Along with this intuition, condition (i) of Assumption 9 suggests that the asset Z and the diffusive factor f are part of the "common shocks." Conditions (ii) and (iii) further impose conditional independence on the innovation terms $\chi_{j,\eta,\tau}$, formalizing the sense in which these innovations constitute "noise." Note that these conditions do not rule out unconditional dependence. Condition (iv) allows $\chi^2_{j,\eta,\tau}$ to be heterogeneous in mean, but "moderately" so. In particular, the cross-sectional average of the conditional means has a well-defined limit $v(\eta,\tau)$, which is the limit of the estimator $\hat{v}_n$ as we show below. Of course, this condition holds automatically if the conditional mean $\mathbb{E}[\chi^2_{j,\eta,\tau}|\mathcal{C}_{\eta-}]$ does not depend on j. It is perhaps possible to relax this condition to a more general version, but we leave such an extension for future work. Condition (v) imposes a mild requirement that the average finite-sample conditional variance has a limit.

Theorem 4, below, characterizes the asymptotic distribution of $\hat{v}_n$ under the general noisy jump factor model.

Theorem 4. *Suppose that Assumptions 3, 4 and 9 hold and* $(N_n^{1/2}+\Delta_n^{-1/2})q_n^w \to 0$. *The following statement hold:*

(a) if $N_n\Delta_n \to \theta$ *for some constant* $\theta \in [0,\infty)$, *then the sequence* $N_n^{1/2}(\hat{v}_n - v(\eta,\tau))$ *converges* $\mathcal{C}_{\eta-}$*-stably in law to*

$$\mathcal{S} + \frac{2\sqrt{\theta}v(\eta,\tau)(\sqrt{\kappa_\tau}\zeta'_{\tau-} + \sqrt{1-\kappa_\tau}\zeta'_{\tau+})}{\Delta Z_\tau}, \tag{8.3}$$

where the variables $\mathcal{S}$, κ_τ, $\zeta'_{\tau-}$ *and* $\zeta'_{\tau+}$ *are* $\mathcal{C}_{\eta-}$*-conditionally independent,* κ_τ *is uniformly distributed on the unit interval, and* $\mathcal{S}$ *and* $\zeta'_{\tau\pm}$ *are centered Gaussian with conditional variances* Σ_S *and* $\Sigma_{Z,\tau\pm} \equiv \lambda_{Z,\tau\pm}^\top \Sigma_{f,\tau\pm}\lambda_{Z,\tau\pm}$*, respectively;*

(b) if $N_n\Delta_n \to \infty$, *then the sequence* $\Delta_n^{-1/2}(\hat{v}_n - v(\eta,\tau))$ *converges* $\mathcal{F}$*-stably in law to*

$$\frac{2v(\eta,\tau)(\sqrt{\kappa_\tau}\zeta'_{\tau-} + \sqrt{1-\kappa_\tau}\zeta'_{\tau+})}{\Delta Z_\tau}.$$

Comments. The limiting variable in (8.3) consists of two components. The first component $\mathcal{S}$ captures the sampling variability in the infeasible statistic $N_n^{-1}\sum_{j=1}^{N_n}\chi^2_{j,\eta,\tau}$ which would be attained if the jump betas $\beta_{j,\tau}$ were directly observed. The second component accounts for the estimation error in the jump beta estimates. Although the estimation errors are aggregated in the cross-section, they still have an asymptotic effect because of their strong cross-sectional dependence driven by the systematic diffusive factors; by contrast, estimation errors due to idiosyncratic disturbances are "averaged out." It is interesting to note that the second term in (8.3) only involves the estimation error of the jump beta at the (later) jump time τ. Finally, we note that part (a) and part (b) cover all cases for the relative size of N_n and Δ_n. The "balanced" case with $\theta \in (0,\infty)$ is the most interesting one because the limiting distribution captures both cross-sectional and high-frequency sampling variabilities.

We finish this section with describing how to build confidence intervals for the average variance $v(\eta,\tau)$ of the innovation terms based on the above theorem. We focus on the case with $\theta \in (0,\infty)$ so that the two sources of sampling variabilities are both accounted for. From a practical point of view, it is useful to consider more generally the estimation of $g(v(\eta,\tau))$ for some continuously differentiable function $g(\cdot)$. We estimate Σ_S and spot variances $\Sigma_{Z,\tau\pm}$ using, respectively,

$$\begin{cases} \hat{\Sigma}_{S,n} \equiv \dfrac{1}{N_n}\displaystyle\sum_{j=1}^{N_n}(|\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}| \wedge \bar{B}_{n,\eta,\tau})^4 - \hat{v}_n^2, \\ \hat{\Sigma}_{Z,n,\tau\pm} \equiv \dfrac{1}{k_n\Delta_n}\displaystyle\sum_{l=1}^{k_n}\left(\Delta^n_{i(n,\tau\pm)+l}Z\right)^2. \end{cases}$$

Algorithm 2 below provides a simulation-based method for estimating the limit distribution in (8.3), which is theoretically justified by Theorem 5.

Algorithm 2. Step 1. Simulate

$$\tilde{\kappa}_\tau \sim \text{ Uniform}\,[0,1]\,, \quad \tilde{S}_n \sim \mathcal{N}(0, \hat{\Sigma}_{S,n}), \quad \tilde{\zeta}'_{n,\tau\pm} \sim \mathcal{N}(0, \hat{\Sigma}_{Z,n,\tau\pm}).$$

Step 2. Compute

$$\tilde{v}_n \equiv \tilde{S}_n + \frac{2\sqrt{N_n \Delta_n}\hat{v}_n(\sqrt{\tilde{\kappa}_{\tau-}}\tilde{\zeta}'_{n,\tau-} + \sqrt{1-\tilde{\kappa}_\tau}\tilde{\zeta}'_{n,\tau+})}{\Delta^n_{i(n,\tau)}Z}.$$

Step 3. Repeat Step 1 and Step 2 for a large number of times, and use the Monte Carlo distribution of $g'(\hat{v}_n)\tilde{v}_n$ as an approximation of the limit distribution of $N_n^{1/2}\left(g(\hat{v}_n) - g(v(\eta,\tau))\right)$. □

Theorem 5. *Under the same conditions as Theorem 4(a), we have the following:*

(a) the conditional law of $\tilde{v}_n$ given data converges in probability to the $\mathcal{C}_{\eta-}$-conditional law of $\mathcal{S} + 2\sqrt{\theta}v(\eta,\tau)(\sqrt{\kappa_\tau}\zeta'_{\tau-} + \sqrt{1-\kappa_\tau}\zeta'_{\tau+})/\Delta Z_\tau$ under the uniform metric;

(b) consequently, with $\hat{q}_{n,\alpha}$ being the $(1-\alpha/2)$ conditional quantile of $|g'(\hat{v}_n)\tilde{v}_n|$ given data, the two-sided confidence interval $CI_n = [g(\hat{v}_n) - N_n^{-1/2}\hat{q}_{n,\alpha}, g(\hat{v}_n) + N_n^{-1/2}\hat{q}_{n,\alpha}]$ for $g(v(\eta,\tau))$ has asymptotic level $1-\alpha$.

Comment. Part (a) of Theorem 5 shows that the simulated distribution of $\tilde{v}_n$ consistently approximates the limit distribution of $\sqrt{N_n}(\hat{v}_n - v(\eta,\tau))$. As a result, the simulated conditional quantiles of $\tilde{v}_n$ converge in probability to those of the limit distribution. Part (b) specializes this result to establish the validity of a symmetric two-sided confidence interval. Other types of confidence intervals can be justified similarly using part (a).

Finally, Algorithm 2 can be modified to simulate the asymptotic distribution of the normalized mixed-scale estimator $N_n^{1/2}(\hat{v}_n^{(k)} - v(\eta,\tau))$, recall Section 4.2. The modification is to replace $\tilde{v}_n$ with

$$\tilde{v}_n^{(k)} \equiv \tilde{S}_n + \frac{2\sqrt{N_n \Delta_n}\hat{v}_n^{(k)}(\sqrt{\tilde{\kappa}_\tau}\tilde{\zeta}'_{n,\tau-} + \sqrt{k-\tilde{\kappa}_\tau}\tilde{\zeta}'_{n,\tau+})}{\Delta^n_{i(n,\tau),k}Z}.$$

Confidence intervals for $v(\eta,\tau)$ can then be constructed using simulated quantiles of $\tilde{v}_n^{(k)}$ similar to the procedure described in Theorem 5(b) above.

8.2 Proofs

Throughout the proofs, we use K to denote a generic constant that may change from line to line. For a sub σ-field $\mathcal{G} \subseteq \mathcal{F}$ and a sequence X_n of random variables, we write $X_n \xrightarrow{\mathcal{L}|\mathcal{G}} X$ if the $\mathcal{G}$-conditional law of X_n converges in probability to that of X under a metric that is associated with the weak convergence of probability measures. By a standard localization procedure, we can strengthen Assumption 3 as the following without loss of generality:

Assumption 10. *We have Assumption 3 holds with $T_1 = \infty$. Moreover, the processes α_j, λ_j, λ_Z, b_f, σ_f, $\tilde{\sigma}_j^2$ and $\tilde{J}_{Y,j}$ are bounded, uniformly in j.*

8.2.1 Preliminary Results

In this subsection, we introduce some notations and preliminary estimates that are used in the sequel. We consider a sequence Ω_n of random events defined by $\Omega_n \equiv \{$distinct jump times of the Poisson process $t \mapsto \mu([0,t], E)$ are at least $2k_n\Delta_n$ apart$\}$. Since $k_n\Delta_n \to 0$ and the jumps of Z is of finite activity, $\mathbb{P}(\Omega_n) \to 1$. Therefore, we can restrict our calculations to Ω_n without loss of generality. It is (notationally) convenient to extend the definition of the spot jump beta to all $t \in [0,T]$ such that, on each path, $\beta_{j,t} = \beta_{j,\tau}$ for $t \in [\tau - k_n\Delta_n, \tau + k_n\Delta_n]$. This extension is well-behaved on Ω_n and our analysis only concerns the behavior of $\beta_{j,t}$ around shrinking neighborhoods around the jump times. (It should be noted that $\beta_{j,t}$ defined as such is not adapted to $\mathcal{F}_t$.)

We also consider the following sequence of events:

$$\Omega_n' \equiv \left\{ \sum_{j=1}^{N_n} 1_{\left\{\Delta^n_{i(n,\tau)}\tilde{J}_{Y,j} \neq 0\right\}} \leq \lfloor N_n q_n^w \rfloor /2 \text{ for some } \tau \in \mathcal{T} \right\}. \tag{8.4}$$

By Markov's inequality,

$$\mathbb{P}\left(\sum_{j=1}^{N_n} 1_{\left\{\Delta^n_{i(n,\tau)}\tilde{J}_{Y,j} \neq 0\right\}} > \lfloor N_n q_n^w \rfloor /2 \right) \leq \frac{2}{\lfloor N_n q_n^w \rfloor} \sum_{j=1}^{N_n} \mathbb{P}\left(\Delta^n_{i(n,\tau)}\tilde{J}_{Y,j} \neq 0\right) \leq K\Delta_n/q_n^w \to 0.$$

Since $\mathcal{T}$ is finite, we have $\mathbb{P}(\Omega_n') \to 1$.

We denote the continuous part of Y_j and Z as, respectively,

$$Y_{j,t}' \equiv \int_0^t \alpha_{j,u} du + \int_0^t \lambda_{j,u}^\top df_u + \epsilon_{j,t}, \quad Z_t' \equiv \int_0^t \lambda_{Z,u}^\top df_u. \tag{8.5}$$

The diffusive residual process is then defined as

$$\widetilde{Y}_{j,t}' \equiv Y_{j,t}' - \beta_{j,t} Z_t' = \int_0^t \alpha_{j,u} du + \left(\int_0^t \lambda_{j,u}^\top df_u - \beta_{j,t} \int_0^t \lambda_{Z,u}^\top df_u \right) + \epsilon_{j,t}. \tag{8.6}$$

We denote

$$\xi_{n,j,s} \equiv \frac{\Delta^n_{i(n,s)} \widetilde{Y}_j'}{\Delta^n_{i(n,s)} Z}, \tag{8.7}$$

which can be decomposed as

$$\xi_{n,j,s} = \xi_{n,j,s}' + \xi_{n,j,s}'', \tag{8.8}$$

where

$$\begin{cases} \xi'_{n,j,s} \equiv \dfrac{1}{\Delta Z_s}\left(\tilde{\lambda}_{j,s-}^{\top}\left(f_s - f_{(i(n,s)-1)\Delta_n}\right) + \tilde{\lambda}_{j,s}^{\top}\left(f_{i(n,s)\Delta_n} - f_s\right) + \epsilon_{i(n,s)\Delta_n} - \epsilon_{(i(n,s)-1)\Delta_n}\right), \\ \xi''_{n,j,s} \equiv \dfrac{1}{\Delta Z_s}\displaystyle\int_{(i(n,s)-1)\Delta_n}^{i(n,s)\Delta_n} \alpha_{j,u} du + \Delta_{i(n,s)}^{n}\widetilde{Y}'_j\left(\dfrac{1}{\Delta_{i(n,s)}^{n} Z} - \dfrac{1}{\Delta Z_s}\right) \\ \qquad + \dfrac{1}{\Delta Z_s}\displaystyle\int_{(i(n,s)-1)\Delta_n}^{s} \left(\lambda_{j,u} - \lambda_{j,s-}\right)^{\top} df_u - \dfrac{\beta_{j,s}}{\Delta Z_s}\displaystyle\int_{(i(n,s)-1)\Delta_n}^{s} \left(\lambda_{Z,u} - \lambda_{Z,s-}\right)^{\top} df_u \\ \qquad + \dfrac{1}{\Delta Z_s}\displaystyle\int_{s}^{i(n,s)\Delta_n} \left(\lambda_{j,u} - \lambda_{j,s}\right)^{\top} df_u - \dfrac{\beta_{j,s}}{\Delta Z_s}\displaystyle\int_{s}^{i(n,s)\Delta_n} \left(\lambda_{Z,u} - \lambda_{Z,s}\right)^{\top} df_u. \end{cases} \tag{8.9}$$

We further rewrite $\xi'_{n,j,s}$ as

$$\xi'_{n,j,s} = \Delta_n^{1/2} \sum_{q\in\{s-,s+\}} w_{n,q}\left(\tilde{\lambda}_{j,q}^{\top}\Sigma_{f,q}^{1/2}\zeta_{n,q} + R_{n,j,q}\right), \tag{8.10}$$

where we define

$$\begin{cases} \zeta_{n,s-} \equiv \Sigma_{f,s-}^{-1/2}\dfrac{f_s - f_{(i(n,s)-1)\Delta_n}}{\sqrt{s-(i(n,s)-1)\Delta_n}}, & \zeta_{n,s+} \equiv \Sigma_{f,s}^{-1/2}\dfrac{f_{i(n,s)\Delta_n} - f_s}{\sqrt{i(n,s)\Delta_n - s}}, \\ R_{n,j,s-} \equiv \dfrac{\epsilon_{j,s} - \epsilon_{j,(i(n,s)-1)\Delta_n}}{\sqrt{s-(i(n,s)-1)\Delta_n}}, & R_{n,j,s+} \equiv \dfrac{\epsilon_{j,i(n,s)\Delta_n} - \epsilon_{j,s}}{\sqrt{i(n,s)\Delta_n - s}}, \\ w_{n,s-} \equiv \dfrac{1}{\Delta Z_s}\sqrt{\dfrac{s-(i(n,s)-1)\Delta_n}{\Delta_n}}, & w_{n,s+} \equiv \dfrac{1}{\Delta Z_s}\sqrt{\dfrac{i(n,s)\Delta_n - s}{\Delta_n}}. \end{cases} \tag{8.11}$$

Lemma 1. *Under Assumptions 3 and 4, we have for $p, q \in \{\tau-, \tau+, \eta-, \eta+\}$:*

(a) $N_n^{-1}\sum_{j=1}^{N_n} R_{n,j,p}R_{n,j,q} = O_p(N_n^{-1/2})$ *when* $p \neq q$*;*

(b) $N_n^{-1}\sum_{j=1}^{N_n} R_{n,j,q}^2 \xrightarrow{\mathbb{P}} M_\epsilon(q)$*;*

(c) $N_n^{-1}\sum_{j=1}^{N_n} R_{n,j,p}\tilde{\lambda}_{j,q} = O_p(N_n^{-1/2})$*;*

(d) $N_n^{-1}\Delta_n^{-1}\sum_{j=1}^{N_n}(\xi''_{n,j,\tau} - \xi''_{n,j,\eta})^2 = O_p(\Delta_n)$.

Proof of Lemma 1. (a) We prove the case with $p = \tau-$ and $q = \tau+$ in detail, while noting that the other cases can be proved in exactly the same way. Note that the jump times of the Poisson measure μ are necessarily independent of the Brownian motions $\widetilde{W}_j$, $1 \leq j \leq N_n$. Let $\mathcal{G}_t$ be the smallest filtration such that $\mathcal{F}_t \subseteq \mathcal{G}_t$ and the jump times of μ are $\mathcal{G}_t$-measurable. The processes $(\widetilde{W}_j)_{1\leq j\leq N_n}$ remain to be Brownian motions with respect to $(\mathcal{G}_t)_{t\geq 0}$. Consequently, ϵ_j is a $(\mathcal{G}_t)_{t\geq 0}$-martingale and, hence,

$$\mathbb{E}[R_{n,j,\tau-}R_{n,j,\tau+}] = 0. \tag{8.12}$$

Moreover, for $j \neq m$,

$$\mathbb{E}\left[R_{n,j,\tau-}R_{n,j,\tau+}R_{n,m,\tau-}R_{n,m,\tau+}\right] = \mathbb{E}\left[R_{n,j,\tau-}R_{n,m,\tau-}\mathbb{E}\left[R_{n,j,\tau+}R_{n,m,\tau+}|\mathcal{G}_\tau\right]\right] = 0, \tag{8.13}$$

where the first equality holds because $R_{n,j,\tau-}R_{n,m,\tau-}$ is $\mathcal{G}_\tau$-measurable and the second equality holds because $\widetilde{W}_j$ and $\widetilde{W}_m$ are orthogonal. Since the processes $\tilde{\sigma}_j^2$ are uniformly bounded,

$\mathbb{E}[R_{n,j,\tau\pm}^4] \leq K$ holds due to a standard estimate for continuous Itô processes. By the Cauchy–Schwarz inequality, this further implies that

$$\mathbb{E}\left[R_{n,j,\tau-}^2 R_{n,j,\tau+}^2\right] \leq K. \tag{8.14}$$

From (8.12), (8.13) and (8.14), we deduce

$$\mathbb{E}\left[\left(N_n^{-1}\sum_{j=1}^{N_n} R_{n,j,p}R_{n,j,q}\right)^2\right] \leq K N_n^{-1}.$$

The assertion of part (a) then readily follows.

(b) We consider the case $q = \tau-$ in details while noting that the other cases can be proved in the same way. By using Itô's formula, we can decompose

$$\begin{aligned}
&R_{n,j,\tau-}^2 = U_{n,j} + U'_{n,j}, \quad \text{where} \\
&U_{n,j} \equiv \frac{1}{\tau - (i\,(n,\tau) - 1)\,\Delta_n} \int_{(i(n,\tau)-1)\Delta_n}^{\tau} \tilde{\sigma}_{j,u}^2 du, \\
&U'_{n,j} \equiv \frac{2}{\tau - (i\,(n,\tau) - 1)\,\Delta_n} \int_{(i(n,\tau)-1)\Delta_n}^{\tau} (\epsilon_{j,u} - \epsilon_{j,(i(n,\tau)-1)\Delta_n}) d\widetilde{W}_{j,u}.
\end{aligned}$$

We note that $\mathbb{E}[U'_{n,j}] = 0$ for each j and $\mathbb{E}[U'_{n,j}U'_{n,m}] = 0$ for $j \neq m$. In addition, $\mathbb{E}|U'_{n,j}|^2 \leq K$. From these estimates, it readily follows that

$$\frac{1}{N_n}\sum_{j=1}^{N_n} U'_{n,j} = O_p(N_n^{-1/2}). \tag{8.15}$$

Next, note that by Assumption 3(v),

$$\mathbb{E}\left|U_{n,j} - \tilde{\sigma}_{j,\tau-}^2\right| \leq \mathbb{E}\left[\sup_{s,t,|s-t|\leq\Delta_n} \left|\tilde{\sigma}_{j,s}^2 - \tilde{\sigma}_{j,t}^2\right|\right] \leq K\Delta_n^{1/2}.$$

From this estimate and Assumption 4, we deduce

$$\frac{1}{N_n}\sum_{j=1}^{N_n} U_{n,j} = \frac{1}{N_n}\sum_{j=1}^{N_n} \tilde{\sigma}_{j,\tau-}^2 + o_p(1) = M_\epsilon(\tau-) + o_p(1). \tag{8.16}$$

The assertion of part (b) then follows from (8.15) and (8.16).

(c) By Assumption 4, $\tilde{\lambda}_{j,q}$ is conditionally independent of $R_{n,j,p}$ and, hence, $\mathbb{E}[R_{n,j,p}\tilde{\lambda}_{j,q}|\mathcal{G}_0] = \mathbb{E}[R_{n,j,p}|\mathcal{G}_0]\mathbb{E}[\tilde{\lambda}_{j,q}|\mathcal{G}_0] = 0$. In addition, for $j \neq m$,

$$\mathbb{E}[R_{n,j,p}R_{n,m,p}\tilde{\lambda}_{j,q}\tilde{\lambda}_{m,q}^\top|\mathcal{G}_0] = \mathbb{E}[R_{n,j,p}R_{n,m,p}|\mathcal{G}_0]\mathbb{E}[\tilde{\lambda}_{j,q}\tilde{\lambda}_{m,q}^\top|\mathcal{G}_0] = 0,$$

where the second equality follows from the orthogonality between $\widetilde{W}_j$ and $\widetilde{W}_m$. Since $\tilde{\lambda}_{j,q}$ is bounded, $R_{n,j,p}\tilde{\lambda}_{j,q}$ has bounded second moment. The assertion of part (c) readily follows from these facts.

(d) First, since the α_j's are uniformly bounded, it is easy to see that $(\Delta Z_s)^{-1}\int_{(i(n,s)-1)\Delta_n}^{i(n,s)\Delta_n}\alpha_{j,u}du = O_p(\Delta_n)$ uniformly in j. Hence,

$$\frac{1}{N_n\Delta_n}\sum_{j=1}^{N_n}\left(\frac{1}{\Delta Z_s}\int_{(i(n,s)-1)\Delta_n}^{i(n,s)\Delta_n}\alpha_{j,u}du\right)^2 = O_p(\Delta_n). \tag{8.17}$$

Further note that, uniformly in j, we have $\mathbb{E}|\Delta_{i(n,s)}^n\widetilde{Y}_j'|^2 \le K\Delta_n$ and, hence,

$$\frac{1}{N_n\Delta_n}\sum_{j=1}^{N_n}(\Delta_{i(n,s)}^n\widetilde{Y}_j')^2 = O_p(1). \tag{8.18}$$

It is easy to see that

$$\frac{1}{\Delta_{i(n,s)}^n Z} - \frac{1}{\Delta Z_s} = O_p(\Delta_n^{1/2}). \tag{8.19}$$

From (8.18) and (8.19), we deduce

$$\frac{1}{N_n\Delta_n}\sum_{j=1}^{N_n}\left(\Delta_{i(n,s)}^n\widetilde{Y}_j'\left(\frac{1}{\Delta_{i(n,s)}^n Z} - \frac{1}{\Delta Z_s}\right)\right)^2 = O_p(\Delta_n). \tag{8.20}$$

We then note that, since the processes λ_j's are (1/2)-Hölder continuous under L_2-norm uniformly in j (Assumption 3(v)), the following estimate also holds uniformly

$$\mathbb{E}\left[\left(\int_{(i(n,s)-1)\Delta_n}^{s}(\lambda_{j,u}-\lambda_{j,s-})^\top df_u\right)^2\right] \le K\Delta_n^2.$$

Hence,

$$\frac{1}{N_n\Delta_n}\sum_{j=1}^{N_n}\left(\frac{1}{\Delta Z_s}\int_{(i(n,s)-1)\Delta_n}^{s}(\lambda_{j,u}-\lambda_{j,s-})^\top df_u\right)^2 = O_p(\Delta_n). \tag{8.21}$$

Similarly,

$$\begin{cases} \dfrac{1}{N_n\Delta_n}\displaystyle\sum_{j=1}^{N_n}\left(\frac{\beta_{j,s}}{\Delta Z_s}\int_{(i(n,s)-1)\Delta_n}^{s}(\lambda_{Z,u}-\lambda_{Z,s-})^\top df_u\right)^2 = O_p(\Delta_n), \\ \dfrac{1}{N_n\Delta_n}\displaystyle\sum_{j=1}^{N_n}\left(\frac{1}{\Delta Z_s}\int_{s}^{i(n,s)\Delta_n}(\lambda_{j,u}-\lambda_{j,s})^\top df_u\right)^2 = O_p(\Delta_n), \\ \dfrac{1}{N_n\Delta_n}\displaystyle\sum_{j=1}^{N_n}\left(\frac{\beta_{j,s}}{\Delta Z_s}\int_{s}^{i(n,s)\Delta_n}(\lambda_{Z,u}-\lambda_{Z,s})^\top df_u\right)^2 = O_p(\Delta_n). \end{cases} \tag{8.22}$$

With an appeal to the Cauchy–Schwarz inequality, the assertion of part (d) then follows from (8.17), (8.20), (8.21) and (8.22). *Q.E.D.*

Next, we set

$$\begin{cases} A_n(s) \equiv \dfrac{1}{N_n\Delta_n}\displaystyle\sum_{j=1}^{N_n}\left(\xi_{n,j,s}'\right)^2, \quad s\in\{\eta,\tau\}, \\ B_n(\eta,\tau) \equiv \dfrac{1}{N_n\Delta_n}\displaystyle\sum_{j=1}^{N_n}\xi_{n,j,\eta}'\xi_{n,j,\tau}'. \end{cases} \tag{8.23}$$

The following lemma collects some convergence results that we use for deriving limiting distributions.

Lemma 2. *Suppose that Assumptions 3 and 4 hold. Then,*

$$(A_n(\eta), A_n(\tau), B_n(\eta,\tau)) \xrightarrow{\mathcal{L}\text{-}s} (\mathcal{A}(\eta), \mathcal{A}(\tau), \mathcal{B}(\eta,\tau)),$$

where $\xrightarrow{\mathcal{L}\text{-}s}$ *denotes* $\mathcal{F}$*-stable convergence in law.*

Proof of Lemma 2. By Theorem 4.3.1 in Jacod and Protter (2012),

$$(w_{n,q}, \zeta_{n,q})_{q\in\{\eta-,\eta+,\tau-,\tau+\}} \xrightarrow{\mathcal{L}\text{-}s} (w_q, \zeta_q)_{q\in\{\eta-,\eta+,\tau-,\tau+\}}. \tag{8.24}$$

Recall the definitions in (8.10) and (8.23). We have, for $s \in \{\eta, \tau\}$,

$$\begin{aligned} A_n(s) &= \frac{1}{N_n}\sum_{j=1}^{N_n}\left(\sum_{q\in\{s-,s+\}} w_{n,q}\left(\tilde{\lambda}_{j,q}^{\top}\Sigma_{f,q}^{1/2}\zeta_{n,q} + R_{n,j,q}\right)\right)^2 \\ &= \sum_{p,q\in\{s-,s+\}} w_{n,p}w_{n,q}\zeta_{n,p}^{\top}\Sigma_{f,p}^{1/2}\left(\frac{1}{N_n}\sum_{j=1}^{N_n}\tilde{\lambda}_{j,p}\tilde{\lambda}_{j,q}^{\top}\right)\Sigma_{f,q}^{1/2}\zeta_{n,q} \\ &\quad + \sum_{q\in\{s-,s+\}} w_{n,q}^2\left(\frac{1}{N_n}\sum_{j=1}^{N_n}R_{n,j,q}^2\right) + O_p(N_n^{-1/2}), \end{aligned} \tag{8.25}$$

where the rate for the $O_p(N_n^{-1})$ term in the last line is obtained using Lemma 1(a,c). Similarly,

$$\begin{aligned} &B_n(\eta,\tau) \\ &= \frac{1}{N_n}\sum_{j=1}^{N_n}\left(\sum_{p\in\{\tau-,\tau+\}} w_{n,p}\left(\tilde{\lambda}_{j,p}^{\top}\Sigma_{f,p}^{1/2}\zeta_{n,p} + R_{n,j,p}\right)\right) \\ &\qquad\qquad \times\left(\sum_{q\in\{\eta-,\eta+\}} w_{n,q}\left(\tilde{\lambda}_{j,q}^{\top}\Sigma_{f,q}^{1/2}\zeta_{n,q} + R_{n,j,q}\right)\right) \\ &= \sum_{p\in\{\tau-,\tau+\}}\sum_{q\in\{\eta-,\eta+\}} w_{n,p}w_{n,q}\zeta_{n,p}^{\top}\Sigma_{f,p}^{1/2}\left(\frac{1}{N_n}\sum_{j=1}^{N_n}\tilde{\lambda}_{j,p}\tilde{\lambda}_{j,q}^{\top}\right)\Sigma_{f,q}^{1/2}\zeta_{n,q} + O_p(N_n^{-1/2}). \end{aligned} \tag{8.26}$$

By Assumption 4 and Lemma 1(b),

$$\Sigma_{f,p}^{1/2}\left(\frac{1}{N_n}\sum_{j=1}^{N_n}\tilde{\lambda}_{j,p}\tilde{\lambda}_{j,q}^{\top}\right)\Sigma_{f,q}^{1/2} \xrightarrow{\mathbb{P}} M_C(p,q), \quad \frac{1}{N_n}\sum_{j=1}^{N_n}R_{n,j,q}^2 \xrightarrow{\mathbb{P}} M_\epsilon(q). \tag{8.27}$$

We further note that the limiting variables $M_C(p,q)$ and $M_\epsilon(q)$ are $\mathcal{F}$-measurable. Hence, by the property of stable convergence in law, we can deduce the assertion of Lemma 2 from (8.24), (8.25), (8.26) and (8.27). *Q.E.D.*

Finally, we show in Lemma 3 some consistency results for the spot jump beta estimates.

Lemma 3. *Under Assumptions 3 and 5, the following holds for* $s \in \mathcal{T}$*:*

(a) $\sup_{1\leq j\leq N_n} |\hat{\beta}_{n,j,s} - \beta_{j,s}| 1_{\{\Delta^n_{i(n,s)} \tilde{J}_{Y,j}=0\}} = o_p(1)$*;*

(b) $N_n^{-1} \sum_{j=1}^{N_n} |\hat{\beta}_{n,j,\tau} - \beta_{j,\tau}|^2 = o_p(1)$*.*

Proof of Lemma 3. (a) Note that

$$\hat{\beta}_{n,j,s} - \beta_{j,s} = \frac{\Delta^n_{i(n,s)} Y_j - \beta_{j,s} \Delta^n_{i(n,s)} Z}{\Delta^n_{i(n,s)} Z}. \tag{8.28}$$

By localization, we can assume that $\tilde{\sigma}_j^2$, Σ_f, λ_j and β_j are bounded. By a standard estimate for continuous Itô semimartingales (applied to the continuous parts of Y_j and Z), we have for any $p \geq 1$,

$$\mathbb{E}\left[|\Delta^n_{i(n,s)} Y_j - \beta_{j,s} \Delta^n_{i(n,s)} Z|^p 1_{\{\Delta^n_{i(n,s)} \tilde{J}_{Y,j}=0\}} \right] \leq K_p \Delta_n^{p/2},$$

for some constant K_p. By using a maximal inequality, we deduce that

$$\sup_{1\leq j\leq N_n} |\Delta^n_{i(n,s)} Y_j - \beta_{j,s} \Delta^n_{i(n,s)} Z| 1_{\{\Delta^n_{i(n,s)} \tilde{J}_{Y,j}=0\}} = O_p(\Delta_n^{1/2} N_n^{\iota}) \tag{8.29}$$

for some arbitrarily small (but fixed) constant $\iota > 0$. Then, by Assumption 5,

$$\sup_{1\leq j\leq N_n} |\Delta^n_{i(n,s)} Y_j - \beta_{j,s} \Delta^n_{i(n,s)} Z| 1_{\{\Delta^n_{i(n,s)} \tilde{J}_{Y,j}=0\}} = o_p(1).$$

Note that $1/\Delta^n_{i(n,s)} Z = O_p(1)$. The assertion of the lemma then readily follows from the above estimate and equation (8.28).

(b) It is easy to see that $\hat{\beta}_{n,j,\tau}$, $1 \leq j \leq N_n$, are uniformly bounded with probability approaching one. We then note that

$$\begin{aligned}
&\frac{1}{N_n} \sum_{j=1}^{N_n} |\hat{\beta}_{n,j,\tau} - \beta_{j,\tau}|^2 \\
&= \frac{1}{N_n} \sum_{j=1}^{N_n} |\hat{\beta}_{n,j,\tau} - \beta_{j,\tau}|^2 1_{\{\Delta^n_{i(n,s)} \tilde{J}_{Y,j}=0\}} + \frac{1}{N_n} \sum_{j=1}^{N_n} |\hat{\beta}_{n,j,\tau} - \beta_{j,\tau}|^2 1_{\{\Delta^n_{i(n,s)} \tilde{J}_{Y,j}\neq 0\}} \\
&\leq \left(\sup_{1\leq j\leq N_n} |\hat{\beta}_{n,j,s} - \beta_{j,s}| 1_{\{\Delta^n_{i(n,s)} \tilde{J}_{Y,j}=0\}} \right)^2 + \frac{K}{N_n} \sum_{j=1}^{N_n} 1_{\{\Delta^n_{i(n,s)} \tilde{J}_{Y,j}\neq 0\}} \\
&= o_p(1),
\end{aligned}$$

as claimed in part (b). *Q.E.D.*

8.2.2 Proof of Proposition 1

Proof of Proposition 1. Recall that the spot jump betas $\beta_{j,s}$ are bounded by assumption. By Lemma 3 and the boundedness of $\tilde{J}_{Y,j}$, we further deduce that the beta estimates $\hat{\beta}_{n,j,s}$ are uniformly (in j) bounded with probability approaching one. Since the loss function $L(\cdot)$ is Lipschitz on bounded sets (Assumption 1), we can now assume that $L(\cdot)$ is globally Lipschitz without loss of generality. Hence, by Lemma 3,

$$\begin{aligned}
&\frac{1}{N_n}\sum_{j=1}^{N_n}\left|L(\hat{\beta}_{n,j,\tau}-\hat{\beta}_{n,j,\eta})-L(\chi_{j,\eta,\tau})\right| \\
&\quad\leq \frac{1}{N_n}\sum_{j=1}^{N_n}\left|L(\hat{\beta}_{n,j,\tau}-\hat{\beta}_{n,j,\eta})-L(\chi_{j,\eta,\tau})\right| 1_{\left\{\left|\Delta^n_{i(n,\tau)}\tilde{J}_{Y,j}\right|+\left|\Delta^n_{i(n,\eta)}\tilde{J}_{Y,j}\right|=0\right\}} \\
&\qquad+\frac{K}{N_n}\sum_{j=1}^{N_n} 1_{\left\{\left|\Delta^n_{i(n,\tau)}\tilde{J}_{Y,j}\right|+\left|\Delta^n_{i(n,\eta)}\tilde{J}_{Y,j}\right|>0\right\}} \\
&\quad\leq K\max_{s\in\{\eta,\tau\},1\leq j\leq N_n}\left|\hat{\beta}_{n,j,s}-\beta_{j,s}\right| 1_{\left\{\Delta^n_{i(n,s)}\tilde{J}_{Y,j}=0\right\}}+O_p(\Delta_n)=o_p(1).
\end{aligned} \tag{8.30}$$

Next, we set

$$\xi_n=\frac{1}{N_n}\sum_{j=1}^{N_n}\left(L(\chi_{j,\eta,\tau})-\mathbb{E}[L(\chi_{j,\eta,\tau})|\mathcal{F}_{\eta-}]\right).$$

Under Assumption 6, ξ_n is the average of $\mathcal{F}_{\eta-}$-conditionally independent variables with zero conditional mean. Hence,

$$\begin{aligned}
\mathbb{E}\left[\xi_n^2|\mathcal{F}_{\eta-}\right] &= \frac{1}{N_n^2}\sum_{j=1}^{N_n}\mathbb{E}\left[\left(L(\chi_{j,\eta,\tau})-\mathbb{E}[L(\chi_{j,\eta,\tau})|\mathcal{F}_{\eta-}]\right)^2\middle|\mathcal{F}_{\eta-}\right] \\
&\leq \frac{1}{N_n^2}\sum_{j=1}^{N_n}\mathbb{E}\left[L(\chi_{j,\eta,\tau})^2|\mathcal{F}_{\eta-}\right]=O_p(N_n^{-1})=o_p(1).
\end{aligned}$$

In particular, this implies that $\mathbb{E}\left[|\xi_n|\wedge 1|\mathcal{F}_{\eta-}\right]=o_p(1)$. By the bounded convergence theorem, we further deduce $\mathbb{E}\left[|\xi_n|\wedge 1\right]\to 0$. But this is equivalent to $\xi_n=o_p(1)$. This, together with (8.30), implies that

$$\frac{1}{N_n}\sum_{j=1}^{N_n}L(\hat{\beta}_{n,j,\tau}-\hat{\beta}_{n,j,\eta})=\frac{1}{N_n}\sum_{j=1}^{N_n}\mathbb{E}[L(\chi_{j,\eta,\tau})|\mathcal{F}_{\eta-}]+o_p(1). \tag{8.31}$$

Since $q_n^w\to 0$, the winsorized estimator $\hat{V}_n$ differs from $N_n^{-1}\sum_{j=1}^{N_n}L(\hat{\beta}_{n,j,\tau}-\hat{\beta}_{n,j,\eta})$ by an $o_p(1)$ term. The assertion of the proposition then follows from (8.31). *Q.E.D.*

8.2.3 Proof of Theorem 1

Proof of Theorem 1. Step 1. The proof proceeds in two steps. Recall Ω'_n from (8.4). Since $\mathbb{P}(\Omega'_n)\to 1$, we can restrict our calculations to Ω'_n without loss of generality. In this step, we show that

$$\Delta_n^{-1}\hat{V}_n=\frac{1}{N_n\Delta_n}\sum_{j=1}^{N_n}L\left(\xi_{n,j,\tau}-\xi_{n,j,\eta}\right)+o_p(1). \tag{8.32}$$

From (8.29), we see that

$$\sup_{1\leq j\leq N_n}|\hat{\beta}_{n,j,s}-\beta_{j,s}|1_{\left\{\Delta^n_{i(n,s)}\tilde{J}_{Y,j}=0\right\}}=O_p(\Delta_n^{1/2}N_n^{\iota}) \tag{8.33}$$

for some fixed but arbitrarily small constant $\iota > 0$. In restriction to Ω'_n and the null hypothesis, $\bar{B}_{n,\eta,\tau}$ is bounded by two times of the left-hand of the above display. Hence,

$$\bar{B}_{n,\eta,\tau} = O_p(\Delta_n^{1/2} N_n^{\iota}). \tag{8.34}$$

We note that

$$\begin{aligned} &\left| \frac{1}{\Delta_n N_n} \sum_{j=1}^{N_n} \Big| L\left(|\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}| \wedge \bar{B}_{n,\eta,\tau} \right) \right. \\ &\qquad \left. - L\left(|\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}| \right) \Big| 1_{\left\{ \left| \Delta^n_{i(n,\tau)} \tilde{J}_{Y,j} \right| + \left| \Delta^n_{i(n,\eta)} \tilde{J}_{Y,j} \right| = 0 \right\}} \right| \\ &\leq \frac{\lceil q_n^w N_n \rceil}{\Delta_n N_n} \sup_{1 \leq j \leq N_n} L\left(|\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}| \right) 1_{\left\{ \left| \Delta^n_{i(n,\tau)} \tilde{J}_{Y,j} \right| + \left| \Delta^n_{i(n,\eta)} \tilde{J}_{Y,j} \right| = 0 \right\}} \\ &= O_p(q_n^w N_n^{2\iota}) = o_p(1), \end{aligned} \tag{8.35}$$

where the inequality follows from the fact that the winsorization is active for at most $\lceil q_n^w N_n \rceil$ terms ($\lceil \cdot \rceil$ denotes the ceiling function); the first equality follows from (8.33); the second equality follows from Assumptions 2 and 5 with ι chosen sufficiently small. Note that in restriction to $\{|\Delta^n_{i(n,\tau)} \tilde{J}_{Y,j}| + |\Delta^n_{i(n,\eta)} \tilde{J}_{Y,j}| = 0\}$ and the null hypothesis, $\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta} = \xi_{n,j,\tau} - \xi_{n,j,\eta}$. Hence,

$$\begin{aligned} &\frac{1}{\Delta_n N_n} \sum_{j=1}^{N_n} L\left(|\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}| \wedge \bar{B}_{n,\eta,\tau} \right) 1_{\left\{ \left| \Delta^n_{i(n,\tau)} \tilde{J}_{Y,j} \right| + \left| \Delta^n_{i(n,\eta)} \tilde{J}_{Y,j} \right| = 0 \right\}} \\ &\quad = \frac{1}{\Delta_n N_n} \sum_{j=1}^{N_n} L\left(\xi_{n,j,\tau} - \xi_{n,j,\eta} \right) 1_{\left\{ \left| \Delta^n_{i(n,\tau)} \tilde{J}_{Y,j} \right| + \left| \Delta^n_{i(n,\eta)} \tilde{J}_{Y,j} \right| = 0 \right\}} + o_p(1). \end{aligned} \tag{8.36}$$

Next, we note that

$$\begin{aligned} &\frac{1}{N_n \Delta_n} \sum_{j=1}^{N_n} L\left(|\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}| \wedge \bar{B}_{n,\eta,\tau} \right) 1_{\left\{ \left| \Delta^n_{i(n,\tau)} \tilde{J}_{Y,j} \right| + \left| \Delta^n_{i(n,\eta)} \tilde{J}_{Y,j} \right| > 0 \right\}} \\ &\quad \leq \frac{L\left(\bar{B}_{n,\eta,\tau} \right)}{N_n \Delta_n} \sum_{j=1}^{N_n} 1_{\left\{ \left| \Delta^n_{i(n,\tau)} \tilde{J}_{Y,j} \right| + \left| \Delta^n_{i(n,\eta)} \tilde{J}_{Y,j} \right| > 0 \right\}} \\ &\quad = O_p(\Delta_n N_n^{2\iota}) = o_p(1), \end{aligned} \tag{8.37}$$

where the inequality follows from the monotonicity of $L(\cdot)$ and the last line follows from (8.34) and the fact that $\mathbb{P}(\Delta^n_{i(n,s)} \tilde{J}_{Y,j} \neq 0) \leq K \Delta_n$. Similarly, we can show that

$$\frac{1}{N_n \Delta_n} \sum_{j=1}^{N_n} L\left(\xi_{n,j,\tau} - \xi_{n,j,\eta} \right) 1_{\left\{ \left| \Delta^n_{i(n,\tau)} \tilde{J}_{Y,j} \right| + \left| \Delta^n_{i(n,\eta)} \tilde{J}_{Y,j} \right| > 0 \right\}} = o_p(1). \tag{8.38}$$

From (8.36), (8.37) and (8.38), we deduce (8.32) as wanted.

Step 2. It remains to derive the convergence of $(N_n \Delta_n)^{-1} \sum_{j=1}^{N_n} L\left(\xi_{n,j,\tau} - \xi_{n,j,\eta} \right)$. Recall the definition of $\xi'_{n,j,s}$ from (8.9). Let L_n be defined as

$$L_n \equiv \frac{1}{N_n \Delta_n} \sum_{j=1}^{N_n} \left(\xi'_{n,j,\tau} - \xi'_{n,j,\eta} \right)^2. \tag{8.39}$$

Recalling the definitions in (8.23), we can rewrite L_n as

$$L_n = A_n(\eta) + A_n(\tau) - 2B_n(\eta, \tau). \tag{8.40}$$

Then, by Lemma 2,

$$L_n \xrightarrow{\mathcal{L}\text{-}s} \mathcal{L}(\eta, \tau) \equiv \mathcal{A}(\eta) + \mathcal{A}(\tau) - 2\mathcal{B}(\eta, \tau). \tag{8.41}$$

From (8.8), we further see that

$$\begin{aligned} &\frac{1}{N_n \Delta_n} \sum_{j=1}^{N_n} L(\xi_{n,j,\tau} - \xi_{n,j,\eta}) \\ &\quad = L_n + \frac{2}{N_n \Delta_n} \sum_{j=1}^{N_n} (\xi'_{n,j,\tau} - \xi'_{n,j,\eta})(\xi''_{n,j,\tau} - \xi''_{n,j,\eta}) + \frac{1}{N_n \Delta_n} \sum_{j=1}^{N_n} (\xi''_{n,j,\tau} - \xi''_{n,j,\eta})^2. \end{aligned} \tag{8.42}$$

By Lemma 1(d), the last term in (8.42) is $O_p(\Delta_n)$. By the Cauchy–Schwarz inequality, this estimate and (8.41) further imply that the second term on the right-hand side of (8.42) is $o_p(1)$. Therefore,

$$\frac{1}{N_n \Delta_n} \sum_{j=1}^{N_n} L(\xi_{n,j,\tau} - \xi_{n,j,\eta}) = L_n + o_p(1).$$

The assertion of the theorem then follows from (8.32) and (8.41). *Q.E.D.*

8.2.4 Proof of Theorem 2

We start with the proof of part (a) and part (b). We provide details for the case with $q = \tau-$, while noting that the case with $q = \tau+$ only requires a change of notation. Hence, we suppress (in most cases) the dependence on q in our notations for simplicity. More specifically, we write $\hat{X}_n$, $\hat{F}_n$, $\hat{\Lambda}_n$, Λ_n^*, $\mathcal{E}_n$, H, Σ_f, M_Λ^* and M_C^* in place of $\hat{X}_n(q)$, $\hat{F}_n(q)$, $\hat{\Lambda}_n(q)$, $\Lambda_n^*(q)$, $\mathcal{E}_n(q)$, H_q, $\Sigma_{f,q}$, $M_\Lambda^*(q,q)$ and $M_C^*(q,q)$, respectively. We denote the jth column of a generic matrix A by $A_{\cdot j}$. Recall the sequence Ω_n of events defined as in Section 8.2.1. Since $\mathbb{P}(\Omega_n) \to 1$, we can restrict our calculations below in Ω_n without loss of generality.

Below, we denote $\Gamma_n \equiv \{\gamma \in \mathbb{R}^{k_n} : \gamma^\top \gamma = k_n\}$. Note that each column of $\hat{F}_n$ is an element of Γ_n. We collect some useful estimates in Lemma 4, where we denote

$$\tilde{\Lambda}_n^* \equiv \left(\lambda_{1,\tau-} - \tilde{\beta}_{n,1,\tau}\lambda_{Z,\tau-}, \ldots, \lambda_{N_n,\tau-} - \tilde{\beta}_{n,N_n,\tau}\lambda_{Z,\tau-}\right)^\top. \tag{8.43}$$

We also consider an $N_n \times k_n$ matrix $\mathcal{E}'_n = [e'_{j,l}]_{1 \le j \le N_n, 1 \le l \le k_n}$ defined as

$$\begin{aligned} e'_{j,l} \equiv \Delta_n^{-1/2} &\int_{(i(n,\tau-)+l-1)\Delta_n}^{(i(n,\tau-)+l)\Delta_n} \alpha_{j,s} ds + \Delta_n^{-1/2} \int_{(i(n,\tau-)+l-1)\Delta_n}^{(i(n,\tau-)+l)\Delta_n} (\lambda_{j,u} - \lambda_{j,\tau-})^\top df_u \\ &- \tilde{\beta}_{n,j,\tau} \Delta_n^{-1/2} \int_{(i(n,\tau-)+l-1)\Delta_n}^{(i(n,\tau-)+l)\Delta_n} (\lambda_{Z,u} - \lambda_{Z,\tau-})^\top df_u. \end{aligned} \tag{8.44}$$

Lemma 4. *Under the conditions of Theorem 2, the following statements hold:*

(a) $\sup_{\gamma\in\Gamma_n} k_n^{-2}N_n^{-1}\gamma^\top\mathcal{E}_n^\top\mathcal{E}_n\gamma = o_p(1)$*;*

(b) $\sup_{\gamma\in\Gamma_n} k_n^{-2}N_n^{-1}\gamma^\top\mathcal{E}_n'^\top\mathcal{E}_n'\gamma = o_p(1)$*;*

(c) $\sup_{\gamma\in\Gamma_n} k_n^{-1}N_n^{-1}\left|\gamma^\top\mathcal{E}_n^\top\Lambda_n^*\right| = o_p(1)$*;*

(d) $\sup_{\gamma\in\Gamma_n} k_n^{-1}N_n^{-1}\left|\gamma^\top\mathcal{E}_n'^\top\Lambda_n^*\right| = o_p(1)$*;*

(e) $N_n^{-1}\tilde{\Lambda}_n^{*\top}\Lambda_n^* = M_\Lambda^* + o_p(1)$ *and* $N_n^{-1}\tilde{\Lambda}_n^{*\top}\tilde{\Lambda}_n^* = M_\Lambda^* + o_p(1)$.

Proof of Lemma 4. (a) Recall that the (j,l) element of $\mathcal{E}_n$ is given by $e_{j,l} \equiv \Delta^n_{i(n,\tau-)+l}\epsilon_j/\Delta_n^{1/2}$. We observe

$$
\begin{aligned}
&\frac{1}{k_n^2N_n}\gamma^\top\mathcal{E}_n^\top\mathcal{E}_n\gamma \\
&\quad= \frac{1}{k_n^2N_n}\sum_{l=1}^{k_n}\sum_{m=1}^{k_n}\gamma_l\gamma_m\sum_{j=1}^{N_n}e_{j,l}e_{j,m} \\
&\quad\le \left(\frac{1}{k_n^2}\sum_{l=1}^{k_n}\sum_{m=1}^{k_n}\gamma_l^2\gamma_m^2\right)^{1/2}\left(\frac{1}{k_n^2}\sum_{l=1}^{k_n}\sum_{m=1}^{k_n}\left(\frac{1}{N_n}\sum_{j=1}^{N_n}e_{j,l}e_{j,m}\right)^2\right)^{1/2} \\
&\quad= \left(\frac{1}{k_n^2}\sum_{l=1}^{k_n}\sum_{m=1}^{k_n}\left(\frac{1}{N_n}\sum_{j=1}^{N_n}e_{j,l}e_{j,m}\right)^2\right)^{1/2},
\end{aligned}
\tag{8.45}
$$

where the first equality is by definition, the inequality is by the Cauchy–Schwarz inequality, and the last line follows from $\gamma^\top\gamma = k_n$.

We decompose the majorant side of (8.45) as

$$
\begin{aligned}
&\frac{1}{k_n^2}\sum_{l=1}^{k_n}\sum_{m=1}^{k_n}\left(\frac{1}{N_n}\sum_{j=1}^{N_n}e_{j,l}e_{j,m}\right)^2 \\
&\quad= \frac{1}{k_n^2}\sum_{l=1}^{k_n}\left(\frac{1}{N_n}\sum_{j=1}^{N_n}e_{j,l}^2\right)^2 + \frac{1}{k_n^2}\sum_{l,m,l\neq m}\left(\frac{1}{N_n}\sum_{j=1}^{N_n}e_{j,l}e_{j,m}\right)^2.
\end{aligned}
\tag{8.46}
$$

By a standard estimate for continuous Itô semimartingales, $\mathbb{E}[e_{j,l}^4] \le K$; this holds uniformly in $j \in \{1,\ldots,N_n\}$ because the idiosyncratic variances $\tilde{\sigma}_j^2$ are uniformly (locally) bounded under Assumption 3(iii). Hence, by Jensen's inequality,

$$
\mathbb{E}\left[\left(\frac{1}{N_n}\sum_{j=1}^{N_n}e_{j,l}^2\right)^2\right] \le \mathbb{E}\left[\frac{1}{N_n}\sum_{j=1}^{N_n}e_{j,l}^4\right] \le K.
$$

From here, it follows that the first term on the right-hand side of (8.46) is $o_p(1)$. In view of (8.45) and (8.46), it remains to show that the second term on the right-hand side of (8.46) is also $o_p(1)$.

To this end, we observe the following for $l \neq m$: (i) $\mathbb{E}\left[e_{j,l}e_{j,m}\right] = 0$ because the process ϵ_j is a martingale; (ii) $\mathbb{E}[e_{j,l}^2e_{j,m}^2] \le K$; and (iii) the variables $(e_{j,l}e_{j,m})_{1\le j\le N_n}$ are uncorrelated, which can be shown by using repeated conditioning and the orthogonality among the Brownian motions $(\widetilde{W}_j)_{j\geq 1}$. Hence,

$$\mathbb{E}\left[\left(\frac{1}{N_n}\sum_{j=1}^{N_n} e_{j,l}e_{j,m}\right)^2\right] \leq KN_n^{-1} \to 0,$$

which implies, as wanted,

$$\frac{1}{k_n^2}\sum_{l,m,l\neq m}\left(\frac{1}{N_n}\sum_{j=1}^{N_n} e_{j,l}e_{j,m}\right)^2 = O_p(N_n^{-1}) = o_p(1).$$

This finishes the proof of part (a).

(b) Similar to (8.45), we can derive

$$\sup_{\gamma\in\Gamma_n}\frac{1}{k_n^2 N_n}\gamma^\top \mathcal{E}_n'^\top \mathcal{E}_n'\gamma \leq \left(\frac{1}{k_n^2}\sum_{l=1}^{k_n}\sum_{m=1}^{k_n}\left(\frac{1}{N_n}\sum_{j=1}^{N_n} e'_{j,l}e'_{j,m}\right)^2\right)^{1/2}. \tag{8.47}$$

In addition, we observe

$$\begin{aligned}
&\mathbb{E}\left|\Delta_n^{-1/2}\int_{(i(n,\tau-)+l-1)\Delta_n}^{(i(n,\tau-)+l)\Delta_n}(\lambda_{j,u}-\lambda_{j,\tau-})^\top df_u\right|^4 \\
&\quad\leq K\Delta_n^{-2}\mathbb{E}\left[\left(\int_{(i(n,\tau-)+l-1)\Delta_n}^{(i(n,\tau-)+l)\Delta_n}\|\lambda_{j,u}-\lambda_{j,\tau-}\|^2\,du\right)^2\right] \\
&\quad\leq K\Delta_n^{-1}\mathbb{E}\left[\int_{(i(n,\tau-)+l-1)\Delta_n}^{(i(n,\tau-)+l)\Delta_n}\|\lambda_{j,u}-\lambda_{j,\tau-}\|^4\,du\right] \\
&\quad\leq K\Delta_n^{-1}\mathbb{E}\left[\int_{(i(n,\tau-)+l-1)\Delta_n}^{(i(n,\tau-)+l)\Delta_n}\|\lambda_{j,u}-\lambda_{j,\tau-}\|^2\,du\right] \leq K\Delta_n,
\end{aligned} \tag{8.48}$$

where the first inequality is by the Burkholder–Davis–Gundy inequality, the second inequality is by Jensen's inequality, and the last line holds because $\lambda_{j,u}$ is bounded and (1/2)-Hölder continuous under L_2-norm uniformly in j. Similarly,

$$\mathbb{E}\left|\Delta_n^{-1/2}\int_{(i(n,\tau-)+l-1)\Delta_n}^{(i(n,\tau-)+l)\Delta_n}(\lambda_{Z,u}-\lambda_{Z,\tau-})^\top df_u\right|^4 \leq K\Delta_n. \tag{8.49}$$

Under Assumption 7, $(\tilde{\beta}_{j,n,\tau})_{1\leq j\leq N_n}$ are uniformly bounded with probability approaching one, so we can assume that these variables are bounded without loss of generality. Hence, from (8.48) and (8.49), we deduce that

$$\mathbb{E}|e'_{j,l}|^4 \leq K\Delta_n. \tag{8.50}$$

Hence, by the Cauchy–Schwarz inequality, we further have

$$\mathbb{E}\left[\left(\frac{1}{N_n}\sum_{j=1}^{N_n} e'_{j,l}e'_{j,m}\right)^2\right] \leq K\Delta_n. \tag{8.51}$$

The assertion of part (b) then follows from (8.47) and (8.51).

(c) We denote the (j,k) element of Λ_n^* by $\lambda_{j,k}^*$. We note that for each $k \in \{1,\ldots,r\}$ (recalling that $\Lambda_{n,\cdot k}^*$ denotes the kth column of Λ_n^*),

$$\begin{aligned}
\frac{1}{k_n N_n}\left|\gamma^\top \mathcal{E}_n^\top \Lambda_{n,\cdot k}^*\right| &= \left|\frac{1}{k_n}\sum_{l=1}^{k_n}\gamma_l\left(\frac{1}{N_n}\sum_{j=1}^{N_n} e_{j,l}\lambda_{j,k}^*\right)\right| \\
&\leq \left(\frac{1}{k_n}\sum_{l=1}^{k_n}\gamma_l^2\right)^{1/2}\left(\frac{1}{k_n}\sum_{l=1}^{k_n}\left(\frac{1}{N_n}\sum_{j=1}^{N_n} e_{j,l}\lambda_{j,k}^*\right)^2\right)^{1/2} \\
&= \left(\frac{1}{k_n}\sum_{l=1}^{k_n}\left(\frac{1}{N_n}\sum_{j=1}^{N_n} e_{j,l}\lambda_{j,k}^*\right)^2\right)^{1/2}, \qquad (8.52)
\end{aligned}$$

where the first line is by definition, the second line is by the Cauchy–Schwarz inequality and the last line follows from $\gamma \in \Gamma_n$. Under Assumption 8, $e_{j,l}$ is independent of $\lambda_{j,k}^*$; hence, the variables $(e_{j,l}\lambda_{j,k}^*)_{1\leq j\leq N_n}$ are uncorrelated and have zero mean and bounded second moment. It is then easy to see that

$$\mathbb{E}\left[\frac{1}{k_n}\sum_{l=1}^{k_n}\left(\frac{1}{N_n}\sum_{j=1}^{N_n} e_{j,l}\lambda_{j,k}^*\right)^2\right] \leq K/N_n.$$

Therefore,

$$\frac{1}{k_n}\sum_{l=1}^{k_n}\left(\frac{1}{N_n}\sum_{j=1}^{N_n} e_{j,l}\lambda_{j,k}^*\right)^2 = O_p(N_n^{-1}) = o_p(1). \qquad (8.53)$$

The assertion of part (c) then follows from (8.52) and (8.53).

(d) Like (8.52), we can derive

$$\frac{1}{k_n N_n}\left|\gamma^\top \mathcal{E}_n'^\top \Lambda_{n,\cdot k}^*\right| \leq \left(\frac{1}{k_n}\sum_{l=1}^{k_n}\left(\frac{1}{N_n}\sum_{j=1}^{N_n} e'_{j,l}\lambda_{j,k}^*\right)^2\right)^{1/2}. \qquad (8.54)$$

We further note that,

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{1}{N_n}\sum_{j=1}^{N_n} e'_{j,l}\lambda_{j,k}^*\right)^2\right] &\leq \mathbb{E}\left[\frac{1}{N_n}\sum_{j=1}^{N_n}\left(e'_{j,l}\lambda_{j,k}^*\right)^2\right] \\
&\leq \frac{K}{N_n}\sum_{j=1}^{N_n}\mathbb{E}\left[(e'_{j,l})^2\right] \leq K\Delta_n,
\end{aligned}$$

where the first inequality is by Jensen's inequality, the second inequality holds because $\lambda_{j,k}^*$ is bounded and the last inequality can be derived similarly as (8.50). In view of (8.54), the assertion of part (d) readily follows.

(e) From the definitions of Λ_n^* and $\tilde{\Lambda}_n^*$ respectively from (3.9) and (8.43), we see that (recall $q = \tau-$)

$$\tilde{\Lambda}_n^* - \Lambda_n^* = \left(\left(\beta_{1,\tau}^* - \tilde{\beta}_{n,1,\tau}\right)\lambda_{Z,\tau-}, \ldots, \left(\beta_{n,N_n,\tau}^* - \tilde{\beta}_{n,N_n,\tau}\right)\lambda_{Z,\tau-}\right)^\top.$$

Therefore, by Assumption 7,

$$\frac{1}{N_n}(\tilde{\Lambda}_n^* - \Lambda_n^*)^\top(\tilde{\Lambda}_n^* - \Lambda_n^*) = o_p(1). \tag{8.55}$$

That is, $N_n^{-1}\|\tilde{\Lambda}_n^* - \Lambda_n^*\|^2 = o_p(1)$. Since $N_n^{-1}\Lambda_n^{*\top}\Lambda_n^* \xrightarrow{\mathbb{P}} M_\Lambda^*$ by Assumption 8, the estimate above readily implies the assertions in part (e). *Q.E.D.*

We are now ready to prove part (a) and part (b) of Theorem 2. We remind the reader that we fix $q = \tau-$ for proving these parts.

PROOF OF THEOREM 2(a). Step 1. We prove part (a) of Theorem 2 in several steps. In this step, we show that

$$\sup_{\gamma \in \Gamma_n} |\Xi_n(\gamma) - \Xi_n^*(\gamma)| = o_p(1), \tag{8.56}$$

where $\Xi_n(\cdot)$ and $\Xi_n^*(\cdot)$ are defined as

$$\Xi_n(\gamma) \equiv \frac{1}{k_n^2 N_n}\gamma^\top \hat{X}_n^\top \hat{X}_n \gamma, \quad \Xi_n^*(\gamma) \equiv \frac{1}{k_n^2 N_n}\gamma^\top F_n \Lambda_n^{*\top}\Lambda_n^* F_n^\top \gamma. \tag{8.57}$$

Below, we denote the (j,l) element of $\hat{X}_n$ by

$$\xi_{n,j,l} \equiv \frac{\Delta_{i(n,\tau-)+l}^n Y_j \wedge u_n \vee (-u_n) - \tilde{\beta}_{n,j,\tau}\Delta_{i(n,\tau-)+l}^n Z}{\sqrt{\Delta_n}}.$$

We set

$$\begin{aligned}\xi'_{n,j,l} \quad \equiv \quad & \Delta_n^{-1/2}\int_{(i(n,\tau-)+l-1)\Delta_n}^{(i(n,\tau-)+l)\Delta_n} \alpha_{j,s} ds \\ & +\Delta_n^{-1/2}\int_{(i(n,\tau-)+l-1)\Delta_n}^{(i(n,\tau-)+l)\Delta_n} (\lambda_{j,s} - \tilde{\beta}_{n,j,\tau}\lambda_{Z,s})^\top df_s + \Delta_n^{-1/2}\Delta_{i(n,\tau-)+l}^n \epsilon_j.\end{aligned}$$

Note that

$$\mathbb{E}\left|\xi_{n,j,l} - \xi'_{n,j,l}\right|^2 \le K\Delta_n. \tag{8.58}$$

We now define $\hat{X}'_n$ as a $N_n \times k_n$ matrix whose (j,l) element is given by $\xi'_{n,j,l}$ and let

$$\Xi'_n(\gamma) = \frac{1}{k_n^2 N_n}\gamma^\top \hat{X}_n'^\top \hat{X}'_n \gamma.$$

By (8.58),

$$\frac{1}{k_n N_n}\left\|\hat{X}_n - \hat{X}'_n\right\|^2 = \frac{1}{k_n N_n}\sum_{j=1}^{N_n}\sum_{l=1}^{k_n}\left|\xi_{n,j,l} - \xi'_{n,j,l}\right|^2 = o_p(1). \tag{8.59}$$

By the Cauchy–Schwarz inequality and the triangle inequality,

$$\begin{aligned}\sup_{\gamma\in\Gamma_n}\left|\Xi_n(\gamma)-\Xi_n'(\gamma)\right| &= \frac{1}{k_n^2N_n}\sup_{\gamma\in\Gamma_n}\left|\gamma^\top\left(\hat{X}_n^\top\hat{X}_n-\hat{X}_n'^\top\hat{X}_n'\right)\gamma\right| \\ &\leq \frac{1}{k_n^2N_n}\sup_{\gamma\in\Gamma_n}\|\gamma\|^2\left\|\hat{X}_n^\top\hat{X}_n-\hat{X}_n'^\top\hat{X}_n'\right\| \\ &= \frac{1}{k_nN_n}\left\|\hat{X}_n^\top\hat{X}_n-\hat{X}_n'^\top\hat{X}_n'\right\| \\ &\leq \frac{2}{k_nN_n}\left\|\hat{X}_n'\right\|\left\|\hat{X}_n-\hat{X}_n'\right\|+\frac{1}{k_nN_n}\left\|\hat{X}_n-\hat{X}_n'\right\|^2.\end{aligned}$$

It is easy to see that $\|\hat{X}_n'\|=O_p(\sqrt{k_nN_n})$. Hence, by (8.59),

$$\sup_{\gamma\in\Gamma_n}\left|\Xi_n(\gamma)-\Xi_n'(\gamma)\right|=o_p(1). \tag{8.60}$$

To show (8.56), it remains to show that $\sup_{\gamma\in\Gamma_n}|\Xi_n'(\gamma)-\Xi_n^*(\gamma)|=o_p(1)$. We note that, by a standard result for spot covariance estimation

$$F_n^\top F_n/k_n\xrightarrow{\mathbb{P}}\Sigma_f. \tag{8.61}$$

In particular, $\|F_n\|=O_p(k_n^{1/2})$. Hence,

$$\sup_{\gamma\in\Gamma_n}\left\|\gamma^\top F_n/k_n\right\|\leq\sup_{\gamma\in\Gamma_n}\|\gamma\|\,\|F_n\|/k_n=O_p(1). \tag{8.62}$$

Under Assumption 8, $\Lambda_n^{*\top}\Lambda_n^*=O_p(N_n)$. It then follows that

$$\sup_{\gamma\in\Gamma_n}\Xi_n^*(\gamma)=O_p(1). \tag{8.63}$$

Recall the definitions in (3.9), (8.43) and (8.44). We can decompose $\hat{X}_n'$ as

$$\hat{X}_n'=\tilde{\Lambda}_n^*F_n^\top+\mathcal{E}_n+\mathcal{E}_n'. \tag{8.64}$$

Hence,

$$\hat{X}_n'-\Lambda_n^*F_n^\top=(\tilde{\Lambda}_n^*-\Lambda_n^*)F_n^\top+\mathcal{E}_n+\mathcal{E}_n'. \tag{8.65}$$

We can then decompose

$$\begin{aligned}\Xi_n'(\gamma)-\Xi_n^*(\gamma) &= \frac{2}{k_n^2N_n}\gamma^\top F_n\Lambda_n^{*\top}\left[(\tilde{\Lambda}_n^*-\Lambda_n^*)F_n^\top+\mathcal{E}_n+\mathcal{E}_n'\right]\gamma \\ &\quad+\frac{1}{k_n^2N_n}\gamma^\top\left((\tilde{\Lambda}_n^*-\Lambda_n^*)F_n^\top+\mathcal{E}_n+\mathcal{E}_n'\right)^\top\left((\tilde{\Lambda}_n^*-\Lambda_n^*)F_n^\top+\mathcal{E}_n+\mathcal{E}_n'\right)\gamma.\end{aligned} \tag{8.66}$$

By Lemma 4(a,b),

$$\frac{1}{k_n^2N_n}\sup_{\gamma\in\Gamma_n}\gamma^\top\mathcal{E}_n^\top\mathcal{E}_n\gamma=o_p(1),\quad \frac{1}{k_n^2N_n}\sup_{\gamma\in\Gamma_n}\gamma^\top\mathcal{E}_n'^\top\mathcal{E}_n'\gamma=o_p(1). \tag{8.67}$$

Further using the Cauchy–Schwarz inequality, we can deduce that $\sup_{\gamma\in\Gamma_n}\left|\gamma^\top\mathcal{E}_n^\top\mathcal{E}_n'\gamma\right| = o_p(1)$; hence,

$$\frac{1}{k_n^2 N_n}\sup_{\gamma\in\Gamma_n}\gamma^\top\left(\mathcal{E}_n+\mathcal{E}_n'\right)^\top\left(\mathcal{E}_n+\mathcal{E}_n'\right)\gamma = o_p(1). \tag{8.68}$$

In addition, by Lemma 4(e) and (8.62)

$$\sup_{\gamma\in\Gamma_n}\frac{1}{k_n^2 N_n}\gamma^\top F_n^\top(\tilde{\Lambda}_n^*-\Lambda_n^*)^\top(\tilde{\Lambda}_n^*-\Lambda_n^*)F_n^\top\gamma = o_p(1). \tag{8.69}$$

By (8.68) and (8.69), as well as the Cauchy–Schwarz inequality, we deduce

$$\frac{1}{k_n^2 N_n}\sup_{\gamma\in\Gamma_n}\gamma^\top\left(\left(\tilde{\Lambda}_n^*-\Lambda_n^*\right)F_n^\top+\mathcal{E}_n+\mathcal{E}_n'\right)^\top\left(\left(\tilde{\Lambda}_n^*-\Lambda_n^*\right)F_n^\top+\mathcal{E}_n+\mathcal{E}_n'\right)\gamma = o_p(1). \tag{8.70}$$

By (8.63) and the Cauchy–Schwarz inequality, (8.70) further implies that

$$\frac{2}{k_n^2 N_n}\sup_{\gamma\in\Gamma_n}\left|\gamma^\top F_n\Lambda_n^{*\top}\left[(\tilde{\Lambda}_n^*-\Lambda_n^*)F_n^\top+\mathcal{E}_n+\mathcal{E}_n'\right]\gamma\right| = o_p(1). \tag{8.71}$$

By (8.66), (8.70) and (8.71), we deduce $\sup_{\gamma\in\Gamma_n}\left|\Xi_n'(\gamma)-\Xi_n^*(\gamma)\right| = o_p(1)$ and, hence, (8.56) as wanted.

Step 2. In this step, we show that

$$S_n^*(\hat{F}_n^{*\top}F_n/k_n)\Sigma_f^{-1/2}H \xrightarrow{\mathbb{P}} I_r, \tag{8.72}$$

where we recall that $S_n^* = \mathrm{diag}(\mathrm{sign}(\hat{F}_n^{*\top}F_n(F_n^\top F_n/k_n)^{-1/2}H))$ and H is the ordered eigenvector matrix of M_C^*. Below, we denote by D_j the jth largest eigenvalue of M_C^* and write $D = \mathrm{diag}(D_1,\ldots,D_r)$.

We first show that

$$\sup_{\gamma\in\Gamma_n}\Xi_n^*(\gamma) \xrightarrow{\mathbb{P}} D_1. \tag{8.73}$$

To see this, we note that we can represent $\gamma\in\Gamma_n$ as

$$\gamma = F_n(F_n^\top F_n/k_n)^{-1/2}H\delta+\tilde{\gamma}, \tag{8.74}$$

where $\tilde{\gamma}$ is the projection error of γ onto the column space of F_n such that $F_n^\top\tilde{\gamma} = 0$. We can then rewrite

$$\begin{aligned}\sup_{\gamma\in\Gamma_n}\Xi_n^*(\gamma) &= \sup_{\|\delta\|\leq 1}\delta^\top H^\top M_{C,n}^* H\delta, \quad \text{where}\\ M_{C,n}^* &\equiv \left(\frac{F_n^\top F_n}{k_n}\right)^{1/2}\left(\frac{\Lambda_n^{*\top}\Lambda_n^*}{N_n}\right)\left(\frac{F_n^\top F_n}{k_n}\right)^{1/2}.\end{aligned}$$

Hence, $\sup_{\gamma\in\Gamma_n}\Xi_n^*(\gamma)$ is the largest eigenvalue of $M_{C,n}^*$. By (8.61) and Assumption 8,

$$M_{C,n}^* \xrightarrow{\mathbb{P}} \Sigma_f^{1/2}M_\Lambda^*\Sigma_f^{1/2} \equiv M_C^*.$$

Since the mapping for calculating the unique largest eigenvalue is continuous, we deduce (8.73) by using the continuous mapping theorem.

By the construction of $\hat{F}_n$, its first column $\hat{F}_{n,\cdot 1}$ satisfies

$$\Xi_n(\hat{F}_{n,\cdot 1}) = \sup_{\gamma\in\Gamma_n} \Xi_n(\gamma).$$

By (8.56), $\sup_{\gamma\in\Gamma_n} \Xi_n(\gamma) = \sup_{\gamma\in\Gamma_n} \Xi_n^*(\gamma) + o_p(1)$, which implies $\Xi_n(\hat{F}_{n,\cdot 1}) \xrightarrow{\mathbb{P}} D_1$ because of (8.73). Using the uniform convergence result in (8.56), we further deduce

$$\Xi_n^*(\hat{F}_{n,\cdot 1}) \xrightarrow{\mathbb{P}} D_1. \tag{8.75}$$

We now represent $\hat{F}_{n,\cdot 1}$ in the format of (8.74), that is,

$$\hat{F}_{n,\cdot 1} = F_n(F_n^\top F_n/k_n)^{-1/2} H\hat{\delta}_1 + \tilde{\gamma}_1, \tag{8.76}$$

such that $F_n^\top \tilde{\gamma}_1 = 0$. From (8.75) and (8.76), we see

$$\begin{aligned} o_p(1) &= \Xi_n^*(\hat{F}_{n,\cdot 1}) - D_1 \\ &= \hat{\delta}_1^\top H^\top M_{C,n}^* H\hat{\delta}_1 - D_1 \\ &= \hat{\delta}_1^\top H^\top \left(M_{C,n}^* - M_C^*\right) H\hat{\delta}_1 + \hat{\delta}_1^\top H^\top M_C^* H\hat{\delta}_1 - D_1 \\ &= \hat{\delta}_1^\top H^\top \left(M_{C,n}^* - M_C^*\right) H\hat{\delta}_1 + \hat{\delta}_1^\top D\hat{\delta}_1 - D_1, \end{aligned}$$

where the last line follows from the eigenvalue decomposition $M_C^* = HDH^\top$. Since $\|\hat{\delta}_1\| \le 1$ and $M_{C,n}^* - M_C^* = o_p(1)$, the above display implies that

$$\hat{\delta}_1^\top D\hat{\delta}_1 - D_1 = o_p(1).$$

Since D_1 is the unique largest eigenvalue, this further implies that $\hat{\delta}_{11}^2 \xrightarrow{\mathbb{P}} 1$ and $\hat{\delta}_{1j}^2 \xrightarrow{\mathbb{P}} 0$ for $j \ge 2$. In particular, $\|\hat{\delta}_1\| \xrightarrow{\mathbb{P}} 1$ which implies that $\tilde{\gamma}_1^\top \tilde{\gamma}_1/k_n \xrightarrow{\mathbb{P}} 0$.

Let $S_{n,j}^*$ denote the jth diagonal element of S_n^*. Note that by (8.76),

$$\hat{F}_{n,\cdot 1}^\top F_n/k_n = \hat{\delta}_1^\top H^\top (F_n^\top F_n/k_n)^{1/2}.$$

Hence,

$$\hat{\delta}_1^\top = (\hat{F}_{n,\cdot 1}^\top F_n/k_n)(F_n^\top F_n/k_n)^{-1/2} H.$$

By the definition of $S_{n,1}^*$, the first element of $S_{n,1}^*(\hat{F}_{n,\cdot 1}^\top F_n/k_n)(F_n^\top F_n/k_n)^{-1/2} H$ is nonnegative. Hence,

$$\begin{aligned} &S_{n,1}^*(\hat{F}_{n,\cdot 1}^\top F_n/k_n)(F_n^\top F_n/k_n)^{-1/2} H \\ &\quad = \left(|\hat{\delta}_{11}|, S_{n,1}^*\hat{\delta}_{12}, \ldots, S_{n,1}^*\hat{\delta}_{1r}\right) \xrightarrow{\mathbb{P}} (1, 0, \ldots, 0). \end{aligned}$$

By (8.61), we further deduce that

$$S^*_{n,1}(\hat{F}^\top_{n,\cdot 1}F_n/k_n)\Sigma_f^{-1/2}H \xrightarrow{\mathbb{P}} (1,0,\ldots,0)\,,$$

which shows the convergence in (8.72) for the first row.

By repeating the same argument (by setting Γ_n as the subspace orthogonal to previous eigenvectors), we can prove the convergence in (8.72) for the jth row, $2 \le j \le r$.

Step 3. In this step, we finish the proof for part (a) of Theorem 2. We denote

$$\tilde{D}_n = N_n^{-1}(\hat{\Lambda}^*_n - \Lambda^*_n\Sigma_f^{1/2}HS^*_n)^\top(\hat{\Lambda}^*_n - \Lambda^*_n\Sigma_f^{1/2}HS^*_n).$$

The assertion of part (a) can be rewritten as $\text{Trace}[\tilde{D}_n] = o_p(1)$.

We decompose

$$\tilde{D}_n = \tilde{D}_{n,1} - \tilde{D}_{n,2} - \tilde{D}^\top_{n,2} + \tilde{D}_{n,3},$$

where

$$\begin{cases} \tilde{D}_{n,1} \equiv N_n^{-1}\hat{\Lambda}^{*\top}_n\hat{\Lambda}^*_n, \quad \tilde{D}_{n,2} \equiv N_n^{-1}\hat{\Lambda}^{*\top}_n\Lambda^*_n\Sigma_f^{1/2}HS^*_n, \\ \tilde{D}_{n,3} \equiv N_n^{-1}S^*_nH^\top\Sigma_f^{1/2}\Lambda^{*\top}_n\Lambda^*_n\Sigma_f^{1/2}HS^*_n. \end{cases}$$

To prove $\text{Trace}[\tilde{D}_n] = o_p(1)$, it suffices to show that

$$\tilde{D}_{n,k} \xrightarrow{\mathbb{P}} D, \quad k = 1,2,3, \tag{8.77}$$

where we recall that D is the diagonal matrix that collects the ordered eigenvalues of M^*_C. Below, we prove (8.77) for each case.

Case $k = 1$: Recall that we partition $\hat{F}_n = [\hat{F}^*_n \vdots \hat{F}^0_n]$, where $\hat{F}^*_n$ collects the first r columns of $\hat{F}_n$. We set

$$\hat{\Lambda}'^*_n = \frac{1}{k_n}\hat{X}'_n\hat{F}^*_n = \frac{1}{k_n}\left(\tilde{\Lambda}^*_nF^\top_n + \mathcal{E}_n + \mathcal{E}'_n\right)\hat{F}^*_n. \tag{8.78}$$

Note that

$$\begin{aligned} \left\|\tilde{D}_{n,1} - N_n^{-1}\hat{\Lambda}'^{*\top}_n\hat{\Lambda}'^*_n\right\| &= N_n^{-1}\left\|\hat{\Lambda}^{*\top}_n\hat{\Lambda}^*_n - \hat{\Lambda}'^{*\top}_n\hat{\Lambda}'^*_n\right\| \\ &= k_n^{-2}N_n^{-1}\left\|\hat{F}^{*\top}_n\hat{X}^\top_n\hat{X}_n\hat{F}^*_n - \hat{F}^{*\top}_n\hat{X}'^\top_n\hat{X}'_n\hat{F}^*_n\right\| = o_p(1), \end{aligned} \tag{8.79}$$

where the first two equalities are by definition and the last one is by (8.60). Subsequently, by (8.78), we can decompose $\tilde{D}_{n,1}$ as

$$\begin{aligned} \tilde{D}_{n,1} &= \frac{1}{k_n^2N_n}\hat{F}^{*\top}_n\left(\tilde{\Lambda}^*_nF^\top_n + \mathcal{E}_n + \mathcal{E}'_n\right)^\top\left(\tilde{\Lambda}^*_nF^\top_n + \mathcal{E}_n + \mathcal{E}'_n\right)\hat{F}^*_n + o_p(1) \\ &= \tilde{D}_{n,1,1} + \tilde{D}_{n,1,2} + \tilde{D}^\top_{n,1,2} + \tilde{D}_{n,1,3} + o_p(1), \end{aligned}$$

where

$$\begin{cases} \tilde{D}_{n,1,1} \equiv \left(\hat{F}_n^{*\top} F_n/k_n\right) \left(\tilde{\Lambda}_n^{*\top}\tilde{\Lambda}_n^*/N_n\right) \left(F_n^\top \hat{F}_n^*/k_n\right), \\ \tilde{D}_{n,1,2} \equiv \left(k_n^{-1} N_n^{-1} \hat{F}_n^{*\top} \left(\mathcal{E}_n + \mathcal{E}_n'\right)^\top \tilde{\Lambda}_n^*\right) \left(F_n^\top \hat{F}_n^*/k_n\right), \\ \tilde{D}_{n,1,3} \equiv k_n^{-2} N_n^{-1} \hat{F}_n^{*\top} \left(\mathcal{E}_n + \mathcal{E}_n'\right)^\top \left(\mathcal{E}_n + \mathcal{E}_n'\right) \hat{F}_n^*. \end{cases}$$

From (8.72),

$$\frac{1}{k_n} \hat{F}_n^{*\top} F_n - S_n^* H^\top \Sigma_f^{1/2} = o_p(1), \quad \frac{1}{k_n} \hat{F}_n^{*\top} F_n = O_p(1). \tag{8.80}$$

Hence, recalling that H is the eigenvector matrix of $M_C^* = \Sigma_f^{1/2} M_\Lambda^* \Sigma_f^{1/2}$ and S_n^* is a diagonal matrix with ± 1 on its diagonal, we deduce

$$\tilde{D}_{n,1,1} = S_n^* H^\top \Sigma_f^{1/2} M_\Lambda^* \Sigma_f^{1/2} H S_n^* + o_p(1) = D + o_p(1).$$

By Lemma 4, we see that $\tilde{D}_{n,1,2}$ and $\tilde{D}_{n,1,3}$ are both $o_p(1)$. From these estimates, (8.77) for the case $k = 1$ readily follows.

Case $k = 2$: By (8.79) and the Cauchy–Schwarz inequality,

$$\tilde{D}_{n,2} \equiv N_n^{-1} \hat{\Lambda}_n^{\prime*\top} \Lambda_n^* \Sigma_f^{1/2} H S_n^* + o_p(1).$$

By (8.78), we can thus decompose $\tilde{D}_{n,2}$ as $\tilde{D}_{n,2} = \tilde{D}_{n,2,1} + \tilde{D}_{n,2,2} + o_p(1)$ where

$$\begin{cases} \tilde{D}_{n,2,1} \equiv \left(\hat{F}_n^{*\top} F_n/k_n\right) \left(\tilde{\Lambda}_n^{*\top} \Lambda_n^*/N_n\right) \Sigma_f^{1/2} H S_n^*, \\ \tilde{D}_{n,2,2} \equiv \left(k_n^{-1} N_n^{-1} \hat{F}_n^{*\top} \left(\mathcal{E}_n + \mathcal{E}_n'\right)^\top \Lambda_n^*\right) \Sigma_f^{1/2} H S_n^*. \end{cases}$$

By (8.80) and Lemma 4(e), we deduce

$$\tilde{D}_{n,2,1} = S_n^* H^\top \Sigma_f^{1/2} M_\Lambda^* \Sigma_f^{1/2} H S_n^* + o_p(1) = D + o_p(1).$$

By Lemma 4(c,d), $\tilde{D}_{n,2,2} = o_p(1)$. This proves (8.77) for the case $k = 2$.

Case $k = 3$: By Assumption 8, it is obvious that

$$\tilde{D}_{n,3} = S_n^* H^\top \Sigma_f^{1/2} M_\Lambda^* \Sigma_f^{1/2} H S_n^* + o_p(1) = D + o_p(1).$$

This finishes the proof of (8.77) and, hence, part (a) of Theorem 2. *Q.E.D.*

Proof of Theorem 2(b). We fix $j \in \{r+1, \ldots, \bar{r}\}$. Recall that $\hat{\Lambda}_{n,\cdot j}$ denote the jth column of $\hat{\Lambda}_n$. By the definitions of $\hat{\Lambda}_n$ and $\hat{F}_n$,

$$\frac{1}{N_n} \hat{\Lambda}_{n,\cdot j}^\top \hat{\Lambda}_{n,\cdot j} = \Xi_n(\hat{F}_{n,\cdot j}). \tag{8.81}$$

Like in (8.76), for each $k \in \{1, \ldots, r\}$, we can represent

$$\hat{F}_{n,\cdot k} = F_n (F_n^\top F_n/k_n)^{-1/2} H \hat{\delta}_k + \tilde{\gamma}_k, \tag{8.82}$$

where $F_n^\top \tilde{\gamma}_k = 0$. Following a similar argument as in Step 2 of the proof of Theorem 2(a), we can show that, for each $k, k' \in \{1, \dots, r\}$ with $k \neq k'$,

$$\hat{\delta}_{kk}^2 \xrightarrow{\mathbb{P}} 1, \quad \hat{\delta}_{kk'} \xrightarrow{\mathbb{P}} 0, \quad \tilde{\gamma}_k^\top \tilde{\gamma}_k / k_n \xrightarrow{\mathbb{P}} 0. \tag{8.83}$$

We also represent

$$\hat{F}_{n,\cdot j} = F_n (F_n^\top F_n / k_n)^{-1/2} H \hat{\delta}_j + \tilde{\gamma}_j, \tag{8.84}$$

where $F_n^\top \tilde{\gamma}_j = 0$. Since $\hat{F}_{n,\cdot j}^\top \hat{F}_{n,\cdot k} / k_n = 0$ for $1 \leq k \leq r$ (because $\hat{F}_n$ collects the eigenvectors of $\hat{X}_n^\top \hat{X}_n$), we have

$$\hat{\delta}_j^\top \hat{\delta}_k + \tilde{\gamma}_j^\top \tilde{\gamma}_k / k_n = 0. \tag{8.85}$$

Since $\tilde{\gamma}_k^\top \tilde{\gamma}_k / k_n \xrightarrow{\mathbb{P}} 0$ and $\tilde{\gamma}_j^\top \tilde{\gamma}_j / k_n \leq 1$, we have $\tilde{\gamma}_j^\top \tilde{\gamma}_k / k_n = o_p(1)$ by the Cauchy–Schwarz inequality. Therefore, $\hat{\delta}_j^\top \hat{\delta}_k = o_p(1)$ for $1 \leq k \leq r$. By (8.83) above, this implies $\hat{\delta}_j = o_p(1)$. Hence,

$$\Xi_n^*(\hat{F}_{n,\cdot j}) = \hat{\delta}_j^\top H M_{C,n}^* H^\top \hat{\delta}_j = o_p(1). \tag{8.86}$$

By (8.56), $\Xi_n(\hat{F}_{n,\cdot j}) = o_p(1)$. The assertion of part (b) readily follows from (8.81). *Q.E.D.*

Proof of Theorem 2(c). By Assumption 8,

$$H_p^\top \Sigma_{f,p}^{1/2} \frac{\Lambda_n^*(p)^\top \Lambda_n^*(q)}{N_n} \Sigma_{f,q}^{1/2} H_q \xrightarrow{\mathbb{P}} H_p^\top M_C^*(p,q) H_q. \tag{8.87}$$

We observe

$$\begin{aligned}
&\frac{1}{N_n} \left\| \hat{\Lambda}_n^*(p)^\top \hat{\Lambda}_n^*(q) - S_n^*(p) H_p^\top \Sigma_{f,p}^{1/2} \Lambda_n^*(p)^\top \Lambda_n^*(q) \Sigma_{f,q}^{1/2} H_q S_n^*(q) \right\| \\
&\quad \leq \frac{1}{N_n} \left\| \left(\hat{\Lambda}_n^*(p) - \Lambda_n^*(p) \Sigma_{f,p}^{1/2} H_p S_n^*(p) \right)^\top \Lambda_n^*(q) \Sigma_{f,q}^{1/2} H_q \right\| \\
&\qquad + \frac{1}{N_n} \left\| H_p^\top \Sigma_{f,p}^{1/2} \Lambda_n^*(p)^\top \left(\hat{\Lambda}_n^*(q) - \Lambda_n^*(q) \Sigma_{f,q}^{1/2} H_q S_n^*(q) \right) \right\| \\
&\qquad + \frac{1}{N_n} \left\| \left(\hat{\Lambda}_n^*(p) - \Lambda_n^*(p) \Sigma_{f,p}^{1/2} H_p S_n^*(p) \right)^\top \left(\hat{\Lambda}_n^*(q) - \Lambda_n^*(q) \Sigma_{f,q}^{1/2} H_q S_n^*(q) \right) \right\|.
\end{aligned}$$

By the Cauchy–Schwarz inequality and Theorem 2(a), we deduce that the terms on the majorant side of the above display are all $o_p(1)$. Hence, by (8.87),

$$\frac{1}{N_n} \hat{\Lambda}_n^*(p)^\top \hat{\Lambda}_n^*(q) - S_n^*(p) H_p^\top M_C^*(p,q) H_q S_n^*(q) = o_p(1). \tag{8.88}$$

In particular,

$$\frac{1}{N_n} \hat{\Lambda}_n^*(p)^\top \hat{\Lambda}_n^*(q) = O_p(1). \tag{8.89}$$

By Theorem 2(b),

$$\frac{1}{N_n} \hat{\Lambda}_n^0(p)^\top \hat{\Lambda}_n^0(q) = o_p(1). \tag{8.90}$$

By the Cauchy–Schwarz inequality, (8.89) and (8.90), we deduce

$$\frac{1}{N_n}\hat{\Lambda}_n^*(p)^\top \hat{\Lambda}_n^0(q) = o_p(1). \tag{8.91}$$

The assertion of part (c) then follows from (8.88), (8.90) and (8.91). *Q.E.D.*

Proof of Theorem 2(d). By part (c) of Theorem 2,

$$\begin{aligned}
\text{Trace}[\widehat{M}_{C,n}(q,q)] &= \text{Trace}[S_n^*(q)H_q^\top M_C^*(q,q)H_q S_n^*(q)] + o_p(1) \\
&= \text{Trace}[M_C^*(q,q)] + o_p(1) \\
&= \text{Trace}[M_\Lambda^*(q,q)\Sigma_{f,q}] + o_p(1),
\end{aligned}$$

where the second inequality follows from the orthogonality of $H_q S_n^*(q)$ and the last line holds because $M_C^*(q,q) = \Sigma_{f,q}^{1/2} M_\Lambda^*(q,q)\Sigma_{f,q}^{1/2}$. We also note from (8.59) that

$$\frac{1}{k_n N_n}\|\hat{X}_n(q)\|^2 = \frac{1}{k_n N_n}\|\hat{X}_n'(q)\|^2 + o_p(1).$$

Hence, it remains to show that

$$\frac{1}{k_n N_n}\|\hat{X}_n'(q)\|^2 \xrightarrow{\mathbb{P}} \text{Trace}[M_\Lambda^*(q,q)\Sigma_{f,q}] + M_\epsilon(q). \tag{8.92}$$

To show (8.92), we consider the following decomposition:

$$\begin{aligned}
\|\hat{X}_n'(q)\|^2 &= \text{Trace}\left[\hat{X}_n'(q)^\top \hat{X}_n'(q)\right] \\
&= \text{Trace}\left[\tilde{\Lambda}_n^*(q)^\top \tilde{\Lambda}_n^*(q) F_n(q)^\top F_n(q)\right] \\
&\quad + \text{Trace}\left[\left(\mathcal{E}_n(q) + \mathcal{E}_n'(q)\right)^\top \left(\mathcal{E}_n(q) + \mathcal{E}_n'(q)\right)\right] \\
&\quad + 2\,\text{Trace}\left[F_n(q)\tilde{\Lambda}_n^*(q)^\top \left(\mathcal{E}_n(q) + \mathcal{E}_n'(q)\right)\right].
\end{aligned} \tag{8.93}$$

By Lemma 4(e) and (8.61),

$$\frac{1}{k_n N_n}\text{Trace}\left[\tilde{\Lambda}_n^*(q)^\top \tilde{\Lambda}_n^*(q) F_n(q)^\top F_n(q)\right] \xrightarrow{\mathbb{P}} \text{Trace}\left[M_\Lambda^*(q,q)\Sigma_{f,q}\right]. \tag{8.94}$$

In the proof of Lemma 4(c,d), we have shown that

$$\frac{1}{k_n}\left\|\frac{1}{N_n}\left(\mathcal{E}_n(q) + \mathcal{E}_n'(q)\right)^\top \Lambda_n^*(q)\right\|^2 = o_p(1).$$

In addition, by (8.55),

$$\begin{aligned}
&\frac{1}{k_n}\left\|\frac{1}{N_n}\left(\mathcal{E}_n(q) + \mathcal{E}_n'(q)\right)^\top \left(\tilde{\Lambda}_n^*(q) - \Lambda_n^*(q)\right)\right\|^2 \\
&\qquad \le \frac{\|\mathcal{E}_n(q) + \mathcal{E}_n'(q)\|^2}{k_n N_n} \cdot \frac{\left\|\tilde{\Lambda}_n^*(q) - \Lambda_n^*(q)\right\|^2}{N_n} = o_p(1).
\end{aligned}$$

Hence, $\| \left(\mathcal{E}_n(q)+\mathcal{E}_n'(q)\right)^\top \tilde{\Lambda}_n^*(q)\| = o_p(N_n k_n^{1/2})$. Also note that $\|F_n(q)\| = O_p(k_n^{1/2})$. Therefore, by the Cauchy–Schwarz inequality,

$$\left\| \frac{1}{k_n N_n} \left(\mathcal{E}_n(q)+\mathcal{E}_n'(q)\right)^\top \tilde{\Lambda}_n^*(q) F_n(q)^\top \right\| \leq \frac{1}{k_n N_n} \|F_n(q)\| \left\| \left(\mathcal{E}_n(q)+\mathcal{E}_n'(q)\right)^\top \tilde{\Lambda}_n^*(q) \right\| = o_p(1).$$

Consequently,

$$\frac{1}{k_n N_n} \text{Trace} \left[F_n(q) \tilde{\Lambda}_n^*(q)^\top \left(\mathcal{E}_n(q)+\mathcal{E}_n'(q)\right) \right] = o_p(1). \tag{8.95}$$

In view of (8.93), (8.94) and (8.95), (8.92) will be implied by

$$\frac{1}{k_n N_n} \text{Trace} \left[\left(\mathcal{E}_n(q)+\mathcal{E}_n'(q)\right)^\top \left(\mathcal{E}_n(q)+\mathcal{E}_n'(q)\right) \right] \xrightarrow{\mathbb{P}} M_\epsilon(q). \tag{8.96}$$

Finally, we show (8.96). For each j, we denote

$$\begin{aligned} \xi_{n,j} &\equiv \frac{1}{k_n} \sum_{l=1}^{k_n} \left(\frac{\Delta_{i(n,q)+l}^n \epsilon_j}{\sqrt{\Delta_n}} \right)^2, \\ \xi_{n,j}' &\equiv \frac{1}{k_n \Delta_n} \int_{i(n,q)\Delta_n}^{i(n,q)\Delta_n + k_n \Delta_n} \tilde{\sigma}_{j,u}^2 du, \quad \xi_{n,j}'' \equiv \xi_{n,j} - \xi_{n,j}'. \end{aligned}$$

Then, we can decompose

$$\begin{aligned} \frac{1}{k_n N_n} \text{Trace} \left[\mathcal{E}_n(q)^\top \mathcal{E}_n(q) \right] &= \frac{1}{k_n N_n} \sum_{j=1}^{N_n} \sum_{l=1}^{k_n} \left(\frac{\Delta_{i(n,q)+l}^n \epsilon_j}{\sqrt{\Delta_n}} \right)^2 \\ &= \frac{1}{N_n} \sum_{j=1}^{N_n} \xi_{n,j}' + \frac{1}{N_n} \sum_{j=1}^{N_n} \xi_{n,j}''. \end{aligned}$$

We note that conditional on $\mathcal{F}_{i(n,q)\Delta_n}$, the variables $(\xi_{n,j}'')_{1 \leq j \leq N_n}$ are uncorrelated with zero mean and bounded variances. Hence,

$$\frac{1}{N_n} \sum_{j=1}^{N_n} \xi_{n,j}'' = o_p(1). \tag{8.97}$$

In addition, we note that

$$\begin{aligned} \frac{1}{N_n} \sum_{j=1}^{N_n} \xi_{n,j}' - \frac{1}{N_n} \sum_{j=1}^{N_n} \tilde{\sigma}_{j,q}^2 &= \frac{1}{N_n} \sum_{j=1}^{N_n} \frac{1}{k_n \Delta_n} \int_{i(n,q)\Delta_n}^{i(n,q)\Delta_n + k_n \Delta_n} \left(\tilde{\sigma}_{j,u}^2 - \tilde{\sigma}_{j,q}^2 \right) du \\ &= O_p(k_n^{1/2} \Delta_n^{1/2}) = o_p(1). \end{aligned}$$

It readily follows that

$$\frac{1}{N_n} \sum_{j=1}^{N_n} \xi_{n,j}' \xrightarrow{\mathbb{P}} M_\epsilon(q). \tag{8.98}$$

By (8.97) and (8.98),

$$\frac{1}{k_n N_n} \text{Trace} \left[\mathcal{E}_n(q)^\top \mathcal{E}_n(q) \right] \xrightarrow{\mathbb{P}} M_\epsilon(q). \tag{8.99}$$

We further note that

$$\frac{1}{k_n N_n}\text{Trace}\left[\mathcal{E}_n'(q)^\top \mathcal{E}_n'(q)\right] = \frac{1}{k_n N_n}\sum_{j=1}^{N_n}\sum_{l=1}^{k_n}\left(e_{j,l}'\right)^2 = O_p(\Delta_n). \tag{8.100}$$

With an appeal to the Cauchy–Schwarz inequality, we deduce (8.96) from (8.99) and (8.100). This finishes the proof of part (d) of Theorem 2. *Q.E.D.*

8.2.5 Proof of Theorem 3

(a) Firstly, by Theorem 2(c,d), it is obvious that $\tilde{L}_n(\eta,\tau) = O_p(1)$. Hence, the quantile $cv_{n,\alpha} = O_p(1)$. Next, we consider the case under the null hypothesis, so $M_C^*(p,q)$ coincides with $M_C(p,q)$.

We partition $\tilde{\zeta}_q^\top = (\tilde{\zeta}_q^{*\top}, \tilde{\zeta}_q^{0\top})$, where $\tilde{\zeta}_q^*$ is r-dimensional. By Theorem 2(c,d), we have, for $s \in \{\eta,\tau\}$,

$$\begin{cases} \tilde{A}_n(s) = \sum\limits_{p,q\in\{s-,s+\}} \tilde{w}_{n,p}\tilde{w}_{n,q}\tilde{\zeta}_p^{*\top} S_n^*(p) H_p^\top M_C(p,q) H_q S_n^*(q)\tilde{\zeta}_q^* + \sum\limits_{q\in\{s-,s+\}} \tilde{w}_{n,q}^2 M_\epsilon(q) + o_p(1), \\ \tilde{B}_n(\eta,\tau) = \sum\limits_{p\in\{\tau-,\tau+\}}\sum\limits_{q\in\{\eta-,\eta+\}} \tilde{w}_{n,p}\tilde{w}_{n,q}\tilde{\zeta}_p^{*\top} S_n^*(p) H_p^\top M_C(p,q) H_q S_n^*(q)\tilde{\zeta}_q^* + o_p(1). \end{cases}$$

We note that the r-dimensional vectors $H_q S_n^*(q)\tilde{\zeta}_q^*$ are, conditionally on $\mathcal{F}$, standard normal and mutually independent across $q \in \{\tau-,\tau+,\eta-,\eta+\}$. We also observe that for $s \in \{\eta,\tau\}$, $\Delta_{i(n,s)}^n Z \xrightarrow{\mathbb{P}} \Delta Z_s$. Hence,

$$\left(H_q S_n^*(q)\tilde{\zeta}_q^*, \tilde{w}_{n,q}\right)_{q\in\{\tau-,\tau+,\eta-,\eta+\}} \xrightarrow{\mathcal{L}|\mathcal{F}} (\zeta_q, w_q)_{q\in\{\tau-,\tau+,\eta-,\eta+\}}, \tag{8.101}$$

where $\xrightarrow{\mathcal{L}|\mathcal{F}}$ denotes the convergence of conditional law in probability. It follows that

$$\left(\tilde{A}_n(\eta), \tilde{A}_n(\tau), \tilde{B}_n(\eta,\tau)\right) \xrightarrow{\mathcal{L}|\mathcal{F}} \left(\mathcal{A}(\eta), \mathcal{A}(\tau), \mathcal{B}(\eta,\tau)\right).$$

Consequently, $\tilde{L}_n(\eta,\tau) \xrightarrow{\mathcal{L}|\mathcal{F}} \mathcal{L}(\eta,\tau)$. We further note that the $\mathcal{F}$-conditional distribution function of $\mathcal{L}(\eta,\tau)$ is continuous and strictly increasing. Hence, $cv_{n,\alpha} \xrightarrow{\mathbb{P}} cv_\alpha$.

(b) The assertion on the asymptotic level follows from part (a) and Theorem 1. Under the alternative, $\Delta_n^{-1}\hat{V}_n$ diverges to $+\infty$ in probability by Proposition 1. The power property then follows from $cv_{n,\alpha} = O_p(1)$. *Q.E.D.*

8.2.6 Proof of Theorem 4

Proof of Theorem 4. We write v in place of $v(\eta,\tau)$ for simplicity. We note that the winsorization affects at most $\lceil q_n^w N_n \rceil$ terms in the summation. Hence, under the condition $N_n^{1/2} q_n^w \to 0$,

$$\sqrt{N_n}\left(\hat{v}_n - v\right) = \frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\left(\left(\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}\right)^2 - v\right) + o_p(1).$$

Recall that $\hat{\beta}_{n,j,s} - \beta_{j,s} = \xi_{n,j,s} + \Delta^n_{i(n,s)}\tilde{J}_{Y,j}/\Delta^n_{i(n,s)}Z$ and $\chi_{j,\eta,\tau} = \beta_{j,\tau} - \beta_{j,\eta}$, where $\xi_{n,j,s}$ is defined in (8.7). We can rewrite

$$\begin{aligned}
&\frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\left(\left(\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}\right)^2 - v\right) \\
&\quad = \frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\left(\left(\xi_{n,j,\tau} - \xi_{n,j,\eta} + \frac{\Delta^n_{i(n,\tau)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\tau)}Z} - \frac{\Delta^n_{i(n,\eta)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\eta)}Z} + \chi_{j,\eta,\tau}\right)^2 - v\right) \\
&\quad = \frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\left(\chi^2_{j,\eta,\tau} - v\right) \\
&\qquad + \frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\left(\xi_{n,j,\tau} - \xi_{n,j,\eta} + \frac{\Delta^n_{i(n,\tau)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\tau)}Z} - \frac{\Delta^n_{i(n,\eta)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\eta)}Z}\right)^2 \\
&\qquad + \frac{2}{\sqrt{N_n}}\sum_{j=1}^{N_n}\chi_{j,\eta,\tau}\left(\frac{\Delta^n_{i(n,\tau)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\tau)}Z} - \frac{\Delta^n_{i(n,\eta)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\eta)}Z}\right) \\
&\qquad + \frac{2}{\sqrt{N_n}}\sum_{j=1}^{N_n}\chi_{j,\eta,\tau}\left(\xi_{n,j,\tau} - \xi_{n,j,\eta}\right).
\end{aligned} \tag{8.102}$$

Under Assumption 9, the variables $(\chi^2_{j,\eta,\tau})_{j\geq 1}$ are $\mathcal{C}_{\eta-}$-conditionally independent. By using a central limit theorem under the $\mathcal{C}_{\eta-}$-conditional probability, we deduce that

$$\frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\left(\chi^2_{j,\eta,\tau} - \mathbb{E}[\chi^2_{j,\eta,\tau}|\mathcal{C}_{\eta-}]\right) \stackrel{\mathcal{L}|\mathcal{C}_{\eta-}}{\longrightarrow} \mathcal{S}.$$

By Assumption 9(iv), we further deduce that

$$S_n \equiv \frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\left(\chi^2_{j,\eta,\tau} - v\right) \stackrel{\mathcal{L}|\mathcal{C}_{\eta-}}{\longrightarrow} \mathcal{S}. \tag{8.103}$$

We further note that the second term on the right-hand side of (8.102) satisfies

$$\frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\left(\xi_{n,j,\tau} - \xi_{n,j,\eta} + \frac{\Delta^n_{i(n,\tau)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\tau)}Z} - \frac{\Delta^n_{i(n,\eta)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\eta)}Z}\right)^2 = O_p(N_n^{1/2}\Delta_n) = o_p(1). \tag{8.104}$$

Turning to the third term on the right-hand side of (8.102), we note that the variables

$$\chi_{j,\eta,\tau}\left(\frac{\Delta^n_{i(n,\tau)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\tau)}Z} - \frac{\Delta^n_{i(n,\eta)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\eta)}Z}\right), \quad 1 \leq j \leq N_n,$$

are, conditionally on $\mathcal{C}_{\eta-}$, uncorrelated with zero mean. Hence,

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\chi_{j,\eta,\tau}\left(\frac{\Delta^n_{i(n,\tau)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\tau)}Z} - \frac{\Delta^n_{i(n,\eta)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\eta)}Z}\right)\right)^2\middle|\mathcal{C}_{\eta-}\right] \\
&\quad = \frac{1}{N_n}\sum_{j=1}^{N_n}\mathbb{E}\left[\chi^2_{j,\eta,\tau}\left(\frac{\Delta^n_{i(n,\tau)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\tau)}Z} - \frac{\Delta^n_{i(n,\eta)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\eta)}Z}\right)^2\middle|\mathcal{C}_{\eta-}\right] = O_p(\Delta_n),
\end{aligned} \tag{8.105}$$

which further implies that

$$\frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\chi_{j,\eta,\tau}\left(\frac{\Delta^n_{i(n,\tau)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\tau)}Z}-\frac{\Delta^n_{i(n,\eta)}\tilde{J}_{Y,j}}{\Delta^n_{i(n,\eta)}Z}\right)=O_p(\Delta_n^{1/2}). \tag{8.106}$$

It remains to consider the fourth term on the right-hand side of (8.102). Recall that

$$\begin{aligned}\xi_{n,j,\eta} &= \frac{\Delta^n_{i(n,\eta)}Y'_j-\beta_{j,\eta}\Delta^n_{i(n,\eta)}Z'}{\Delta^n_{i(n,\eta)}Z},\\ \xi_{n,j,\tau} &= \frac{\Delta^n_{i(n,\tau)}Y'_j-\beta_{j,\tau}\Delta^n_{i(n,\tau)}Z'}{\Delta^n_{i(n,\tau)}Z}=\frac{\Delta^n_{i(n,\tau)}Y'_j-(\beta_{j,\eta}+\chi_{j,\eta,\tau})\,\Delta^n_{i(n,\tau)}Z'}{\Delta^n_{i(n,\tau)}Z}.\end{aligned}$$

We first observe that

$$\frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\chi_{j,\eta,\tau}\xi_{n,j,\eta}=\frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\frac{\chi_{j,\eta,\tau}\left(\Delta^n_{i(n,\eta)}Y'_j-\beta_{j,\eta}\Delta^n_{i(n,\eta)}Z'\right)}{\Delta^n_{i(n,\eta)}Z}.$$

Under Assumption 9(iii), we see that

$$\mathbb{E}[\chi_{j,\eta,\tau}\Delta^n_{i(n,\eta)}Y'_j|\mathcal{C}_{\eta-}]=\mathbb{E}[\chi_{j,\eta,\tau}|\mathcal{C}_{\eta-}]\mathbb{E}[\Delta^n_{i(n,\eta)}Y'_j|\mathcal{C}_{\eta-}]=0. \tag{8.107}$$

Since $\Delta^n_{i(n,\eta)}Z'$ is $\mathcal{C}$-measurable and $\chi_{j,\eta,\tau}$ is $\mathcal{C}_{\eta-}$-conditionally independent of $\beta_{j,\eta}$, we have

$$\mathbb{E}\left[\chi_{j,\eta,\tau}\beta_{j,\eta}\Delta^n_{i(n,\eta)}Z'|\mathcal{C}_{\eta-}\right]=\mathbb{E}\left[\chi_{j,\eta,\tau}|\mathcal{C}_{\eta-}\right]\mathbb{E}\left[\beta_{j,\eta}|\mathcal{C}_{\eta-}\right]\Delta^n_{i(n,\eta)}Z'=0. \tag{8.108}$$

From (8.107) and (8.108), and the fact that $\Delta^n_{i(n,\eta)}Z$ is $\mathcal{C}$-measurable, we deduce that

$$\mathbb{E}\left[\chi_{j,\eta,\tau}\xi_{n,j,\eta}|\mathcal{C}_{\eta-}\right]=0. \tag{8.109}$$

By using a similar argument, we can show that $\mathbb{E}\left[\chi_{j,\eta,\tau}\xi_{n,j,\eta}\chi_{m,\eta,\tau}\xi_{n,m,\eta}|\mathcal{C}_{\eta-}\right]=0$ for $j\neq m$. Hence,

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\chi_{j,\eta,\tau}\xi_{n,j,\eta}\right)^2\middle|\mathcal{C}_{\eta-}\right]=\frac{1}{N_n}\sum_{j=1}^{N_n}\mathbb{E}\left[\chi^2_{j,\eta,\tau}\xi^2_{n,j,\eta}|\mathcal{C}_{\eta-}\right]=O_p(\Delta_n),$$

which implies that

$$\frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\chi_{j,\eta,\tau}\xi_{n,j,\eta}=O_p(\Delta_n^{1/2}). \tag{8.110}$$

Next, we observe

$$\begin{aligned}\frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\chi_{j,\eta,\tau}\xi_{n,j,\tau} &= \frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\chi_{j,\eta,\tau}\left(\frac{\Delta^n_{i(n,\tau)}Y'_j-(\beta_{j,\eta}+\chi_{j,\eta,\tau})\,\Delta^n_{i(n,\tau)}Z'}{\Delta^n_{i(n,\tau)}Z}\right)\\ &= \frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\chi_{j,\eta,\tau}\left(\frac{\Delta^n_{i(n,\tau)}Y'_j-\beta_{j,\eta}\Delta^n_{i(n,\tau)}Z'}{\Delta^n_{i(n,\tau)}Z}\right)\\ &\quad-\sqrt{N_n\Delta_n}\left(\frac{1}{N_n}\sum_{j=1}^{N_n}\chi^2_{j,\eta,\tau}\right)\left(\frac{\Delta^n_{i(n,\tau)}Z'/\sqrt{\Delta_n}}{\Delta^n_{i(n,\tau)}Z}\right).\end{aligned} \tag{8.111}$$

Following a similar argument that leads to (8.110), we can show that

$$\frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\chi_{j,\eta,\tau}\left(\frac{\Delta_{i(n,\tau)}^n Y_j' - \beta_{j,\eta}\Delta_{i(n,\tau)}^n Z'}{\Delta_{i(n,\tau)}^n Z}\right) = O_p(\Delta_n^{1/2}).$$

In addition, we note that

$$\frac{\Delta_{i(n,\tau)}^n Z'/\sqrt{\Delta_n}}{\Delta_{i(n,\tau)}^n Z} \xrightarrow{\mathcal{L}\text{-}s} -\frac{\sqrt{\kappa_\tau}\zeta_{\tau-}' + \sqrt{1-\kappa_\tau}\zeta_{\tau+}'}{\Delta Z_\tau}. \tag{8.112}$$

Since the sequence in (8.112) is $\mathcal{C}_{\eta-}$-measurable, by Proposition 5 of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) and (8.103), we further deduce the following $\mathcal{C}_{\eta-}$-stable convergence in law:

$$\left(S_n, \frac{\Delta_{i(n,\tau)}^n Z'/\sqrt{\Delta_n}}{\Delta_{i(n,\tau)}^n Z}\right) \xrightarrow{\mathcal{L}\text{-}s} \left(\mathcal{S}, -\frac{\sqrt{\kappa_\tau}\zeta_{\tau-}' + \sqrt{1-\kappa_\tau}\zeta_{\tau+}'}{\Delta Z_\tau}\right). \tag{8.113}$$

Since $N_n^{-1}\sum_{j=1}^{N_n}\chi_{j,\eta,\tau}^2 \xrightarrow{\mathbb{P}} v$ and v is $\mathcal{C}_{\eta-}$-measurable,

$$\left(S_n, \frac{1}{\sqrt{N_n}}\sum_{j=1}^{N_n}\chi_{j,\eta,\tau}\xi_{n,j,\tau}\right) \xrightarrow{\mathcal{L}\text{-}s} \left(\mathcal{S}, \frac{\sqrt{\theta}v\left(\sqrt{\kappa_\tau}\zeta_{\tau-}' + \sqrt{1-\kappa_\tau}\zeta_{\tau+}'\right)}{\Delta Z_\tau}\right). \tag{8.114}$$

From (8.104), (8.106), (8.110) and (8.114), the assertion of Theorem 4(a) readily follows. Part (b) can be proved similarly. *Q.E.D.*

8.2.7 Proof of Theorem 5

Proof of Theorem 5. We first observe that

$$\frac{1}{N_n}\sum_{j=1}^{N_n}Var[\chi_{j,\eta,\tau}^2|\mathcal{C}_{\eta-}] = \frac{1}{N_n}\sum_{j=1}^{N_n}\mathbb{E}[\chi_{j,\eta,\tau}^4|\mathcal{C}_{\eta-}] - \frac{1}{N_n}\sum_{j=1}^{N_n}\left(\mathbb{E}[\chi_{j,\eta,\tau}^2|\mathcal{C}_{\eta-}]\right)^2.$$

Under Assumption 9(iv),

$$\begin{aligned}\frac{1}{N_n}\sum_{j=1}^{N_n}\left(\mathbb{E}[\chi_{j,\eta,\tau}^2|\mathcal{C}_{\eta-}]\right)^2 &= \frac{1}{N_n}\sum_{j=1}^{N_n}\left(\mathbb{E}[\chi_{j,\eta,\tau}^2|\mathcal{C}_{\eta-}] - v(\eta,\tau)\right)^2 \\ &\quad + \frac{2v(\eta,\tau)}{N_n}\sum_{j=1}^{N_n}\left(\mathbb{E}[\chi_{j,\eta,\tau}^2|\mathcal{C}_{\eta-}] - v(\eta,\tau)\right) + v(\eta,\tau)^2 \xrightarrow{\mathbb{P}} v(\eta,\tau)^2.\end{aligned}$$

Then by Assumption 9(v),

$$\frac{1}{N_n}\sum_{j=1}^{N_n}\mathbb{E}[\chi_{j,\eta,\tau}^4|\mathcal{C}_{\eta-}] \xrightarrow{\mathbb{P}} \Sigma_S + v(\eta,\tau)^2. \tag{8.115}$$

By a similar proof as that in Proposition 1,

$$\frac{1}{N_n}\sum_{j=1}^{N_n}(|\hat{\beta}_{n,j,\tau} - \hat{\beta}_{n,j,\eta}| \wedge \bar{B}_{n,\eta,\tau})^4 = \frac{1}{N_n}\sum_{j=1}^{N_n}\mathbb{E}[\chi_{j,\eta,\tau}^4|\mathcal{C}_{\eta-}] + o_p(1). \tag{8.116}$$

By (8.115) and (8.116), $N_n^{-1}\sum_{j=1}^{N_n}(|\hat{\beta}_{n,j,\tau}-\hat{\beta}_{n,j,\eta}|\wedge\bar{B}_{n,\eta,\tau})^4 \xrightarrow{\mathbb{P}} \Sigma_S+v(\eta,\tau)^2$. We also note that $\hat{v}_n \xrightarrow{\mathbb{P}} v(\eta,\tau)$. Therefore, $\hat{\Sigma}_{S,n} \xrightarrow{\mathbb{P}} \Sigma_S$. Since $\hat{\Sigma}_{Z,n,\tau\pm} \xrightarrow{\mathbb{P}} \hat{\Sigma}_{Z,\tau\pm}$, the assertion of part (a) readily follows from the construction of Algorithm 2. Part (b) follows as a corollary of part (a). *Q.E.D.*