

# Unified Inference for Nonlinear Factor Models from Panels with Fixed and Large Time Span\*

Torben G. Andersen<sup>†</sup> Nicola Fusari<sup>‡</sup> Viktor Todorov<sup>§</sup> Rasmus T. Varneskov<sup>¶</sup>

March 1, 2018

## Abstract

We provide unifying inference theory for parametric nonlinear factor models based on a panel of noisy observations. The panel has a large cross-section and a time span that may be either small or large. Moreover, we incorporate an additional source of information, provided by noisy observations on some known functions of the factor realizations. The estimation is carried out via penalized least squares, i.e., by minimizing the  $L_2$  distance between observations from the panel and their model-implied counterparts, augmented by a penalty for the deviation of the extracted factors from the noisy signals of them. When the time dimension is fixed, the limit distribution of the parameter vector is mixed Gaussian with conditional variance depending on the path of the factor realizations. On the other hand, when the time span is large, the convergence rate is faster and the limit distribution is Gaussian with a constant variance. In this case, however, we incur an incidental parameter problem since, at each point in time, we need to recover the concurrent factor realizations. This leads to an asymptotic bias that is absent in the setting with a fixed time span. In either scenario, the limit distribution of the estimates for the factor realizations is mixed Gaussian, but is related to the limiting distribution of the parameter vector only in the scenario with a fixed time horizon. Although the limit behavior is very different for the small versus large time span, we develop a feasible inference theory that applies, without modification, in either case. Hence, the user need not take a stand on the relative size of the time dimension of the panel. Similarly, we propose a time-varying data-driven weighting of the penalty in the objective function, which enhances efficiency by adapting to the relative quality of the signal for the factor realizations.

**Keywords:** Asymptotic Bias, Incidental Parameter Problem, Inference, Large Data Sets, Nonlinear Factor Model, Options, Panel Data, Stable Convergence, Stochastic Volatility.

**JEL classification:** C51, C52, G12.

---

\*Andersen and Varneskov gratefully acknowledge support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. The work is partially supported by NSF Grant SES-1530748. We would like to thank the guest editor (Frank Diebold), three anonymous referees, Marcelo Fernandez (our discussant), Christian Gourieroux as well as seminar participants at various conferences and seminars for many helpful comments and suggestions.

<sup>†</sup>Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208; NBER, Cambridge, MA; and CREATES, Aarhus, Denmark; e-mail: t-andersen@northwestern.edu.

<sup>‡</sup>The Johns Hopkins University Carey Business School, Baltimore, MD 21202; e-mail: nicola.fusari@jhu.edu.

<sup>§</sup>Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208; e-mail: v-todorov@northwestern.edu.

<sup>¶</sup>Department of Finance, Kellogg School of Management, Northwestern University, Evanston, IL 60208; CREATES, Aarhus, Denmark; Multi Assets at Nordea Asset Management, Copenhagen, Denmark e-mail: rasmus.varneskov@kellogg.northwestern.edu.

# 1 Introduction

We develop unified inference theory for parametric nonlinear factor models from panels of noisy observations, which have an asymptotically increasing cross-sectional dimension and a time span that is either fixed or increasing. In particular, we provide inference for the time-invariant parameters of the model as well as the period-by-period latent factor realizations. Importantly, although the limit theories for the fixed and increasing time span scenarios are very different, our feasible implementation is identical across the two cases, enabling empirical work that is robust to the relative length of the time series. The theory is presented in the context of an option panel, i.e., a collection of option prices observed across time and written on a given underlying asset. Our option prices are recorded over a sample period, where each observation date typically will have a different set of option strikes and tenors (times-to-maturity) available. With the rapid increase of trading in derivatives over recent years, such option panels are readily accessible and provide a unique source of information regarding the state of the economy and the pricing of volatility and tail risk in financial markets. Moreover, with suitable adaptation, our theory is applicable for analysis of economic systems with similar panel structures, such as household and firm data or term structure and large-scale macro DSGE models.

Given a specific parametric model, the theoretical value of an option is a nonlinear function of the (known) option strike and tenor as well as of the (unknown) time-invariant parameters and the latent state vector (factor) realization.<sup>1</sup> Under correct specification, the observed option price equals the model-implied theoretical value plus an observation error, centered at zero. Therefore, unknown parameters and factor realizations can be estimated by minimizing the  $L_2$  distance between observed and model-implied option prices (i.e., nonlinear least squares). We further utilize nonparametric volatility measures based on high-frequency data of the underlying asset and penalize deviations of model-implied volatilities (determined by the factor realizations) from their nonparametric high-frequency estimates in the objective function. We refer to this estimation approach as penalized least squares (PLS). In alternative applications, penalization of this form may be useful if the latent factors can be associated directly with economic variables that are observed with measurement error.

Andersen et al. (2015a) develop the inference theory for PLS and a fixed time span. In this setting, both the parameters and factor realizations are estimated at rate  $\sqrt{N}$ , where  $N$  denotes the average number of options per observation date. The limiting distribution is mixed Gaussian, i.e., it is Gaussian

---

<sup>1</sup>As is common, we rely on a fully parametric specification of the risk-neutral return distribution. This ensures a tractable representation for which we are able to develop comprehensive estimation and inference techniques, in spite of the highly nonlinear dependence of the option prices on the underlying latent state variables (or factors), and the strong dependence in the option prices across nearby strikes and tenors as well as in the time series dimension. In addition, without an explicit dynamic representation of the risk-neutral distribution, it is hard to impose fundamental no-arbitrage conditions on the system. Importantly, a battery of diagnostic tests is available to guard against model misspecification, following the procedures of Andersen et al. (2015a).

conditional on the observed path of the factors. In particular, this enables the approach to convey explicitly how the precision of the estimates for the option prices and factor realizations changes from period-to-period, depending on the level of volatility of the underlying asset, the option coverage over the strike ranges and tenors, the size of the measurement errors, etc.

In this paper, we develop the asymptotic theory for the case where the time span of the panel may increase at a rate equal to that of the cross-section. This scenario induces a so-called incidental parameter problem (Neyman and Scott (1948) and Lancaster (2000)). As the time span increases, the quantities to estimate (the factor realizations in our setting) also increase. For the long time span ( $T$  large), we show that the rate of convergence of the parameter vector rises to  $\sqrt{NT}$ . Unlike for the fixed time span, the limiting distribution for the parameters is now Gaussian. In addition, due to the incidental parameter problem, there is an asymptotic bias of order  $O_p(1/N)$ . We propose model-based de-biasing, which allows for feasible inference on the parameters. On the other hand, the rate of recovery for the factor realizations remains  $\sqrt{N}$  and the asymptotic limit is still mixed Gaussian. We derive the joint asymptotic distribution of the parameters and the factor realizations. This is critical for exploring various economic hypotheses. For example, this allows for testing model performance across different time periods of interest, e.g., calm versus turbulent market conditions.

Despite the fact that the behavior of our PLS estimator is very different in the fixed and increasing time span scenarios – both in terms of rate of convergence and limiting distribution – we show that feasible inference can be performed in an identical manner for either of the two asymptotic schemes, provided it suitably accommodates features of both limit theories. The reason is that the estimate of the asymptotic bias for the long span case is asymptotically negligible if the time span actually is fixed. Likewise, estimates for the dependence between the parameters and factor realizations in the fixed span case become asymptotically negligible in the long span setting. Finally, the estimates for the conditional variance of the parameters automatically represent long-run asymptotic variance estimates in the long span case. The above results depend critically on the fact that the limit theory is  $\mathcal{F}$ -conditionally stable ( $\mathcal{F}$  denotes the sigma-algebra of the underlying probability space). Among other things, this implies that we can apply standard normalization techniques for the construction of confidence intervals and testing, even when the asymptotic variances are random in the limit. From a practical point of view, this set of findings is important because the applied econometrician need not take a stand on whether the time span of the panel is of a length that justifies appealing to one limit theory versus the other – the procedure automatically adapts to the relevant features of the data.

Another novel aspect of our inference procedure is the use of extra information on the latent factor. This is reminiscent of the factor augmented vector autoregression (FAVAR) proposed by Bernanke et al. (2005). In our context, the signal for the factor realization takes the form of a high-frequency

nonparametric estimator of volatility. The asymptotic role of the penalization based on this volatility estimator depends on the relative size of the option cross-section versus the sampling frequency of the underlying asset price. This determines the relative (asymptotic) efficiency in recovering the volatility from options and high-frequency data, respectively. If the former is more efficient, the specification of the penalization term ensures it has no asymptotic impact while, if the latter is more efficient, the penalization will play an asymptotic role. Therefore, the user need not take a stand on whether the option or high-frequency data are more efficient for recovering volatility. As for the span of the panel, the procedure simply adapts to the situation at hand.

To further improve efficiency, we extend the estimation to weighted penalized least squares. We do so by forming a consistent estimate for the option observation error variance across each cross-section from a first-stage PLS estimate of the parameters and factor realizations. We then weight the squared (pricing) errors in the objective function by the estimated option error variance and the penalization term by a measure of its (conditional) asymptotic variance. Again, the approach applies for both the fixed and long span case, without modification. It allows the relative weight of the penalization in the objective function to shift over the days in the sample, reflecting the time-variation in the informativeness of the option versus high-frequency data.

Our main theoretical contributions can be summarized as follows. First, we analyze nonlinear panel models in the joint asymptotic setting where the time and cross-sectional dimensions can grow at the same rate in the presence of stochastic latent factors (fixed effects). Second, we augment the estimation by including noisy factor signals in the form of nonparametric high-frequency volatility estimates. Third, we derive the joint asymptotic limit of the latent factors, the parameters and the high-frequency volatility estimates in the various asymptotic setups ( $T$  fixed or increasing up to the rate of  $N$ ), which is mixed Gaussian. Fourth, and perhaps most importantly, we show that feasible inference can be conducted in a way that works irrespective of whether  $T$  is fixed or increasing.

The above contributions can be compared with prior work in the following way. First, Andersen et al. (2015a) (see also Andrews (2005) and Kuersteiner and Prucha (2013)) consider only the fixed time span case. Second, Hahn and Kuersteiner (2002), Hahn and Newey (2004), Arellano and Hahn (2007), Arellano and Bonhomme (2009), Hahn and Kuersteiner (2011) and Fernandez-Val and Weidner (2016) consider only the case when  $T$  and  $N$  grow at the same rate, while Gagliardini and Gourieroux (2014) study the case of  $T$  growing at a slower rate than  $N$ . These papers do not allow for noisy factor signals in the estimation nor do they consider the joint asymptotic behavior of factor realizations and parameters. Third, Bai and Ng (2002) and Bai (2003) consider estimation of linear factor models. The major difference from that strand of work is that our factor loadings (i.e., the sensitivity of option prices to factors) are “structural,” i.e., driven by a parameter vector of fixed size, while Bai and Ng

(2002) and Bai (2003) treat them as unknown, unrestricted coefficients. Hence, in their case, the number of parameters (factor loadings) grows in proportion to the cross-sectional dimension of the panel. This implies a very different asymptotic behavior of the factor estimates depending on the relative size of  $T$  versus  $N$ . In particular, when  $T$  is fixed, Bai (2003) shows that relatively strong assumptions, including homoskedasticity of the observation error, are necessary for consistent factor extraction. Finally, our paper relates to Gagliardini et al. (2011), who consider estimation using data on options as well as the underlying asset. Their asymptotic setup is different as they treat the case of  $N$  fixed and depend importantly on (exact) nonparametric inference from the underlying asset data.<sup>2</sup>

We conduct numerical experiments to assess the practical relevance of the developed limit theory. We rely on a popular two-factor stochastic volatility model for the underlying asset price and consider strike and tenor configurations that mimic typical option data sets. The findings illustrate the advantages of our unified inference framework. For the factor extraction, we find the “stabilization” of the standard error, captured in the long-span asymptotic setting, to become effective across widely different horizons for different factors. Thus, neither the fixed nor long span asymptotics is satisfactory across all factors simultaneously. In addition, we find fairly strong dependence in the extraction error for the factor realizations. Such dependence is of higher order in the long-span asymptotic setup and is, thus, ignored if the associated asymptotic is used. Finally, the asymptotic biases, which arise only in the long-span asymptotic setting, play a nontrivial role for some parameters, even over relatively short time spans. As noted, our feasible unified inference procedure adapts by accommodating those features of the fixed or long span asymptotics that are relevant for the situation at hand.

The rest of the paper is organized as follows. In Section 2, we introduce the parametric nonlinear factor model for the option panel, the option observation scheme and the nonparametric high-frequency volatility estimator. In Section 3, we derive the asymptotic behavior of the PLS estimator in the long time span case. Section 4 develops unified feasible inference theory for panels with long and short time spans. In Section 5, we extend the analysis to weighted penalized least squares. Section 6 contains numerical experiments, illustrating the gains from the unified inference in the context of a popular parametric asset pricing model. Section 7 concludes. The Appendix in Section A contains the formal assumptions for the theoretical analysis as well as the proofs. Finally, Appendices B and C provide details about the numerical analysis and additional simulation results.

---

<sup>2</sup>There is a large body of earlier empirical option pricing work, see e.g., Bates (2003), Christoffersen and Jacobs (2004), Eraker (2004), Pan (2002) and the survey by Garcia et al. (2010). The emphasis of these studies, however, is on extracting information from a limited (fixed  $N$ ) number of options, which differs fundamentally from our intention of using the full set of available options, leading to the setup involving an asymptotically increasing cross-section. Moreover, unlike our analysis, the fixed cross-section typically necessitates parametric assumptions for the observation error and a parametric specification for the factor evolution.

## 2 Setup

We start our analysis by introducing the nonlinear factor model for the panel of option prices. We then describe the observation scheme for the options as well as the high-frequency return data for the nonparametric volatility measurement.

### 2.1 A Parametric Nonlinear Factor Model for the Option Panel

The options are written on an underlying process, denoted by  $X$ , which is defined on a filtered probability space  $(\Omega^{(0)}, \mathcal{F}^{(0)}, (\mathcal{F}_t^{(0)})_{t \geq 0}, \mathbb{P}^{(0)})$ . The process  $X$  is governed by the following general dynamics,

$$\frac{dX_t}{X_{t-}} = \alpha_t dt + \sqrt{V_t} dW_t^{\mathbb{P}} + \int_{x > -1} x \tilde{\mu}^{\mathbb{P}}(dt, dx), \quad (1)$$

where  $\alpha_t$  and  $V_t$  are càdlàg;  $W_t^{\mathbb{P}}$  is a  $\mathbb{P}^{(0)}$ -Brownian motion;  $\mu^{\mathbb{P}}$  is a random measure counting the jumps in  $X$ , with compensator  $\tilde{\nu}^{\mathbb{P}}(dt, dx) = a_t dt \otimes \nu^{\mathbb{P}}(dx)$  for some process  $a_t$  and Lévy measure  $\nu^{\mathbb{P}}(dx)$ , and with associated martingale measure  $\tilde{\mu}^{\mathbb{P}} = \mu^{\mathbb{P}} - \tilde{\nu}^{\mathbb{P}}$ . We denote expectations under  $\mathbb{P}^{(0)}$  by  $\mathbb{E}(\cdot)$ . This formulation is very general; essentially all parametric models in the literature satisfy equation (1).

The so-called jump activity is bounded by the parameter  $r$ , i.e., we have,

$$\int_{\mathbb{R}} (|x|^r \wedge 1) \nu^{\mathbb{P}}(dx) < \infty, \quad (2)$$

for some  $r \in [0, 1)$  and we define  $x \wedge y = \min(x, y)$ . The leading case is  $r = 0$ , corresponding to jumps of finite activity, which have been predominantly used in prior work, while values of  $r > 0$  allow for infinitely active jumps. The parameter  $r$  appears explicitly in our asymptotic results.

Under no-arbitrage, a risk-neutral probability measure,  $\mathbb{Q}$ , is guaranteed to exist, see, e.g., Section 6.K in Duffie (2001), and is locally equivalent to  $\mathbb{P}^{(0)}$ . It transforms discounted asset prices into (local) martingales. In particular, for  $X$  under  $\mathbb{Q}$ , we have,

$$\frac{dX_t}{X_{t-}} = (r_t - \delta_t) dt + \sqrt{V_t} dW_t^{\mathbb{Q}} + \int_{x > -1} x \tilde{\mu}^{\mathbb{Q}}(dt, dx), \quad (3)$$

where  $r_t$  is the risk-free interest rate and  $\delta_t$  is the dividend yield,  $W_t^{\mathbb{Q}}$  denotes a  $\mathbb{Q}$ -Brownian motion, and the jump martingale measure is defined with respect to a risk-neutral compensator  $\tilde{\nu}^{\mathbb{Q}}(dt, dx)$ .

The theoretical values of derivatives written on  $X$  equal their expected discounted payoffs under the risk-neutral measure,  $\mathbb{Q}$ . In particular, if we denote the prices of European-style out-of-the-money (OTM) options on  $X$  at time  $t$  by  $O_{t,k,\tau}$ , we have,

$$O_{t,k,\tau} = \begin{cases} \mathbb{E}_t^{\mathbb{Q}} \left[ e^{-\int_t^{t+\tau} r_s ds} (X_{t+\tau} - K)^+ \right], & \text{if } K > F_{t,t+\tau}, \\ \mathbb{E}_t^{\mathbb{Q}} \left[ e^{-\int_t^{t+\tau} r_s ds} (K - X_{t+\tau})^+ \right], & \text{if } K \leq F_{t,t+\tau}, \end{cases} \quad (4)$$

where  $\tau$  is the tenor,  $K$  is the strike price,  $F_{t,t+\tau}$  is the futures price of the underlying asset,  $X$ , at time  $t$  for the future date  $t + \tau$ , and  $k = \ln(K/F_{t,t+\tau})$  is the log-moneyness.<sup>3</sup>

We model the risk-neutral probability measure parametrically, rendering the option prices nonlinear functions of  $k$  and  $\tau$  as well as a parameter vector and a latent state vector.<sup>4</sup> In particular, under the risk-neutral model, the diffusive volatility and jump processes are governed by a (latent) state vector, or factors, so that  $V_t = \xi_1(\mathbf{S}_t)$  and  $\tilde{\nu}^{\mathbb{Q}}(dt, dx) = \xi_2(\mathbf{S}_t) \otimes \nu^{\mathbb{Q}}(dx)$ , where  $\nu^{\mathbb{Q}}(dx)$  is a Lévy measure;  $\xi_1(\cdot)$  and  $\xi_2(\cdot)$  are known functions; and  $\mathbf{S}_t$  denotes the  $p \times 1$  state vector taking values in  $\mathcal{S} \subset \mathbb{R}^p$ . Importantly, we only require  $\mathbf{S}_t$  to be an  $\mathcal{F}_t^{(0)}$ -adapted stochastic process, whose dynamics need not be specified explicitly for the purposes of our analysis.<sup>5</sup> We further assume that the model-implied option prices, quoted in terms of Black-Scholes implied variance (BSIV), are given by the functions  $\kappa(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$ , where  $k$  and  $\tau$  denote the moneyness and tenor of the option, and  $\mathbf{Z}$  and  $\boldsymbol{\theta}$  denote a particular value of the state and parameter vectors, respectively. The parameter vector is of dimension  $q \times 1$  and takes values  $\boldsymbol{\theta} \in \Theta$ , with  $\Theta$  being a compact subset of  $\mathbb{R}^q$ . We denote the true values of the latent factors at time  $t$  by  $\mathbf{S}_t$  and the true value of the parameter vector by  $\boldsymbol{\theta}_0$ . We note that the only time-dependence of the option prices is through the value of the state vector.<sup>6</sup> Our primary interest in this paper is on inference for  $\boldsymbol{\theta}_0$  as well as the factor realizations  $\{\mathbf{S}_t\}_{t \geq 0}$ .

The above framework for option pricing nests most continuous-time models used in empirical work, including the affine jump-diffusion class of Duffie et al. (2000).<sup>7</sup> The only requirement for implementation of our inference procedures is knowledge of the function  $\kappa(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$ , either in closed-form, up to numerical integration, or via simulation.<sup>8</sup> We stress, in particular, that we do allow for risk-neutral specifications in which there is dependence between the latent state vector  $\mathbf{S}_t$  and either  $W_t^{\mathbb{Q}}$  or the jump measure  $\mu^{\mathbb{Q}}$ . That is, we accommodate the so-called leverage effect, working through either the diffusive or the jump component of  $X_t$ , or both.

## 2.2 Observation Scheme

The time span of the panel is  $[0, T]$ , and we have observations from the option surface at given integer times  $t = 1, \dots, T$ . For each date, the options cover a significant range of strikes  $k$  and several tenors,

<sup>3</sup>Note that by including the subscript  $t$  in the notation of the option price, we allow for the option price to change from one day to another, even when the log-moneyness of the option,  $k$ , stays the same.

<sup>4</sup>Our focus is on the option panel, which is fully characterized through a model for the risk-neutral probability. We impose no assumptions on the  $\mathbb{P}$  dynamics beyond what is implied by our model for the  $\mathbb{Q}$  dynamics and no-arbitrage.

<sup>5</sup>For convenience, we assume the functions  $\xi_1(\cdot)$  and  $\xi_2(\cdot)$  do not depend on the parameter vector. This is almost universally satisfied in empirical applications. In addition,  $r_t$  and  $\delta_t$  should be known functions of the state vector  $\mathbf{S}_t$ .

<sup>6</sup>Thus, we implicitly assume stationarity of the risk-neutral distribution through the state vector.

<sup>7</sup>We only assume that the diffusion coefficient and the jump intensity of  $X$  are governed by a finite-dimensional state vector under the risk-neutral probability measure, but make no such assumption for the statistical measure.

<sup>8</sup>In our numerical exercise below, it takes 0.001 seconds to evaluate an entire option cross-section on a given trading day, and 0.06 seconds to optimize over the factors on a single day, i.e., to perform the optimization in equation (12) below on a single day.

$\tau$ . Specifically, at time  $t$ , we observe the option prices  $\{O_{t,k_j,\tau_j}\}_{j=1,\dots,N_t}$  for some (large) integer  $N_t$ , where the index  $j$  runs across the full set of strikes and tenors available on that day. The number of options for the maturity  $\tau$  is denoted  $N_t^\tau$ , so  $N_t = \sum_\tau N_t^\tau$ , where  $N_t^\tau$  and  $N_t$  are  $\mathcal{F}_t^{(0)}$ -adapted.

Given the heterogeneity of the available option data, we allow the panel to be unbalanced, so both the total number of options and the available strike-tenor combinations  $(k, \tau)$  are free to vary across time. In particular, we assume,

$$\frac{N_t^\tau}{N_t} \approx \pi_t^\tau \quad \text{and} \quad \frac{N_t}{N} \approx \varsigma_t, \quad (5)$$

where  $\pi_t^\tau$  and  $\varsigma_t$  are positive-valued processes, and  $N$  is an unobserved non-random number, representing the ‘‘average size of the cross-section’’. We refer to Section A.1 of the Appendix for a formal statement regarding the nature of the approximations in equation (5).

For each pair,  $(t, \tau)$ , we denote the minimum and maximum log-moneyness by  $\underline{k}(t, \tau)$  and  $\bar{k}(t, \tau)$ , respectively. At time  $t$  and tenor  $\tau$ , we define the  $\mathcal{F}_t^{(0)}$ -adapted moneyness grid as,

$$\underline{k}(t, \tau) = k_{t,\tau}(1) < k_{t,\tau}(2) \dots < k_{t,\tau}(N_t^\tau) = \bar{k}(t, \tau),$$

and we further set  $\Delta_{t,\tau}(i) = k_{t,\tau}(i) - k_{t,\tau}(i-1)$ , for  $i = 2, \dots, N_t^\tau$ . Our asymptotic scheme sequentially adds new strikes to the existing ones within  $[\underline{k}(t, \tau), \bar{k}(t, \tau)]$ , so that  $\Delta_{t,\tau}(i)$  shrinks towards zero, asymptotically, while allowing for the strikes to cluster relatively more densely within certain regions of the range, i.e.,

$$\Delta_{t,\tau}(i) \approx \frac{1}{N_t^\tau} \psi_{t,\tau}(k_{t,\tau}(i)), \quad i = 2, \dots, N_t^\tau, \quad (6)$$

where  $\psi_{t,\tau}(k)$  is some positive valued process. Again, we defer a formal characterization of the approximation in equation (6) to the Appendix.

Finally,  $\mathcal{T}_t$  denotes the tenors available at a given point in time  $t$ , and the vectors  $\underline{\mathbf{k}}_t = (\underline{k}(t, \tau))_{\tau \in \mathcal{T}_t}$  and  $\bar{\mathbf{k}}_t = (\bar{k}(t, \tau))_{\tau \in \mathcal{T}_t}$  indicate the lowest and highest log-moneyness at time  $t$  across the available tenors. These quantities may vary over time and be random, thus accommodating pronounced shifts in the observed option cross-section along these dimensions over the sample.

The above setting allows for a great deal of intertemporal heterogeneity in the observation scheme. For example, the tenors need not be identical across days, and we allow for a different number of options in the panel across days, maturities and moneyness. Also, intuitively, the relative number of options at a given time will impact the precision of the inference for the state vector compared to other dates. Likewise, the relative number of options across the maturities and the local ‘‘sparseness’’ of the strikes impact the quality of inference for parameters and state variables differentially depending on their sensitivity to tenor and moneyness. The quantities  $\varsigma_t$ ,  $\pi_t^\tau$  and  $\psi_{t,\tau}(k)$  capture these facets of the panel configuration, and they appear explicitly in the asymptotic distribution theory below.



We stipulate that option prices, quoted in terms of BSIV, are observed with error, i.e., we have,

$$\tilde{\kappa}_{t,k,\tau} = \kappa_{t,k,\tau} + \epsilon_{t,k,\tau}, \quad (7)$$

where we use the shorthand  $\kappa_{t,k,\tau} = \kappa(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0)$  and the observation (measurement) errors,  $\epsilon_{t,k,\tau}$ , are defined on a space  $\Omega^{(1)} = \prod_{t \in \mathbb{N}, k \in \mathbb{R}, \tau \in \Gamma} \mathcal{A}_{t,k,\tau}$ , for  $\mathcal{A}_{t,k,\tau} = \mathbb{R}$ , with  $\Gamma$  denoting the set of all possible tenors. Moreover,  $\Omega^{(1)}$  is equipped with the product Borel  $\sigma$ -field  $\mathcal{F}^{(1)}$  as well as with a transition probability  $\mathbb{P}^{(1)}(\omega^{(0)}, d\omega^{(1)})$  from the original probability space  $\Omega^{(0)}$  – on which  $X$  is defined – to  $\Omega^{(1)}$ . Finally, we define the information filtration on  $\Omega^{(1)}$  via  $\mathcal{F}_t^{(1)} = \sigma(\epsilon_{s,k,\tau} : s \leq t)$ . Then, the filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$  is given as follows,  $\Omega = \Omega^{(0)} \times \Omega^{(1)}$ ,  $\mathcal{F} = \mathcal{F}^{(0)} \times \mathcal{F}^{(1)}$ ,

$$\mathcal{F}_t = \cap_{s > t} \mathcal{F}_s^{(0)} \times \mathcal{F}_s^{(1)}, \quad \text{and} \quad \mathbb{P}(d\omega^{(0)}, d\omega^{(1)}) = \mathbb{P}^{(0)}(d\omega^{(0)})\mathbb{P}^{(1)}(\omega^{(0)}, d\omega^{(1)}).$$

We assume the extension has the property that any local martingale and semimartingale remain a local martingale and semimartingale (with the same characteristics) on the extended probability space. Formally, this is assured if the extension is “very good”.<sup>9</sup>

Intuitively, we motivate this representation by the following considerations. The option errors are defined on the space  $\Omega^{(1)}$ . We equip this space with the simple product topology as, at any point in time, only a countable number of them appear in our estimation. Since we want to allow the errors and the underlying process  $X$  to be dependent, we define the probability measure via a transition probability distribution from  $\Omega^{(0)}$  to  $\Omega^{(1)}$ . Importantly, we assume that the observation errors are  $\mathcal{F}^{(0)}$ -conditionally centered and possibly heteroskedastic of unknown form, i.e.,

$$\mathbb{E}(\epsilon_{t,k,\tau} | \mathcal{F}^{(0)}) = 0, \quad \text{and} \quad \mathbb{E}(\epsilon_{t,k,\tau}^2 | \mathcal{F}^{(0)}) = \phi_{t,k,\tau} \text{ for some } \phi_{t,k,\tau}, \quad (8)$$

and, in addition, they are  $\mathcal{F}^{(0)}$ -conditionally independent:

$$\epsilon_{t,k,\tau} \perp\!\!\!\perp \epsilon_{t',k',\tau'} | \mathcal{F}^{(0)} \quad \text{as soon as } (t, k, \tau) \neq (t', k', \tau'), \quad (9)$$

where, as is common for two vectors  $\mathbf{a}$  and  $\mathbf{b}$ ,  $\mathbf{a} \neq \mathbf{b}$  signifies that they differ in at least one of their elements. We refer to the Appendix for additional assumptions on the observation error process, but note that these are satisfied in the case of “modulated” observation noise, i.e., for  $\epsilon_{t,k,\tau} = \sqrt{\phi_{t,k,\tau}} \zeta_{t,k,\tau}$ , where  $\phi_{t,k,\tau}$  is  $\mathcal{F}_t^{(0)}$ -adapted and  $\zeta_{t,k,\tau}$  is defined on  $\Omega^{(1)}$  and is independent from  $\mathcal{F}^{(0)}$ .<sup>10</sup>

<sup>9</sup>That is,  $\omega^{(0)} \rightarrow \int 1_{\{A\}}(\omega^{(0)}, \omega^{(1)})\mathbb{P}^{(1)}(\omega^{(0)}, d\omega^{(1)})$  is  $\mathcal{F}_t^{(0)}$ -measurable for all  $A \in \mathcal{F}^{(1)}$  (see Definition II.7.1 of Jacod and Shiryaev (2003)).

<sup>10</sup>We also note that the quantities associated with the sampling scheme,  $\underline{k}$ ,  $\bar{k}$  and  $N_t$ , are allowed to be random and depend on  $\mathcal{F}^{(0)}$ . Due to the assumption  $\mathbb{E}(\epsilon_{t,k,\tau} | \mathcal{F}^{(0)}) = 0$ , this does not cause “endogeneity” biases in the estimation.

### 2.3 Volatility Estimation from High-Frequency Asset Prices

We conclude this section by introducing the second source of information, consisting of high-frequency data on the underlying asset  $X$ . These data are used only for nonparametric estimation of the volatility process, which, in turn, assist in the recovery of the state vector (or at least part of it). To this end, we assume to have an equidistant high-frequency recording of  $X$  with grid size  $\Delta_n = 1/n$ , i.e., we observe  $X$  at times  $0, 1/n, \dots, i/n, \dots, T+1$ . The log-price is denoted by  $x_t = \log(X_t)$  and, similarly, its increment by  $\Delta_i^n x = x_{i/n} - x_{(i-1)/n}$ . Using this, our nonparametric volatility estimator at time  $t$  is defined as

$$\hat{V}_t^n = \frac{n}{k_n} \sum_{i \in I_t^n} (\Delta_i^n x)^2 \mathbf{1}_{\{|\Delta_i^n x| \leq \alpha \Delta_n^\varpi\}}, \quad I_t^n = [tn + 1, \dots, tn + k_n], \quad (10)$$

for some  $\alpha > 0$ ,  $\varpi \in (0, 1/2)$ , and  $k_n \in \mathbb{N}$  with  $k_n < n$ . This constitutes the spot counterpart of the integrated truncated variation estimator of Mancini (2001). For satisfactory finite sample performance, it is critical to determine  $\alpha$  adaptively, based on a preliminary tuning-free estimate of volatility, such as the bipower variation of Barndorff-Nielsen and Shephard (2004, 2006) or the median realized volatility of Andersen et al. (2012).

The limit theory for the PLS estimator is directly impacted by the degree of information about spot volatility embedded in the high-frequency data relative to the option prices. Asymptotically, this is determined by the rate at which we acquire information from the high-frequency asset returns versus the option cross-section. We summarize this feature of the system through the assumption,

$$\frac{k_n}{N} \rightarrow \varrho \quad \text{for some real number } \varrho \geq 0.$$

Note that we allow  $\varrho = 0$ , but not  $\varrho = \infty$ . This is consistent with the empirical findings, where short-dated option prices are invariably found to be very useful in pinning down the level of volatility, while the high-frequency spot estimator can be quite noisy. Nonetheless, we do expect the high-frequency data to provide auxiliary information, corresponding to a limiting ratio  $\varrho > 0$ .

## 3 Inference for Panels with Long Time Span

This section provides our econometric analysis for the scenario where the large cross-sectional dimension of the panel is accompanied by a large time span. Given the observation equation (7), we estimate the system using least squares. However, we further penalize the deviation of the model-implied volatility from its nonparametric high-frequency counterpart during estimation, so we refer to our approach as penalized least squares. Section 3.1 introduces the formal estimation criterion and establishes consistency. We then provide the notation required for the associated limit theory in

Section 3.2, before presenting the limit distribution results in Section 3.3. Finally, Section 3.4 derives results for the estimation of the factor dynamics from the extracted factor realizations.

### 3.1 The PLS Estimator and Consistency

Our setup is in some respects similar to that of the classical dynamic nonlinear panel models, see, e.g., Hahn and Newey (2004) and Hahn and Kuersteiner (2011), with the latent factors playing the role of the fixed effects. As in this strand of work, we do not model the dynamics of the latent factors parametrically. Consequently, given our assumption for the observation error in equation (7), it is natural to estimate the parameters and the fixed effects (latent factors) via least squares. We further augment the objective function by a penalty term, which reflects information in the high-frequency volatility estimator,  $\widehat{V}_t^n$ , about the latent factors  $\mathbf{S}_t$ .<sup>11</sup>

Formally, we define the  $T \times p$  matrix of factor realizations by  $\mathbf{Z} = \{\mathbf{Z}'_t\}_{t=1}^T$  and use this to write the objective function, for some finite constant  $\lambda \geq 0$ , as,<sup>12</sup>

$$\begin{aligned} \mathcal{L}(\mathbf{Z}, \boldsymbol{\theta}) &\equiv \sum_{t=1}^T \mathcal{L}_t(\mathbf{Z}_t, \boldsymbol{\theta}), \quad \text{with} & (11) \\ \mathcal{L}_t(\mathbf{Z}_t, \boldsymbol{\theta}) &\equiv \left\{ \sum_{j=1}^{N_t} \frac{(\tilde{\kappa}_{t,k_j,\tau_j} - \kappa(k_j, \tau_j, \mathbf{Z}_t, \boldsymbol{\theta}))^2}{N_t} + \lambda \frac{k_n}{N_t} \left( \widehat{V}_t^n - \xi_1(\mathbf{Z}_t) \right)^2 \right\}. \end{aligned}$$

The first part of the objective function is the  $L_2$ -distance between observed and model-implied option prices. The second is a penalization term for the deviation of the model-implied spot volatility estimate from a direct, but noisy, measure of spot volatility obtained from high-frequency return data,  $\widehat{V}_t^n$ .<sup>13</sup> We recall that  $V_t = \xi_1(\mathbf{S}_t)$ . The use of a noisy observation of the state vector (or part of it) in the estimation procedure is reminiscent of the FAVAR approach, proposed by Bernanke et al. (2005).<sup>14</sup> Similarly to the FAVAR, the penalization in equation (11) serves to improve the identification of the latent factors, which, as we show later, may contribute to an improvement in efficiency.

To characterize the properties of our PLS estimator, it is convenient to concentrate out the factor realizations, i.e., we set,

$$\hat{\mathbf{S}}(\boldsymbol{\theta}) = \underset{\mathbf{Z} \in \mathcal{S}^T}{\operatorname{argmin}} \mathcal{L}(\mathbf{Z}, \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}, \quad \mathcal{S} \in \mathbb{R}^p, \quad (12)$$

<sup>11</sup>Note that our option pricing model does contain a parametric assumption for the evolution of  $\mathbf{S}_t$ . However, this is for its behavior under the risk-neutral measure  $\mathbb{Q}$  only. We avoid specifying the  $\mathbb{P}$  dynamics of  $\mathbf{S}_t$  parametrically in this paper (this is an interesting extension that can be pursued in future work). As discussed in Andersen et al. (2015a), this has the distinct advantage of robustness with respect to the modeling of the risk premia, i.e., the  $\mathbb{P} - \mathbb{Q}$  wedge.

<sup>12</sup>Note that, in Section 5, we will generalize the PLS estimator by allowing  $\lambda$  to be time-varying and random, which will lead to efficiency improvements.

<sup>13</sup>Recall from equation (10) that we standardized the estimator  $\widehat{V}_t^n$  by  $k_n$ . This is the reason why we multiply the penalization term in equation (11) by  $k_n$ . It is merely a matter of normalization.

<sup>14</sup>Bernanke et al. (2005) augment a VAR of economic variables with a noisy observation of a latent factor, related to the variables in the VAR system, to exploit auxiliary information regarding the factor realizations.

and the resulting estimator, then, takes the form,

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \mathcal{L}(\hat{\mathbf{S}}(\boldsymbol{\theta}), \boldsymbol{\theta}), \quad \hat{\mathbf{S}} = \hat{\mathbf{S}}(\hat{\boldsymbol{\theta}}). \quad (13)$$

In the case when  $N \rightarrow \infty$  and  $T$  is fixed, the asymptotic properties of the estimator  $(\hat{\mathbf{S}}', \hat{\boldsymbol{\theta}})$  have been derived in Andersen et al. (2015a). In this section, we consider the case where both  $N$  and  $T$  increase asymptotically. We start with the consistency result.

**Theorem 1.** *Let Assumptions 1-6 hold. Assume  $N \rightarrow \infty$ ,  $T \rightarrow \infty$  with  $T/N \rightarrow \gamma$  for some finite  $\gamma \geq 0$ ,  $\Delta_n \rightarrow 0$  and  $k_n \rightarrow \infty$  with  $\frac{k_n}{N} \rightarrow \varrho$  for some  $\varrho \geq 0$ . In addition let  $\varpi \in \left(\frac{1}{4-\tau}, \frac{1}{2}\right)$  and  $k_n \sqrt{\Delta_n} \rightarrow 0$ . Then, we have,*

$$\sup_{t=1, \dots, T} \|\hat{\mathbf{S}}_t - \mathbf{S}_t\| \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \hat{\boldsymbol{\theta}} \xrightarrow{\mathbb{P}} \boldsymbol{\theta}_0. \quad (14)$$

We note that consistency holds under very general conditions for the tuning parameters of the high-frequency volatility estimator,  $\varpi$  and  $k_n$ . Of course, these conditions are absent if the penalization term is left out of the objective function, i.e., when  $\lambda = 0$ .

### 3.2 Auxiliary Notation for the Limiting Distribution

We now turn towards the limit distribution of the parameter and factor estimates. As discussed in the introduction, the model parameters will be estimated at the rate  $\sqrt{NT}$ , but are subject to an asymptotic bias due to the incidental parameter problem, while the factor (or state vector) realizations converge at the slower  $\sqrt{N}$  rate. In addition, given the multiple and potentially interacting sources of unobserved heterogeneity, the PLS estimator clearly will not utilize the optimal weighting of the observations during estimation. As a consequence, the limiting covariance matrices in the asymptotic distribution take the common sandwich form, albeit with a number of inherent features of the system impacting the representation. To state the result in an interpretable format, we require additional notation, which we is introduced in the following.

In what follows, all partial derivatives of  $\kappa$  that appear below are assumed to exist, as implied by our assumptions in Section A.1. The first-order derivatives with respect to the parameter and state vectors are given by the  $q \times 1$  and  $p \times 1$  vectors,

$$\mathbf{D}(k, \tau, \mathbf{Z}, \boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \kappa(k, \tau, \mathbf{Z}, \boldsymbol{\theta}), \quad \mathbf{G}(k, \tau, \mathbf{Z}, \boldsymbol{\theta}) = \frac{\partial}{\partial \mathbf{Z}} \kappa(k, \tau, \mathbf{Z}, \boldsymbol{\theta}),$$

respectively, and, similarly, their first and second derivatives with respect to the state vector  $\mathbf{Z}$  are denoted by  $\mathbf{D}^{(S)}(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$ ,  $\mathbf{G}^{(S)}(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$ ,  $\mathbf{D}^{(SS)}(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$  and  $\mathbf{G}^{(SS)}(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$  (of dimensions  $q \times p$ ,  $p \times p$ ,  $q \times p^2$  and  $p \times p^2$ , respectively), and the corresponding derivatives with respect to  $\boldsymbol{\theta}$  are denoted by  $\mathbf{D}^{(\theta)}(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$  and  $\mathbf{G}^{(\theta)}(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$  (of dimension  $q \times q$  and  $p \times q$ ).

To simplify notation, for each pair  $(k_j, \tau_j)$ , we set  $\tilde{\kappa}_{j,t} \equiv \tilde{\kappa}_{t,k_j,\tau_j}$ ,  $\kappa_j(\mathbf{Z}, \boldsymbol{\theta}) \equiv \kappa(k_j, \tau_j, \mathbf{Z}, \boldsymbol{\theta})$ , and simplify the notation for the partial derivatives of the latter analogously. Furthermore, when we evaluate  $\kappa_j(\mathbf{Z}, \boldsymbol{\theta})$  at the true vector  $(\mathbf{S}'_t, \boldsymbol{\theta}'_0)$ , we denote it by  $\kappa_{j,t,0}$ , and the notation for the partial derivatives of  $\kappa_j(\mathbf{Z}, \boldsymbol{\theta})$  is simplified in the same manner, if evaluated at  $(\mathbf{S}'_t, \boldsymbol{\theta}'_0)$ . Similarly, for the penalty term,

$$\mathbf{C}(\mathbf{Z}) \equiv \frac{\partial}{\partial \mathbf{Z}} \xi_1(\mathbf{Z}) \quad \text{and} \quad \mathbf{C}^{(S)}(\mathbf{Z}) \equiv \frac{\partial^2}{\partial \mathbf{Z} \mathbf{Z}'} \xi_1(\mathbf{Z})$$

denote the  $p \times 1$  gradient vector and  $p \times p$  higher-order gradient matrix, respectively, and we will use the shorthand notation  $\mathbf{C}_{t,0} = \mathbf{C}(\mathbf{S}_t)$  and  $\mathbf{C}_{t,0}^{(S)} = \mathbf{C}^{(S)}(\mathbf{S}_t)$ .

The differential rate of convergence implies that the parameters can be treated as fixed and known, when deriving the limit distribution for the estimates of the factor realizations. As a result, the quantities determining the limiting conditional covariance matrix for the latter take a fairly conventional form. The main complication arises from the fact that the usual component stemming from the least squares criterion is supplemented by a contribution from the penalization term. The relevant matrices for the factor realizations are given in the following equations. First, define the  $p \times p$  matrix,

$$\mathcal{I}_{S,t} = \sum_{\tau} \pi_t^{\tau} \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \mathbf{G}(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0) \mathbf{G}(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0)' dk, \quad (15)$$

which represents an outer product of gradient type term for the least squares criterion. Notice that the relative number of options at each tenor ( $\pi_t^{\tau}$ ) and the relative denseness, or clustering, of options across certain segments of the strike range ( $\psi_{t,\tau}(k)$ ) enter explicitly in the expression. It is important to note that  $\mathcal{I}_{S,t}$  is simply the in-fill asymptotic limit of an average of outer products of gradients, computed using the available values of the pair  $(k, \tau)$  at time  $t$ , i.e., we have,

$$\mathcal{I}_{S,t} \approx \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{G}(k_j, \tau_j, \mathbf{S}_t, \boldsymbol{\theta}_0) \mathbf{G}(k_j, \tau_j, \mathbf{S}_t, \boldsymbol{\theta}_0)', \quad (16)$$

and analogous intuition applies to the rest of the integrals (over  $k$ ) appearing in this section.

A corresponding  $p \times p$  matrix, reflecting the outer product of the gradient for the penalization term, is given by,

$$\mathcal{P}_{S,t} \equiv \frac{\lambda \varrho}{\varsigma_t} \times \mathbf{C}_{t,0} \mathbf{C}'_{t,0}, \quad (17)$$

where the relative number of options available on that date ( $\varsigma_t$ ) and the relative informativeness of the high-frequency volatility estimator relative to the option data ( $\varrho$ ) enter the expression.

Finally, we define the  $p \times p$  asymptotic covariance matrices  $\boldsymbol{\Omega}_{S,t}$  and  $\boldsymbol{\Psi}_{S,t}$ , which further include the variance of the observation error for the option BSIV ( $\phi_{t,k,\tau}$ ), respectively, the variance of the nonparametric volatility estimator ( $2V_t^2$ ),

$$\boldsymbol{\Omega}_{S,t} = \frac{1}{\varsigma_t} \sum_{\tau} \pi_t^{\tau} \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \phi_{t,k,\tau} \mathbf{G}(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0) \mathbf{G}(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0)' dk, \quad \boldsymbol{\Psi}_{S,t} = 2 \frac{\lambda^2 \varrho}{\varsigma_t^2} V_t^2 \mathbf{C}_{t,0} \mathbf{C}'_{t,0}.$$

We turn next to the expressions required to capture the asymptotic behavior of the estimates for the model parameters,  $\boldsymbol{\theta}_0$ . They are more complex and involve interaction terms among the factors and parameters as well as the features of the system determining the magnitude of the asymptotic bias. We start by defining the  $q \times q$  information matrix for the parameter vector,

$$\mathcal{I}_{\boldsymbol{\theta},t} = \sum_{\tau} \pi_t^{\tau} \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \mathbf{D}(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0) \mathbf{D}(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0)' dk, \quad (18)$$

and the  $p \times q$  cross information matrix,

$$\mathcal{I}_{S,\boldsymbol{\theta},t} = \sum_{\tau} \pi_t^{\tau} \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \mathbf{G}(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0) \mathbf{D}(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0)' dk. \quad (19)$$

In turn, these are used to define,

$$\mathbf{R}_{t,0} \equiv -(\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathcal{I}_{S,\boldsymbol{\theta},t} \quad \text{and} \quad \mathbf{U}_t(k, \tau, \mathbf{Z}, \boldsymbol{\theta}) \equiv \mathbf{D}(k, \tau, \mathbf{Z}, \boldsymbol{\theta}) + \mathbf{R}'_{t,0} \mathbf{G}(k, \tau, \mathbf{Z}, \boldsymbol{\theta}), \quad (20)$$

the first of which is the influence function of the gradient with respect  $\boldsymbol{\theta}$ , and  $\mathbf{U}_t(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$  partials out the effect of the state variables on  $\mathbf{D}(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$ . This corresponds to demeaning the score of  $\boldsymbol{\theta}$ , in analogy to eliminating time and individual effects in large linear panel data models. As above, we denote the first and second derivative with respect to  $\mathbf{Z}$  as well as the first derivative with respect to  $\boldsymbol{\theta}$  of  $\mathbf{U}_t(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$  by  $\mathbf{U}_t^{(S)}(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$ ,  $\mathbf{U}_t^{(SS)}(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$ , and  $\mathbf{U}_t^{(\boldsymbol{\theta})}(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$ , respectively. In analogy with our simplified notation for a pair  $(k_j, \tau_j)$ , we set  $\mathbf{U}_{j,t}(\mathbf{Z}, \boldsymbol{\theta}) = \mathbf{U}_t(k_j, \tau_j, \mathbf{Z}, \boldsymbol{\theta})$ , and, when evaluated at the true vector  $(\mathbf{S}'_t, \boldsymbol{\theta}'_0)$ , we further write  $\mathbf{U}_{j,t,0}$ . Similar simplifications are made to the notation for the partial derivatives of  $\mathbf{U}_t(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$ . With this notation in hand, we define the  $q \times q$  information matrix,

$$\mathcal{I}_{U,t} = \sum_{\tau} \pi_t^{\tau} \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \mathbf{U}_t(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0) \mathbf{U}_t(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0)' dk, \quad (21)$$

the  $q \times q$  matrices  $\mathcal{P}_{\boldsymbol{\theta},t} = \mathbf{R}'_{t,0} \mathcal{P}_{S,t} \mathbf{R}_{t,0}$ ,  $\boldsymbol{\Psi}_{\boldsymbol{\theta},t} = \mathbf{R}'_{t,0} \boldsymbol{\Psi}_{S,t} \mathbf{R}_{t,0}$ , and,

$$\begin{aligned} \boldsymbol{\Omega}_{U,t} &= \frac{1}{s_t} \sum_{\tau} \pi_t^{\tau} \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \phi_{t,k,\tau} \mathbf{U}_t(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0) \mathbf{U}_t(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0)' dk, \\ \boldsymbol{\Omega}_{S,U,t} &= \frac{1}{s_t} \sum_{\tau} \pi_t^{\tau} \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \phi_{t,k,\tau} \mathbf{G}(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0) \mathbf{U}_t(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0)' dk, \end{aligned}$$

which all appear, directly or indirectly, in the limiting distributions developed subsequently. Finally, we let  $\boldsymbol{\mathcal{X}}$  be generic notation for any of the matrices  $\mathcal{I}_U$ ,  $\mathcal{P}_{\boldsymbol{\theta}}$ ,  $\boldsymbol{\Omega}_U$ , and  $\boldsymbol{\Psi}_{\boldsymbol{\theta}}$ , which we define by,

$$\boldsymbol{\mathcal{X}} = \begin{cases} \text{plim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \boldsymbol{\mathcal{X}}_t, & \text{if } T \rightarrow \infty, \\ \frac{1}{T} \sum_{t=1}^T \boldsymbol{\mathcal{X}}_t, & \text{if } T \text{ is fixed.} \end{cases} \quad (22)$$

These matrices are guaranteed to exist by our assumptions.<sup>15</sup> Note that the convergence in probability result (22) may be proven by invoking an ergodic theorem. For example, if  $\varsigma_t$ ,  $\pi_t^\top$ ,  $\psi_{t,\tau}(k)$ ,  $\underline{k}(t, \tau)$  and  $\bar{k}(t, \tau)$  are all functions of  $\mathbf{S}_t$ , do not have any other source of randomness or dependence on time, and if the process  $\mathbf{S}_t$  is stationary and ergodic with  $\mathbb{E}|\mathcal{X}_t| < \infty$ , then equation (22) holds.

### 3.3 Limiting Distribution of the PLS Estimator

The representation of the asymptotic distribution for the parameters and factor realizations involves the notion of stable convergence (denoted by  $\xrightarrow{\mathcal{L}-s}$ ). It means that the convergence holds jointly with any bounded random variable defined on the original probability space. This feature is fundamental in establishing the following limiting distribution result.

**Theorem 2.** *Let Assumptions 1-8 hold. Assume  $N \rightarrow \infty$ ,  $T \rightarrow \infty$  with  $T/N \rightarrow \gamma$  for some finite  $\gamma \geq 0$ ,  $\Delta_n \rightarrow 0$  and  $k_n \rightarrow \infty$  with  $\frac{k_n}{N} \rightarrow \varrho$  for some finite  $\varrho \geq 0$ . In addition, let,*

$$\varpi \in \left( \frac{1}{2(2-r)}, \frac{1}{2} \right), \quad k_n \sqrt{\Delta_n} \rightarrow 0, \quad \sqrt{T k_n} \left( k_n \Delta_n \vee \Delta_n^{(2-r)\varpi} \right) \rightarrow 0, \quad \frac{\Delta_n^{(2-\frac{r}{2})\varpi-1}}{k_n} \rightarrow 0. \quad (23)$$

Then, we have,

$$\sqrt{TN} \left( \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 - \mathbf{B}_{T,N} \right) \xrightarrow{\mathcal{L}-s} (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} (\boldsymbol{\Omega}_U + \boldsymbol{\Psi}_\theta)^{1/2} \times \mathbf{E}_\theta, \quad (24)$$

where  $\mathbf{E}_\theta$  is a  $q \times 1$  standard Gaussian vector, defined on an extension of the original probability space and independent from  $\mathcal{F}$ , and the asymptotic bias is given by,

$$\mathbf{B}_{T,N} = (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \times \sum_{j=1}^2 \left( \mathbf{B}_{T,N,1}^{(j)} - \mathbf{B}_{T,N,2}^{(j)} - \mathbf{B}_{T,N,3}^{(j)} \right),$$

whose components are defined in Section A.3. In addition, for arbitrary  $t_1, \dots, t_h \in \mathbb{N}$ , we have,

$$\sqrt{N} \begin{pmatrix} \hat{\mathbf{S}}_{t_1} - \mathbf{S}_{t_1} \\ \vdots \\ \hat{\mathbf{S}}_{t_h} - \mathbf{S}_{t_h} \end{pmatrix} \xrightarrow{\mathcal{L}-s} \text{diag} \begin{pmatrix} (\mathcal{I}_{S,t_1} + \mathcal{P}_{S,t_1})^{-1} (\boldsymbol{\Omega}_{S,t_1} + \boldsymbol{\Psi}_{S,t_1})^{1/2} \\ \vdots \\ (\mathcal{I}_{S,t_h} + \mathcal{P}_{S,t_h})^{-1} (\boldsymbol{\Omega}_{S,t_h} + \boldsymbol{\Psi}_{S,t_h})^{1/2} \end{pmatrix} \times \begin{pmatrix} \mathbf{E}_{\mathbf{S}_{t_1}} \\ \vdots \\ \mathbf{E}_{\mathbf{S}_{t_h}} \end{pmatrix}, \quad (25)$$

where  $\mathbf{E}_{\mathbf{S}_{t_1}}, \dots, \mathbf{E}_{\mathbf{S}_{t_h}}$  is a sequence of standard Gaussian vectors, defined on an extension of the original probability space, with each of them independent of each other, of  $\mathbf{E}_\theta$ , as well as of  $\mathcal{F}$ , and the convergence in equation (25) holds jointly with that in equation (24).

<sup>15</sup>Typical option data sets contain deterministic cycles in the configuration of the available tenors. This feature renders quantities such as  $\mathcal{I}_{U,t}$  non-stationary. However, considering  $\sum_{t \in \mathcal{T}_m} \mathcal{I}_{U,t}$ , for  $\mathcal{T}_m$  denoting a period covering one calendar month, such non-stationarities vanish, and standard laws of large numbers suffice for validation of equation (22).

This limit result deserves a detailed discussion. First, the rate of convergence of the parameter vector is  $\sqrt{NT}$ , so both the increase in the cross-section and time dimension of the panel help in the recovery of the parameters. By contrast, the factor realizations are estimated at the rate  $\sqrt{N}$ . This is quite intuitive as increasing the time span of the panel does not provide extra information for the factor realization at a given point in time.<sup>16</sup> The asymptotic limit distributions of the parameters and factor realizations are fundamentally different. For the former, it is Gaussian with a constant limiting variance, as is common in econometric applications. On the other hand, it is mixed Gaussian for the estimates of the factor realizations, i.e., it is Gaussian only when conditioned on the filtration of the original probability space, and the covariance matrix depends explicitly on the latter. That is,  $\mathcal{I}_{S,t} + \mathcal{P}_{S,t}$  and  $\Omega_{S,t} + \Psi_{S,t}$ , which appear in the limiting distribution of the factor realizations in equation (25), are stochastic. This implies that the precision in recovering the factor realizations is time-varying and depends on unobserved random quantities associated with the cross-section of option data at that particular point in time. For example, periods of elevated volatility may be associated with a larger variance of the observation errors (captured by the function  $\phi$ ), and hence noisier recovery for the factor realizations.

We can contrast the limiting behavior above with the case of a panel with fixed time span, but asymptotically increasing cross-section. In this scenario, both the parameter vector and factor realizations have mixed Gaussian limiting distributions (see also Lemma 1 below). Furthermore, unlike here, the parameter and state vector estimates display asymptotic dependence in the fixed time span case. This dependence becomes asymptotically negligible when the time span is long, since the parameter vector is now recovered at a faster rate.

We note that, given the non-standard mixed Gaussian distribution of the factors, we need a stronger notion of convergence in equation (25) – stable convergence – than what is typically derived in econometric applications, in order to specify the joint asymptotic behavior of the parameters and factors. The stable convergence is also required to accommodate the penalization with the high-frequency volatility estimator and characterize jointly the asymptotic behavior of the option error and the error coming from the high-frequency volatility estimation. To the best of our knowledge, the joint asymptotic behavior of the factor and parameter estimates in the nontrivial case, where the  $\mathcal{F}$ -conditional limiting variance of the factors is stochastic, has not been derived in earlier work on large panels, with both  $N$  and  $T$  diverging. Such results are indispensable for deriving the properties of test statistics for joint model fit across a time period of interest, involving jointly the estimates for the factors, parameters and high-frequency volatility.<sup>17</sup>

<sup>16</sup>We note that this could occur if there were pathwise restrictions on the dynamic evolution of the factors. However, existing dynamic factor models do not feature such restrictions.

<sup>17</sup>We refer to Jacod and Shiryaev (2003) for additional details regarding the notion of stable convergence. In the case where



A further distinguishing feature of the long versus fixed span case is the presence of an asymptotic bias in the parameter estimation, the explicit form of which is given in the Appendix. This bias reflects the effect of uncertainty in the recovery of the factor realizations on the parameter estimation. Intuitively, although the errors in the estimation of the factor realizations average out over time, and thus do not slow down the recovery rate of the parameters, the nonlinearity of the model, both in parameters and the state vector, induces an asymptotic bias of order  $1/N$  in the parameter estimates. Therefore, if  $T \ll N$  ( $\gamma = 0$  in Theorem 2), this bias does not affect the central limit theorem for the parameter vector, i.e., the bias term in the limit result (24) is of higher order. On the other hand, if  $T \asymp N$ , the asymptotic bias does have a non-negligible effect on the limit distribution of the parameter estimate. The form of the bias terms resembles those in the classical panel data literature, see e.g., Hahn and Newey (2004) and Hahn and Kuersteiner (2011), with the difference here arising from the stochastic nature of the terms, the infill asymptotic setting in the cross-section as well as the penalty term in the estimation due to the factor signal from the high-frequency data. Finally, we remark that the limit results in equations (24) and (25) are infeasible because the asymptotic variance and bias term are based on population quantities. In the next section, we show that they are straightforward to estimate consistently from the data, thus enabling feasible inference.

In terms of the asymptotic role of the penalization, we have two cases. When  $k_n/N \rightarrow 0$ , the penalization has no first-order asymptotic effect, while for  $k_n/N \rightarrow \varrho > 0$ , the penalization impacts the asymptotic distribution of the estimators for both the parameters and factor realizations. The penalty condition provides a restriction on the factor realization, namely that the model-implied estimate of volatility  $\xi_1(\mathbf{S}_t)$  should be close to the nonparametric estimate of volatility  $V_t$ . The option panel allows us to recover the factor realizations at the rate  $\sqrt{N}$ , as implied by equation (25). On the other hand, the rate of convergence of the high-frequency volatility estimate is  $\sqrt{k_n}$  (recall,  $k_n$  is the local window for the high-frequency data used to construct the volatility estimator). Therefore, when  $k_n/N \rightarrow 0$ , the option data are more efficient in recovering the volatility, rendering penalization during estimation suboptimal asymptotically.<sup>18</sup> On the other hand, when  $k_n/N \rightarrow \varrho > 0$ , the nonparametric high-frequency volatility estimate and the option-based are equally efficient, in terms of rate of convergence, and hence the penalization may have an asymptotic effect. This explains why, in our objective function, the penalty term is multiplied by  $\lambda \frac{k_n}{N_t}$ . This ensures that the penalty

---

the penalization term does not have an impact on the limiting distribution, then we can strengthen the convergence results of Theorem 2 to almost sure convergence, conditional on the filtration generated by the latent factor  $\mathbf{S}_t$ . When the penalization term matters asymptotically, however, we cannot prove this stronger mode of convergence as, in general, we do not have control over the resulting conditional distribution (for example,  $\mathbf{X}$  and  $\mathbf{S}$  can be strongly dependent, accommodating the so-called “leverage effect” in asset prices). The stable convergence result in Theorem 2 is sufficient for deriving the feasible limiting theory below.

<sup>18</sup>Even though it is suboptimal (asymptotically) to use penalization within this scenario, its presence may greatly facilitate the identification of the state vector in finite samples.

becomes asymptotically negligible exactly when it is suboptimal to utilize high-frequency return data for estimation. Importantly, the user does not need to take a stand on which data source, options or high-frequency data, is more efficient for recovering spot volatility. Our PLS procedure automatically adapts to the scenario at hand. Nevertheless, the user still needs to pick the constant  $\lambda$ . In Section 5, we propose a data-dependent choice for this parameter, which boosts efficiency of the inference.

Above, we rule out the possibility that  $k_n/N \rightarrow \infty$ , i.e., that the high-frequency data are more efficient than the options (in terms of rate of convergence) in recovering spot volatility.<sup>19</sup> This case would heighten the asymptotic role of the penalization term. In fact, we have distinct scenarios depending on the dimension of the state vector and the role of its components in determining spot volatility. For example, in the one-factor stochastic volatility model of Duffie et al. (2000), the state vector equals the level of (spot) volatility. Hence, when  $k_n/N \rightarrow \infty$ , the rate of convergence of the state vector will be the faster  $\sqrt{k_n}$  in lieu of the  $\sqrt{N}$  implied by our limit result (25), and the asymptotic bias for the parameter vector in equation (24) vanishes. In a more general parametric model with multiple factors, one of which is spot volatility, all components of the state vector, except for spot volatility, will be estimated at the rate  $\sqrt{N}$ , while the stochastic volatility will be estimated at the faster  $\sqrt{k_n}$  rate. Since  $k_n/N \rightarrow \infty$  is implausible on empirical grounds, at least for assets with liquid option markets, we do not pursue this case any further.

The relative growth conditions (23) are necessary for our limiting result, although they are required only if the penalty term is present in the objective function, i.e., if  $\lambda > 0$ . In assessing the implications of these conditions, the following points are useful. First, to avoid unnecessary truncation, it is optimal to follow the convention in the literature and set  $\varpi$  very close to  $1/2$ . Second, the higher the  $k_n$ , the faster the rate of convergence of the high-frequency estimator. Therefore, we prefer the largest value of  $k_n$  for which condition (23) is satisfied. In the typical case of finite activity jumps, i.e., when  $r = 0$  (the value of  $r$  is typically known from the specification of the risk-neutral model), and provided  $\varpi$  is set arbitrarily close to  $1/2$ , the binding constraints in condition (23) are  $k_n\sqrt{\Delta_n} \rightarrow 0$  and  $\sqrt{Tk_n}k_n\Delta_n \rightarrow 0$ . Hence, the optimal feasible choice of  $k_n$  depends on the relative size of  $T$  and  $1/\Delta_n$ . If  $T$  is not very large, we can set  $k_n \asymp 1/\Delta_n^{1/2-\iota}$  for some arbitrary small  $\iota > 0$ , guaranteeing the optimal feasible rate for estimating spot volatility from the high-frequency data. On the other hand, if  $T$  is of the same order as  $1/\Delta_n$ , the optimal feasible rate is  $k_n \asymp 1/\Delta_n^{1/3-\iota}$ . From a practical perspective, this means that for panels with a large time span relative to the sampling frequency for the returns, we need to pick a smaller sized local window for volatility estimation, slowing its rate of convergence, but reducing its bias. Given the discussion in the previous paragraph, it implies that the high-frequency volatility

<sup>19</sup>The “limiting” case of  $k_n/N \rightarrow \infty$  implies that part of the state vector is observed by the econometrician. This scenario can be of interest in other applications of factor augmentation than the volatility measurement case considered here.

estimator becomes comparatively noisy ( $k_n/N_t$  is small), then the penalty term may have only an asymptotically vanishing impact on our estimation.

Finally, in cases where the observation errors become negligible, we may have  $\mathbf{\Omega}_U$  as well as  $\mathbf{\Omega}_{S,t}$  vanish for some or all  $t$ . Some of our limit results would then become degenerate. In fact, in the extreme case when options are observed with no error, we obtain exact error-free recovery for the parameter vector and the sequence of factor realizations. Given the large percentage bid-ask spreads observed on option markets, we deem such scenarios empirically implausible.

### 3.4 On the Estimation of the Factor Dynamics

The limit result (25) allows us to quantify the precision in recovering the factor realizations, which is useful in a number of contexts. For example, as discussed in Andersen et al. (2015a), asset pricing models impose pathwise restrictions on the relation between derivative securities and the underlying asset, and our limit result enables us to test these formally. Another important use of the extracted factors is for the estimation of the underlying factor dynamics. However, for that application, the limit result (25) is insufficient since we need to analyze the impact of the errors on the recovery of the factors over a long time span. The following theorem provides the requisite result.

**Theorem 3.** *Let Assumptions 1-8 hold. Assume  $N \rightarrow \infty$ ,  $T \rightarrow \infty$  with  $T/N \rightarrow \gamma$  for some finite  $\gamma \geq 0$ ,  $\Delta_n \rightarrow 0$  and  $k_n \rightarrow \infty$  with  $\frac{k_n}{N} \rightarrow \varrho$  for some finite  $\varrho \geq 0$  as well as condition (23) of Theorem 2. Then for any twice differentiable function  $g : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}^d$  with bounded first and second derivatives, we have*

$$\frac{1}{\sqrt{T}} \sum_{t=2}^T \left( g(\hat{\mathbf{S}}_t, \hat{\mathbf{S}}_{t-1}) - g(\mathbf{S}_t, \mathbf{S}_{t-1}) \right) = o_p(1). \quad (26)$$

We note from Theorem 2 that  $\hat{\mathbf{S}}_t - \mathbf{S}_t = O_p(1/\sqrt{N})$ . In spite of that, the error in recovering the factor realizations on the statistic in equation (26) is only  $o_p(1)$ , even in the case where  $T$  and  $N$  increase at the same rate. The reason is that the error in recovering the factor realizations on different days across the sample effectively gets averaged out in the time series mean in equation (26), thus mitigating its impact on the statistic sufficiently to establish the limiting result.

Under standard stationarity conditions for  $\mathbf{S}_t$  and further assumptions guaranteeing a central limit theorem (e.g., mixing conditions),  $\frac{1}{\sqrt{T}} \sum_{t=2}^T (g(\mathbf{S}_t, \mathbf{S}_{t-1}) - \mathbb{E}(g(\mathbf{S}_t, \mathbf{S}_{t-1})))$  will converge to a normal random variable. Theorem 3 ensures that the recovery of the state vector from the option panel (and high-frequency asset returns) has no asymptotic effect on the long span limit result, i.e., it is valid even after replacing the state vector realizations by their estimates.

If the dynamics for the state vector takes a parametric form and  $\mathbf{S}_t$  is a Markov process under the physical probability measure, an attractive choice of  $g$ , capable of generating asymptotically efficient

estimates for the parameters of the state vector dynamics (with a  $\sqrt{T}$  rate of convergence), is based on the conditional characteristic function, namely  $g(\mathbf{S}_t, \mathbf{S}_{t-1}) = e^{i\mathbf{u}'\mathbf{S}_t} - \mathbb{E}_{t-1}(e^{i\mathbf{u}'\mathbf{S}_t})$ .

The use of the extracted factors for estimation of the state vector dynamics, using the above theorem, is reminiscent of the implied state inference of Pan (2002) and Pastorello et al. (2003). These papers employ a finite set of options to invert for the state vector, and then conduct inference based on the option-implied factor realizations using method of moments or maximum likelihood techniques. The major difference between these procedures and our approach is that we allow for observation errors in the option prices, while letting the number of options increase asymptotically. Because of this difference in the asymptotic setup, we find that, in contrast to Pan (2002) and Pastorello et al. (2003), the estimation of the risk-neutral parameters has no first order asymptotic effect on the state vector extraction - the risk-neutral parameters are estimated at the rate  $\sqrt{NT}$ , while the parameters of the  $\mathbb{P}$  dynamics of  $\mathbf{S}_t$  are estimated at the slower rate of  $\sqrt{T}$ .

Second, in an asymptotic setup with  $T \rightarrow \infty$  and  $T/N \rightarrow 0$  (and without using noisy information for the factor realizations), Gagliardini and Gourieroux (2014) model the dynamics of the state vector parametrically. They perform joint estimation of the parameters for the nonlinear factor model ( $\boldsymbol{\theta}$  in our setting) and for those driving the dynamics of the state vector (for which we make no parametric assumption) via maximum likelihood, and they establish asymptotic independence between the two sets of parameters (and rates of convergence of  $\sqrt{NT}$  and  $\sqrt{T}$ , respectively). However, their approach and associated limit theory exclude the challenging case of  $T \asymp N$ , which can be the pertinent asymptotic scheme for many practical applications.

Finally, Theorem 3 is related to results in the large panel data literature. In particular, Fernandez-Val and Lee (2013) derive the asymptotic distribution of functions of the parameters, data and fixed effects in large panels with random coefficients, c.f. their Corollary 1. As explained above, the fixed effects in classic panel settings can be viewed as counterparts to factor realizations in our dynamic nonlinear factor model, implying that the roles of  $N$  and  $T$  are reversed in the two asymptotic schemes.<sup>20</sup> Unlike Fernandez-Val and Lee (2013), however, we do not specify the dynamics of the factors in the model (again, fixed effects in their setting), and we only establish asymptotic negligibility of the difference  $\frac{1}{\sqrt{T}} \sum_{t=2}^T (g(\hat{\mathbf{S}}_t, \hat{\mathbf{S}}_{t-1}) - g(\mathbf{S}_t, \mathbf{S}_{t-1}))$  instead of deriving the limit distribution for statistics of the form  $\frac{1}{\sqrt{T}} \sum_{t=2}^T (g(\hat{\mathbf{S}}_t, \hat{\mathbf{S}}_{t-1}) - \mathbb{E}[g(\mathbf{S}_t, \mathbf{S}_{t-1})])$ .<sup>21</sup> Importantly, and in contrast with their framework, the present setting allows the factors to display time-series dependence.

<sup>20</sup>To see this, note that the asymptotic bias of the fixed model parameters in our setup is of order  $O_p(1/N)$ , while the corresponding bias in Fernandez-Val and Lee (2013) is of order  $O_p(1/T)$ .

<sup>21</sup>Note that the bias correction term in Corollary 1 of Fernandez-Val and Lee (2013) is negligible when the rate of convergence of the statistic is  $\sqrt{N}$  (which corresponds to  $\sqrt{T}$  in our setup).

## 4 Unified Feasible Inference with Short and Long Time Span

This section proceeds by developing feasible inference techniques for the parameter vector and the factor realizations, which enable us to construct empirical confidence intervals and conduct testing. Our inference covers panels with a fixed or an increasing time span.

### 4.1 The Unified Inference Results

We begin with a lemma that restates the limit distribution of the PLS estimator for the fixed time span, originally derived in Andersen et al. (2015a), in a manner that facilitates comparison with our limit results for the case of a large time span.

**Lemma 1.** *Under the conditions of Theorem 2, but with fixed time span, i.e.,  $T$  fixed, then,*

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}^{-s}} \frac{1}{\sqrt{T}}(\mathcal{I}_U + \mathcal{P}_\theta)^{-1}(\boldsymbol{\Omega}_U + \boldsymbol{\Psi}_\theta)^{1/2} \times \mathbf{E}_\theta,$$

where  $\mathbf{E}_\theta$  is a  $q \times 1$  vector of standard Gaussian variables, independent of each other and of  $\mathcal{F}$ . Furthermore, we have,

$$\sqrt{N} \begin{pmatrix} \hat{\mathbf{S}}_1 - \mathbf{S}_1 \\ \vdots \\ \hat{\mathbf{S}}_T - \mathbf{S}_T \end{pmatrix} \xrightarrow{\mathcal{L}^{-s}} (\mathbf{AVAR}_S)^{1/2} \times \begin{pmatrix} \mathbf{E}_{S_1} \\ \vdots \\ \mathbf{E}_{S_T} \end{pmatrix}, \quad \mathbf{AVAR}_S = \text{diag} \begin{pmatrix} \mathbf{AVAR}_{S_1} \\ \vdots \\ \mathbf{AVAR}_{S_T} \end{pmatrix} + \mathbf{ACOV}_S,$$

where  $\mathbf{E}_{S_t}$ ,  $t = 1, \dots, T$  is a sequence of standard Gaussian vectors, defined on an extension of the original probability space, independent of  $\mathcal{F}$ , but dependent on each other and on  $\mathbf{E}_\theta$ . Moreover, the diagonal elements of asymptotic covariance matrix decomposes as  $\mathbf{AVAR}_{S_t} = \sum_{i=1}^4 \mathbf{AVAR}_{S_t,i}$  where,

$$\mathbf{AVAR}_{S_t,1} = (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1}(\boldsymbol{\Omega}_{S,t} + \boldsymbol{\Psi}_{S,t})(\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1},$$

$$\mathbf{AVAR}_{S_t,2} = \mathbf{R}_{t,0} \mathbf{AVAR}_\theta \mathbf{R}'_{t,0}/T, \quad \mathbf{AVAR}_\theta = (\mathcal{I}_U + \mathcal{P}_\theta)^{-1}(\boldsymbol{\Omega}_U + \boldsymbol{\Psi}_\theta)(\mathcal{I}_U + \mathcal{P}_\theta)^{-1}$$

$$\mathbf{AVAR}_{S_t,3} = (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1}(\boldsymbol{\Omega}_{S,U,t} + \boldsymbol{\Psi}_{S,t} \mathbf{R}_{t,0}) (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \mathbf{R}'_{t,0}/T, \quad \mathbf{AVAR}_{S_t,4} = \mathbf{AVAR}'_{S_t,3}.$$

Finally, the matrix containing the off-diagonal elements is defined as,

$$\mathbf{ACOV}_S = \begin{pmatrix} \mathbf{0}_{p \times p} & \mathbf{R}_{1,0} \mathbf{AVAR}_\theta \mathbf{R}'_{2,0}/T & \dots & \mathbf{R}_{1,0} \mathbf{AVAR}_\theta \mathbf{R}'_{T,0}/T \\ \mathbf{R}_{2,0} \mathbf{AVAR}_\theta \mathbf{R}'_{1,0}/T & \mathbf{0}_{p \times p} & & \vdots \\ \vdots & & \ddots & \\ \mathbf{R}_{T,0} \mathbf{AVAR}_\theta \mathbf{R}'_{1,0}/T & \dots & & \mathbf{0}_{p \times p} \end{pmatrix}.$$

We note that the quantities  $\mathcal{I}_U$ ,  $\mathcal{P}_\theta$ ,  $\boldsymbol{\Omega}_U$  and  $\boldsymbol{\Psi}_\theta$  are random, as  $T$  is fixed (recall equation (22)). Therefore, the limit distribution of the parameter vector is now mixed Gaussian and has no asymptotic

bias. Note also that the expression for the asymptotic variance of the parameter vector looks identical to the one in equation (24). However, when  $T$  is fixed, the asymptotic variance is a random quantity. Only in the asymptotic ( $T \rightarrow \infty$ ) limit will the latter coincide with the one for the long time span. On the other hand, the limiting variance of the factor realizations differs from the long span asymptotic case. In particular, the  $\mathcal{F}$ -conditional limiting variance in equation (25) corresponds to  $\mathbf{AVAR}_{S_t,1}$  in Lemma 1 above, while the additional terms  $\{\mathbf{AVAR}_{S_t,j}\}_{j=2,3,4}$  reflect the asymptotic role of the parameter estimation on the factor extraction. Note that  $\{\mathbf{AVAR}_{S_t,j}\}_{j=2,3,4}$  are all of order  $O(1/T)$  and, as a result, they become asymptotically negligible when  $T$  increases asymptotically, as detailed in the previous section. Likewise,  $\mathbf{ACOV}_S$  in Lemma 1 reflects the asymptotic dependence in the recovery of the factor realizations at different points in time. This dependence is due to parameter uncertainty and, since it is of the order  $O(1/T)$ , negligible in the long time span scenario.

Despite the distinct differences in the limiting behavior of the PLS estimator, we are able to devise a feasible central limit theory that covers both the fixed and long time span setting. The formal result is provided by the following theorem, where  $\bar{N} = \frac{1}{T} \sum_{t=1}^T N_t$  denotes the average number of options per cross-section.<sup>22</sup>

**Theorem 4.** *Under the conditions of Theorem 2 but with  $T$  either fixed or  $T \rightarrow \infty$  with  $T/N \rightarrow \gamma$  for some finite  $\gamma \geq 0$ ,*

$$\sqrt{T\bar{N}} \left( (\hat{\mathcal{I}}_U + \hat{\mathcal{P}}_\theta)^{-1} (\hat{\Omega}_U + \hat{\Psi}_\theta) (\hat{\mathcal{I}}_U + \hat{\mathcal{P}}_\theta)^{-1} \right)^{-1/2} (\hat{\theta} - \theta_0 - \hat{\mathcal{B}}_{T,N}) \xrightarrow{\mathcal{L}} \mathbf{E}_\theta,$$

where  $\mathbf{E}_\theta$  is a  $q \times 1$  vector of standard Gaussian variables, independent of each other, and the estimators  $\hat{\mathcal{I}}_U$ ,  $\hat{\mathcal{P}}_\theta$ ,  $\hat{\Omega}_U$ ,  $\hat{\Psi}_\theta$  and  $\hat{\mathcal{B}}_{T,N}$  are defined in Section A.5. Furthermore, for the state vector,

$$\sqrt{\bar{N}} \left( \widehat{\mathbf{AVAR}}_{S_t} \right)^{-1/2} (\hat{S}_t - S_t) \xrightarrow{\mathcal{L}} \mathbf{E}_{S_t},$$

where  $\mathbf{E}_{S_t}$  is a sequence of  $p \times 1$  vectors of standard Gaussian variables. The estimator of the asymptotic covariance matrix,  $\widehat{\mathbf{AVAR}}_{S_t}$  is, similarly, deferred to the Appendix.

## 4.2 Discussion of the Unified Inference Result

Unlike Theorem 2 and Lemma 1, the limiting variables in the convergence results of Theorem 4 are standard normal vectors. This is because we utilize consistent estimators for the asymptotic variance and bias terms in Theorem 4. Based on this result, inference is feasible regardless of whether the time span is fixed or asymptotically increasing. This is important in practice, as the user need not

<sup>22</sup>Thus,  $\bar{N}$  can be viewed as the feasible counterpart to  $N$ , recall equation (5). We note that one does not need  $\bar{N}$  for feasible implementation of the limit results of Theorem 4. This can be easily checked by inspection of the formulas for the asymptotic biases and variances in the Appendix.

take a stand on whether the time span of the panel renders Lemma 1 or Theorem 2 more suitable for the case at hand. In designing the feasible inference theory, we accommodate features of the limiting distributions of either asymptotic scheme. Several key properties of the limit results for the fixed and increasing time span case are critical in this regard. First, the stable convergence for the factor realizations and parameter vector in the fixed span case enables us to apply a Slutsky type result, even if the probability limits of quantities such as  $\hat{\mathcal{I}}_U$  may be random. Second, the estimates for the asymptotic variance of the parameter vector in the long span case and the  $\mathcal{F}$ -conditional asymptotic variance in the fixed time span case are identical. Third, the estimate of the asymptotic bias is asymptotically negligible for a fixed time span or when the latter increases at a rate slower than that of the cross-sectional dimension of the panel. That is,  $\hat{\mathcal{B}}_{T,N} = O_p(1/N)$ , and thus becomes negligible whenever  $T/N \rightarrow 0$ . Finally, the estimate for the asymptotic dependence between the estimates of the parameter vector and factor realizations is of higher asymptotic order for the long time span case.

We finish this section by discussing the dependence structure imposed on the observation errors and its role in deriving the asymptotic results. In equation (9), we assume  $\mathcal{F}^{(0)}$ -conditional independence of the errors, both in the cross-section and over time. Nevertheless, the errors may be dependent through the information in  $\mathcal{F}^{(0)}$ . Moreover, the (conditional) independence of the errors in equation (9) is much stronger than what we actually need. For example, our results, and in particular Theorem 4, will continue to hold if the observation errors are of the following form,

$$\epsilon_{t,k,\tau} = \sqrt{\phi_{t,k,\tau}} \tilde{\epsilon}_t z_{t,k,\tau}, \quad (27)$$

where  $\phi_{t,k,\tau}$  are  $\mathcal{F}^{(0)}$ -adapted nonnegative random variables, and the sequences  $\{\tilde{\epsilon}_t\}_t$  and  $\{z_{t,k,\tau}\}_{(t,k,\tau)}$  are defined on  $\mathcal{F}^{(1)}$ , and are i.i.d., independent of each other as well as of  $\mathcal{F}^{(0)}$ . In the above specification, and due to the inclusion of  $\tilde{\epsilon}_t$ , the errors  $\epsilon_{t,k,\tau}$  are no longer  $\mathcal{F}^{(0)}$ -conditionally independent in the cross-section. More generally, our results will continue to hold if we replace  $\mathcal{F}^{(0)}$ -conditional independence of the errors in (9) with the following condition on (cross-)products of errors:

$$\begin{aligned} \mathbb{E} \left( \epsilon_{t_1,k_1,\tau_1} \epsilon_{t_2,k_2,\tau_2} \middle| \mathcal{F}^{(0)} \right) &= \mathbb{E} \left( \epsilon_{t_1,k_1,\tau_1}^3 \epsilon_{t_2,k_2,\tau_2} \middle| \mathcal{F}^{(0)} \right) = \mathbb{E} \left( \epsilon_{t_1,k_1,\tau_1}^2 \epsilon_{t_2,k_2,\tau_2} \epsilon_{t_3,k_3,\tau_3} \middle| \mathcal{F}^{(0)} \right) \\ &= \mathbb{E} \left( \epsilon_{t_1,k_1,\tau_1} \epsilon_{t_2,k_2,\tau_2} \epsilon_{t_3,k_3,\tau_3} \epsilon_{t_4,k_4,\tau_4} \middle| \mathcal{F}^{(0)} \right) = 0, \end{aligned} \quad (28)$$

where  $(t_i, k_i, \tau_i)_{i=1}^4$  differ from each other. Condition (28) replaces  $\mathcal{F}^{(0)}$ -conditional independence of the errors with a weaker one imposing  $\mathcal{F}^{(0)}$ -conditional uncorrelatedness of the errors and certain products of them. In particular, this allows for dependence in higher-order  $\mathcal{F}^{(0)}$ -conditional moments of the observation errors, exactly as in the example (27).

Our analysis can be further extended to cover situations in which there are  $\mathcal{F}^{(0)}$ -conditional dependence in the first moments of the errors. To accommodate such extensions, however, we will need to

modify the estimates of the asymptotic variances and biases in Theorem 4 in ways that will account for this dependence. Specifically, we can extend the analysis to cover situations where the errors display time series dependence, e.g., by adopting an error specification as in equation (27), but with  $\tilde{\epsilon}_t$  being allowed to fluctuate systematically over time. In this case, dependence in the limiting distribution for the estimates of the factor realizations will arise, even in the case where the time span increases to infinity. Such time series dependence in the errors may be accounted for by changing our estimators of the asymptotic variances (and of the bias terms) to long-run variance type estimators.

We can also extend our analysis to handle cross-sectional dependence in the observation errors. For this to be feasible, we need both convergence in probability of, and a CLT for, cross-sectional averages of the form  $\frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} m_{j,t}$ , where  $m_{j,t}$  is some  $\mathcal{F}^{(0)}$ -adapted random variable. This implies that we cannot have very strong dependence in the cross-section of the observation errors, and we certainly cannot accommodate scenarios in which the observations in a certain strike domain are perfectly dependent.<sup>23</sup> However, this case is akin to model misspecification and, arguably, such “errors” should, in fact, be part of the option pricing model. On the other hand, we can extend the analysis to handle weak cross-sectional dependence in the errors. For example, we can accommodate a specification for the error of the form  $\epsilon_{t,k,\tau} = \sqrt{\phi_{t,k,\tau}} z_{t,k,\tau}$ , where  $\phi_{t,k,\tau}$  are  $\mathcal{F}^{(0)}$ -adapted nonnegative random variables and for each  $t \in \mathbb{N}$ , the sequences  $\{z_{j,t}\}_{j=1,\dots,N_t}$  are drawn from a stationary mixing process, independently across time. As with time series dependence in the errors, we need to adjust the estimates of the asymptotic variances and biases to ones of long-run types in order to account for the cross-sectional dependence in the errors. We note that this asymptotic setup implies that the dependence between observation errors for fixed strikes and tenors vanishes asymptotically. Of course, in practice, we work with a given panel of option observations and the limit theory is derived simply to proxy the distributional properties of the estimators based on this option panel.

## 5 Weighted Penalized Least Squares

We now propose a weighted counterpart to the PLS estimator, which provides efficiency gains and enables us to choose suitable values for  $\lambda$ , period-by-period (recall equation (11)). In analogy to the reasoning for the classical weighted least squares estimator, we enhance efficiency by weighting the options according to the variance of their observation error. In our case, this implies weighting each option price by an estimate of the  $\mathcal{F}$ -conditional variance  $\phi_{t,k,\tau}$ . While this is feasible, we simplify by assigning all options on a given day the identical weight. This may result in loss of efficiency due to potential heteroskedasticity in the strike and tenor dimension of the option error variance at a given

---

<sup>23</sup>By contrast, since identification and inference for the parameters and the factor realization on a given day can be achieved from a single cross-section of options, we can handle situations with perfect time-series dependence in the errors.



point in time. Nevertheless, it still provides non-trivial efficiency improvements due to the pronounced heteroskedasticity of the  $\mathcal{F}$ -conditional option error variance over time.

Similarly, for efficiency improvements, we should weight the penalization term period-by-period. Unlike the option observation error, whose  $\mathcal{F}$ -conditional variance may be estimated by a preliminary PLS estimator, the error of the nonparametric volatility estimator has an unknown  $\mathcal{F}$ -conditional variance, which cannot be estimated without further assumptions. However, we can instead weight the penalization by an estimate for the  $\mathcal{F}$ -conditional asymptotic variance of the nonparametric volatility estimator. The latter is given by  $2V_t^2$ , and, as a result, it is trivial to construct a consistent estimator for it. Intuitively, since the asymptotic distribution of the nonparametric variance estimator is  $\mathcal{F}$ -conditionally Gaussian with a  $\mathcal{F}$ -conditional variance of  $2V_t^2$ , this is asymptotically equivalent to weighting with the  $\mathcal{F}$ -conditional variance of the variance estimator.<sup>24</sup>

More formally, we define  $\mathcal{L}^w(\mathbf{Z}, \boldsymbol{\theta}) \equiv \sum_{t=1}^T \mathcal{L}_t^w(\mathbf{Z}_t, \boldsymbol{\theta})$  with

$$\mathcal{L}_t^w(\mathbf{Z}_t, \boldsymbol{\theta}) \equiv \left\{ \frac{1}{N_t} \sum_{j=1}^{N_t} \frac{(\tilde{\kappa}_{t,k_j,\tau_j} - \kappa(k_j, \tau_j, \mathbf{Z}_t, \boldsymbol{\theta}))^2}{w(\hat{\phi}_t)} + \frac{k_n}{N_t} \frac{(\hat{V}_t^n - \xi_1(\mathbf{Z}_t))^2}{w(2(\hat{V}_t^n)^2)} \right\}, \quad (29)$$

where

$$\hat{\phi}_t = \frac{1}{N_t} \sum_{j=1}^{N_t} \left( \tilde{\kappa}_{t,k_j,\tau_j} - \kappa(k_j, \tau_j, \hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) \right)^2, \quad t = 1, \dots, T, \quad (30)$$

and  $w(x)$  is a twice continuously differentiable function on  $\mathbb{R}_+$  with bounded first and second derivatives. Moreover,  $w(x) \geq \epsilon$  for some  $\epsilon > 0$ . An example of such a function is a smooth approximation of  $x \vee \epsilon$ . As expected, the above estimator  $\hat{\phi}_t$  is consistent for the cross-sectional average  $\mathcal{F}$ -conditional option variance at time  $t$  (generally a random number). Ideally, we would pick the weight function  $w(x) = x$ , but we rule this out for technical reasons.<sup>25</sup> Even so, with the class of weight functions  $w$ , described above, we may get arbitrary close to the identity function.

Given the above definitions, the resulting weighted penalized least squares (WPLS) estimator is defined as

$$\hat{\mathbf{S}}^w(\boldsymbol{\theta}) = \operatorname{argmin}_{\mathbf{Z} \in \mathcal{S}^T} \mathcal{L}^w(\mathbf{Z}, \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}, \quad (31)$$

and

$$\hat{\boldsymbol{\theta}}^w = \operatorname{argmin}_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathcal{L}^w(\hat{\mathbf{S}}^w(\boldsymbol{\theta}), \boldsymbol{\theta}), \quad \hat{\mathbf{S}}^w = \hat{\mathbf{S}}^w(\hat{\boldsymbol{\theta}}^w). \quad (32)$$

In the next theorem, we characterize its limiting distribution in the long span case, with similar extensions for the feasible central limit theorem readily applying to this scenario as well.

<sup>24</sup>Again, the stable convergence is critical for proving the limit effect from the weighting.

<sup>25</sup>However, under stronger conditions on the boundedness of the variances  $\phi_{t,k,\tau}$  and  $V_t$  from below, we may accommodate the case  $w(x) = x$ .

**Theorem 5.** *Suppose the conditions of Theorem 2 hold. Then, convergence results for  $\hat{\boldsymbol{\theta}}^w$  and  $(\hat{\mathbf{S}}_{t_1}^w, \dots, \hat{\mathbf{S}}_{t_k}^w)$  similar to those in (24) and (25) apply, as long as the quantities for their asymptotic bias and variances are replaced with their weighted analogues, defined in Section A.6.*

We note that, for our WPLS estimator, we do not impose any assumptions regarding the type of heteroskedasticity in the option observation error. This is unlike the classical weighted least squares, see, e.g., White (1982) and Robinson (1987), where heteroskedasticity is modeled as either a known or unknown function of observables. We do assume, however, that the  $\mathcal{F}$ -conditional option error variance is a smooth function of log-moneyness  $k$ . Then, our asymptotic setup of increasingly denser observations, across a given strike range for a specific point in time and tenor, generates consistent estimates of the  $\mathcal{F}$ -conditional option error variance by local (in moneyness) cross-sectional averaging.

The theorem above is stated for a general weight function  $w$ . It is optimal to set  $w$  close to the identity function. If the option error variance  $\phi_{t,k,\tau}$  is independent of  $k$  and  $\tau$ , while  $\phi_{t,k,\tau}$  and  $V_t$  are bounded from below and the option panel is balanced, then we obtain the optimally weighted penalized least squares. The limit results are now greatly simplified and take a familiar form, as evident from the corollary below.

**Corollary 1.** *Suppose the conditions of Theorem 5 hold and, in addition,*

$$\inf_{t \in \mathbb{R}_+} \inf_{\tau \in \mathcal{T}_t} \inf_{k \in [\underline{k}(t,\tau), \bar{k}(t,\tau)]} \phi_{t,k,\tau} > \epsilon, \quad \text{and} \quad \inf_{t \in \mathbb{R}_+} V_t > \epsilon \quad \text{for some finite } \epsilon > 0, \quad \phi_{t,k,\tau} = \phi_t.$$

*Finally, suppose  $\varsigma_t = 1$  for all  $t$ . Then, if  $w(x) = x$  for  $x > \epsilon$ , the asymptotic distribution of  $\hat{\boldsymbol{\theta}}^w$  and  $\hat{\mathbf{S}}^w$  is characterized as follows.*

$$\sqrt{TN} \left( \hat{\boldsymbol{\theta}}^w - \boldsymbol{\theta}_0 - \mathbf{B}_{T,N}^w \right) \xrightarrow{\mathcal{L}^{-s}} (\mathcal{I}_U^w + \mathcal{P}_\theta^w)^{-1/2} \times \mathbf{E}_\theta, \quad (33)$$

*where  $\mathbf{E}_\theta$  is a  $q \times 1$  standard Gaussian vector, defined on an extension of the original probability space and independent from  $\mathcal{F}$ , and the asymptotic bias has the form,*

$$\mathbf{B}_{T,N}^w = -(\mathcal{I}_U^w + \mathcal{P}_\theta^w)^{-1} \left( \mathbf{B}_{T,N,3}^{(1,w)} + \mathbf{B}_{T,N,3}^{(2,w)} \right).$$

*In addition, for arbitrary  $t_1, \dots, t_h \in \mathbb{N}$ , we have,*

$$\sqrt{N} \begin{pmatrix} \hat{\mathbf{S}}_{t_1}^w - \mathbf{S}_{t_1} \\ \vdots \\ \hat{\mathbf{S}}_{t_h}^w - \mathbf{S}_{t_h} \end{pmatrix} \xrightarrow{\mathcal{L}^{-s}} \text{diag} \begin{pmatrix} (\mathcal{I}_{S,t_1}^w + \mathcal{P}_{S,t_1}^w)^{-1/2} \\ \vdots \\ (\mathcal{I}_{S,t_h}^w + \mathcal{P}_{S,t_h}^w)^{-1/2} \end{pmatrix} \times \begin{pmatrix} \mathbf{E}_{S_{t_1}} \\ \vdots \\ \mathbf{E}_{S_{t_h}} \end{pmatrix}, \quad (34)$$

*where  $\mathbf{E}_{S_{t_1}}, \dots, \mathbf{E}_{S_{t_h}}$  is a sequence of standard Gaussian vectors, defined on an extension of the original probability space, with each of them independent of each other, of  $\mathbf{E}_\theta$ , as well as of  $\mathcal{F}$ , and the convergence in (34) holds jointly with that in (33). Finally, all the quantities describing the two asymptotic distributions,  $\mathcal{I}_U^w$ ,  $\mathcal{P}_\theta^w$ ,  $\mathcal{I}_{S,t_h}^w$ ,  $\mathcal{P}_{S,t_h}^w$ , and  $\mathbf{B}_{T,N}^w$ , are defined in Section A.6.*

One convenient aspect of the WPLS estimator is that we need not choose the weight assigned to the penalty term in the objective function. In particular, the tuning parameter  $\lambda$  in the original PLS criterion is replaced by the time-varying ratio of the option variance and the asymptotic variance of the nonparametric variance estimator. As a result, the WPLS estimator alters the objective function period-to-period, adapting to the relative precision of the option pricing and high-frequency volatility estimator. From a practical point of view, this is attractive since the precision of these measures vary substantially over time. The WPLS procedure ensures that the weight of the two components in the objective function reflects their relative information content.

## 6 Numerical Experiments

We now undertake numerical experiments to assess the practical relevance of our unified inference techniques. For this purpose, we rely on a representation that nests many existing models and captures the complex dynamic features of option panels quite well. It takes the form of a two-factor jump-diffusive stochastic volatility model for the risk-neutral distribution,

$$\begin{aligned}\frac{dX_t}{X_{t-}} &= (r_t - \delta_t) dt + \sqrt{V_{1,t}} dW_{1,t}^{\mathbb{Q}} + \sqrt{V_{2,t}} dW_{2,t}^{\mathbb{Q}} + \int_{\mathbb{R}^2} (e^x - 1) \tilde{\mu}^{\mathbb{Q}}(dt, dx), \\ dV_{1,t} &= \kappa_1 (\bar{v}_1 - V_{1,t}) dt + \sigma_1 \sqrt{V_{1,t}} dB_{1,t}^{\mathbb{Q}} + \mu_v \int_{\mathbb{R}^2} x^2 1_{\{x < 0\}} \mu^{\mathbb{Q}}(dt, dx), \\ dV_{2,t} &= \kappa_2 (\bar{v}_2 - V_{2,t}) dt + \sigma_2 \sqrt{V_{2,t}} dB_{2,t}^{\mathbb{Q}},\end{aligned}\tag{35}$$

where  $(W_{1,t}^{\mathbb{Q}}, W_{2,t}^{\mathbb{Q}}, B_{1,t}^{\mathbb{Q}}, B_{2,t}^{\mathbb{Q}})$  is a four-dimensional Brownian motion with  $W_{1,t}^{\mathbb{Q}} \perp\!\!\!\perp W_{2,t}^{\mathbb{Q}}$ ,  $W_{1,t}^{\mathbb{Q}} \perp\!\!\!\perp B_{2,t}^{\mathbb{Q}}$ , and  $W_{2,t}^{\mathbb{Q}} \perp\!\!\!\perp B_{1,t}^{\mathbb{Q}}$ , while  $\text{corr}(W_{1,t}^{\mathbb{Q}}, B_{1,t}^{\mathbb{Q}}) = \rho_{d,1}$  and  $\text{corr}(W_{2,t}^{\mathbb{Q}}, B_{2,t}^{\mathbb{Q}}) = \rho_{d,2}$ . The risk-neutral compensator for the jump measure is given by

$$\begin{aligned}\nu_t^{\mathbb{Q}}(dx, dy) &= \left\{ \left( c^-(t) 1_{\{x < 0\}} \lambda_- e^{-\lambda_- |x|} + c^+(t) 1_{\{x > 0\}} \lambda_+ e^{-\lambda_+ x} \right) \right\} dx, \\ c^-(t) &= c_1^- V_{1,t-} + c_2^- V_{2,t-}, \quad c^+(t) = c_0^+ + c_1^+ V_{1,t-} + c_2^+ V_{2,t-}.\end{aligned}$$

We exploit the parameter estimates for the model from Andersen et al. (2015b), reproduced in Table 1. Of course, these are extracted from actual option panels, thus reflecting the risk-neutral dynamics, and not necessarily the actual statistical (or physical) probability measure. In fact, option prices are known to embody sizable risk premiums, so to generate the actual evolution of the simulated system, we modify the above specification of the volatility processes to more closely resemble the actual observed volatility dynamics, i.e., we account for the risk premiums embedded in the risk-neutral parameter estimates. In particular, we increase  $\kappa_2$  to 0.69, implying that a shock to  $V_{2,t}$  has a (shorter) half-life of one year, and we limit the magnitude of the jumps in  $V_t$  to not exceed 0.40.

Table 1: Parameter Values for model (35) used in Numerical Experiments

Parameter	Value	Parameter	Value	Parameter	Value
$\rho_1$	-0.762	$\kappa_2$	0.169	$c_2^+$	52.795
$\bar{v}_1$	0.004	$\sigma_2$	0.129	$\lambda_-$	16.943
$\kappa_1$	12.831	$c_0^+$	2.315	$\lambda_+$	51.818
$\sigma_1$	0.247	$c_1^-$	69.208	$\mu_v$	6.262
$\rho_2$	-0.945	$c_1^+$	13.161		
$\bar{v}_2$	0.059	$c_2^-$	97.656		

In terms of the option observation scheme, we follow the Monte Carlo setup of Andersen et al. (2015a), but explore the impact of the time dimension of the panel through experiments with different time spans. Specifically, consistent with standard practice, we sample option data every fifth day (time is measured in business days). At each observation date, we obtain option prices at four separate maturities:  $\tau = 10$ ,  $\tau = 45$ ,  $\tau = 120$  and  $\tau = 252$  days. For each tenor, we have 50 OTM option prices for an equispaced log-moneyness grid, covering the range  $[-4, 1] \cdot \sigma\sqrt{\tau}$ , where  $\sigma$  is the current at-the-money BSIV. This implies a time-varying coverage in terms of moneyness, depending on the level of market volatility, roughly mimicking the characteristics of actual option panels. For the option error, we assume  $\epsilon_{t,k,\tau} = \sigma_{t,k,\tau} Z_{t,k,\tau}$ , where  $Z_{t,k,\tau}$  are standard normal variables, independent across time, moneyness and tenor, while  $\sigma_{t,k,\tau} = 0.5 \psi_k / Q_{0.995}$ . For the latter,  $\psi_k$  represents the estimate of the relative bid-ask spread from a kernel regression on actual S&P 500 index options, as reported in Andersen et al. (2015a), and  $Q_{0.995}$  denotes the 0.995-quantile of the standard normal distribution. This noise structure allows for significant fluctuations in the (conditional) noise variance as a function of both the level of volatility and moneyness.<sup>26</sup>

Overall, our simulations involve  $N = 200$  options per day, and we experiment with a span ranging from 1 to 20 years. Since we sample once per week, this generates panels of about 50 to 1,000 cross sections, each containing 200 option observations. Since the cost in time for numerical evaluation of the option prices is nontrivial, we do not perform estimation of the parameters and the factors on the basis of the simulated option data. Instead, in what follows we report our estimates for the asymptotic variance and the asymptotic bias of the parameters and the factors at their true values.

<sup>26</sup>In the absence of parametric assumptions on the actual factor dynamics (as in our theoretical treatment), the Gaussian specification of the option errors implies that the optimally weighted least squares estimator constitutes the maximum likelihood estimator.

## 6.1 Estimation without Penalization

We start with a baseline case where we do not penalize with a high-frequency volatility estimator and abstain from weighting the option observations.

Let us first consider the effect of the time span on the recovery of the factor realizations. Figure 1 plots the estimate for the standard error associated with the recovery of  $V_1$  and  $V_2$ . We consider observations recorded in the last week of the first year of the sample, as this period is present for all time spans explored in our experiments. For ease of interpretation, we report the relative standard error, obtained by dividing the standard error estimate by the true value of the factor (which is random). From Theorem 2 and Lemma 1, we recall that, regardless of the time span, the rate of convergence of the factor extraction is  $\sqrt{N}$ . Moreover, we reiterate that we keep  $N$  fixed in our experiments (mimicking what occurs in practice) and that the limit distribution is mixed Gaussian. These features are reflected in Figure 1. Indeed, the interquartile range of the relative standard error does not shrink with the length of the span. This is due to the fact that the limiting variance of the error in the factor recovery depends on the realized (random) path of the state vector. In addition, the error in recovering the factor realizations tends to be higher for the shorter time spans. This reflects the effect of parameter uncertainty captured by the terms  $\{\mathbf{AVAR}_{S_i,j}\}_{j=2,3,4}$  in Lemma 1. As the time span grows, this effect diminishes, and the quantiles for the relative standard error stabilize and become insensitive to any further lengthening of the span. Interestingly, this occurs at different horizons for the two factors. For  $V_1$ , the “stabilization” of the standard error around its long span value is complete after about 6-7 years. For  $V_2$ , however, this occurs only after 15-16 years. The reason is that the second factor is far more persistent than the first and, hence, it is harder to achieve a proper separation between its realization and long-run parameter-driven properties. Another noteworthy discrepancy is the relatively larger error in the recovery of  $V_1$  and the significant heterogeneity in the precision across samples. This is largely due to the erratic behavior of  $V_1$ , which stems from the random occurrence of large jumps implied by the dynamic specification for this factor.

We next study the intertemporal persistence of the error in the extraction of the factor realizations across the sample. In Figure 2, we plot estimates for this error serial correlation over the last two weeks of the first year in the simulation. Theorem 2 and Lemma 1 imply that this dependence stems exclusively from the finite time span. That is, the intertemporal error correlation is due to imprecision in the estimation of the parameter vector and, as we increase the span, the effect of the latter vanishes asymptotically. Indeed, for both factors, the correlation shrinks as we increase the time span. Nonetheless, the persistence of the error in the recovery of  $V_2$  remains fairly high, even for a panel covering 20 years. This is consistent with the evidence from Figure 1, where we noted that estimation errors for the model parameters have a strong effect on the recovery of this factor,

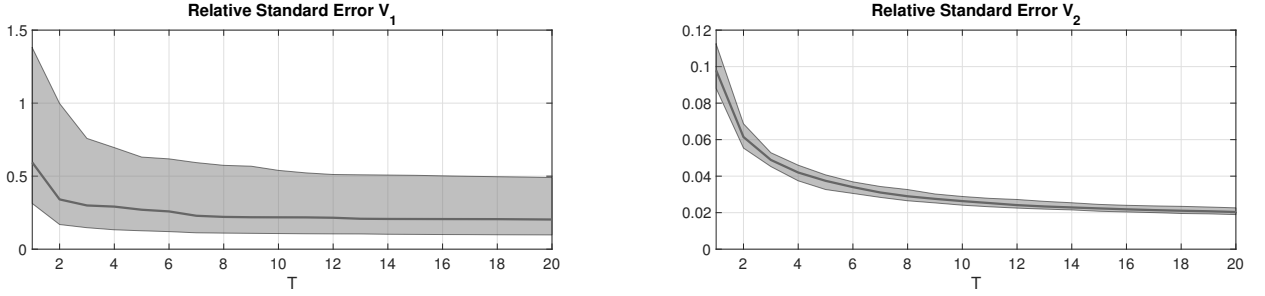


Figure 1: *Relative Standard Errors for Factors: Case of no Penalization.* Standard errors are computed for the factor realization from the last week of the first year in the sample. Standard errors are divided by the absolute value of the true value of the factor realization. The straight line in each of the plots corresponds to the median and the shaded area to the inter-quantile range, both computed from 100 draws of the process  $(V_{1,t}, V_{2,t})$ .

even for panels with a long span. If the econometrician relies on our long span asymptotic result, he or she would be ignoring this nontrivial persistence in the error of the factor extraction. In contrast, our unified approach, advocated in Theorem 4, automatically accommodates this feature and more accurately conveys the degree of temporal dependence in the errors of the recovered factors.

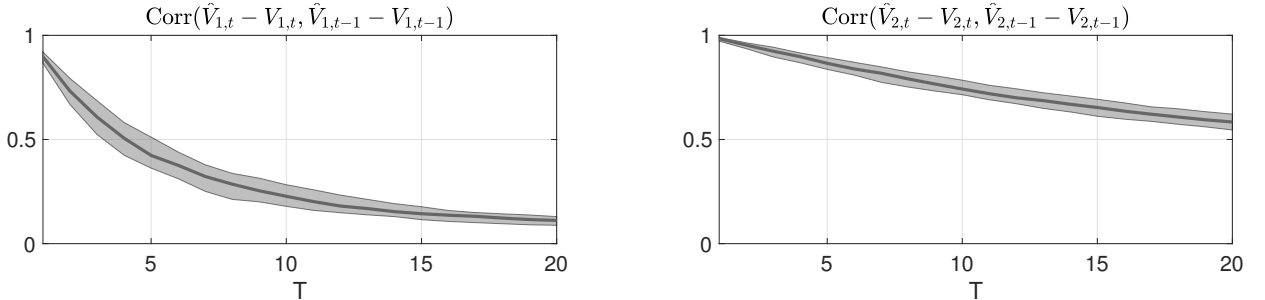


Figure 2: *Persistence of the Error in Factor Extraction: Case of no Penalization.* The plots display correlation between the errors in extracting factors over the last two weeks in the first year of the sample. The straight line in each of the plots corresponds to the median and the shaded area to the inter-quantile range, both computed from 100 draws of the process  $(V_{1,t}, V_{2,t})$ .

We conclude with a discussion of the bias and standard error in parameter estimation. Figure 3 provides relative estimates for these quantities (by dividing them by their true values). To conserve space, we present results for a subset of the parameters only, as the findings for the remaining parameters are qualitatively similar. Contrary to the short time span scenario, the long span limit distribution of the parameter vector is Gaussian, i.e., the limiting variance is independent of the sample path. We note that, for the shorter time spans, the interquantile ranges of the relative standard errors tend to be quite wide but, unlike the case for the factors, they continue to shrink as the time span grows and

eventually become negligible. We also see a large reduction in the bias for some parameters as the span lengthens. However, since we keep  $N$  fixed, the estimate for the bias converges to a constant as the span grows. As for the factor recovery, the “transformation” of the features of the limiting distribution, induced by the transition from the fixed to long span asymptotics, occurs at different times across the parameters. The relative bias is larger for the parameters determining the positive jumps ( $c_1^+$  and  $c_2^+$ ) and smaller for the parameters controlling the mean reversion of the volatility factors ( $\kappa_1$  and  $\kappa_2$ ). The same ranking applies for the relative standard error of the parameters. Comparing the bias to the standard error, we also observe a nontrivial variation across the parameters. For the longest span of 20 years, the bias is approximately of the same size as the standard error for  $c_1^-$  and negligible relative to the standard error for  $\kappa_2$ . Thus, ignoring the bias, i.e., performing inference according to the fixed span asymptotics, will imply poor performance for tests involving parameters such as  $c_1^-$  for time spans as short as 5 years, again corroborating the benefit of relying on the unified inference procedures.

## 6.2 Estimation with Penalization

We proceed by studying the impact of penalizing the estimation with a high-frequency spot volatility measure. To this end, let us set  $k_n = 180$ , which corresponds to using the last three hours of one-minute high frequency observations of the underlying asset. Moreover, we set  $\lambda$  equal to the average value of  $\hat{\phi}_t/(2\hat{V}_t^2)$ , computed over one realization of the options and a volatility path of length  $T = 20$  years. For our parametric setup, this results in  $\lambda$  approximately being 0.04. Note that  $\hat{\phi}_t/(2\hat{V}_t^2)$  is the optimal value of the penalization parameter, and it is generally time-varying.

Figure 4 shows that the introduction of the penalization term in the objective function has little effect on the precision of the factor extraction. The standard errors for  $V_{2,t}$  are almost identical in the two cases, while the standard errors of  $V_{1,t}$  are slightly higher. This may be explained by a constant  $\lambda$  being implemented for the whole sample. This leads to an over-weighting of the penalization term in high volatility periods, which is particularly relevant for  $V_{1,t}$  as it contains frequent jumps. On the other hand, from Figure 5, we notice a nontrivial reduction in the correlation in the error of recovering the factor realizations of the more persistent volatility factor,  $V_{2,t}$ . This is to be expected as the high-frequency volatility estimators contain errors that are uncorrelated across days.

As evident from Figure 6, the penalization also has a nontrivial effect on the parameter recovery. First, the penalization affects the asymptotic bias, and its impact differs across the parameters of the model. Second, the penalization substantially reduces the standard errors of the parameters that control the jump intensity and the jump distribution. In the two-factor volatility model considered here, the jump intensity is a linear function of the two volatility states and the penalization helps to disentangle the jump intensity parameters ( $c_i^\pm$ , for  $i = 0, 1, 2$ ) from the factor realizations.

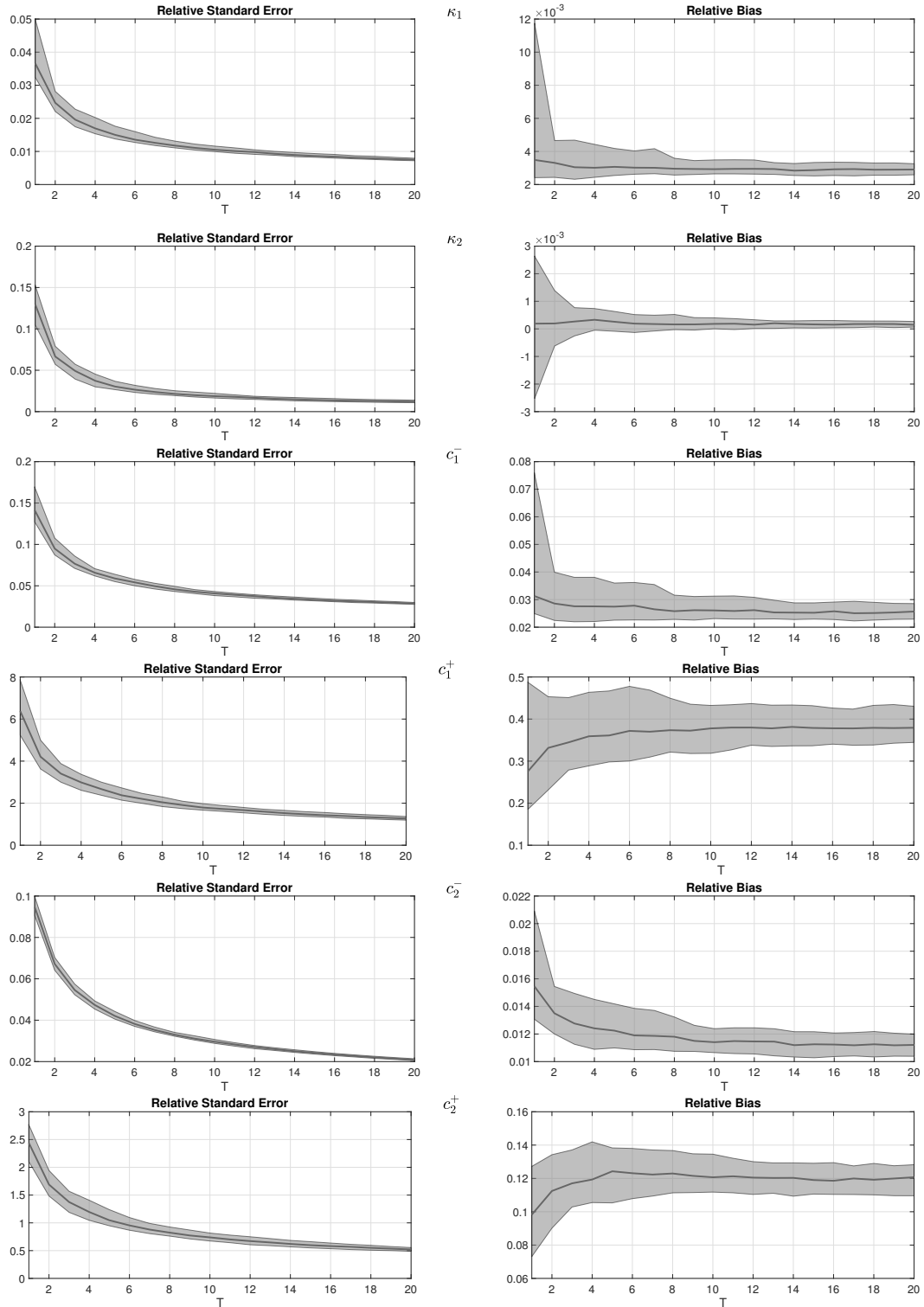


Figure 3: *Relative Standard Error and Bias for Parameters: Case of no Penalization.* Standard errors and biases are divided by the absolute value of the true value of the parameter reported in Table 1. The straight line in each of the plots corresponds to the median and the shaded area to the inter-quartile range, both computed from 100 draws of the process  $(V_{1,t}, V_{2,t})$ .



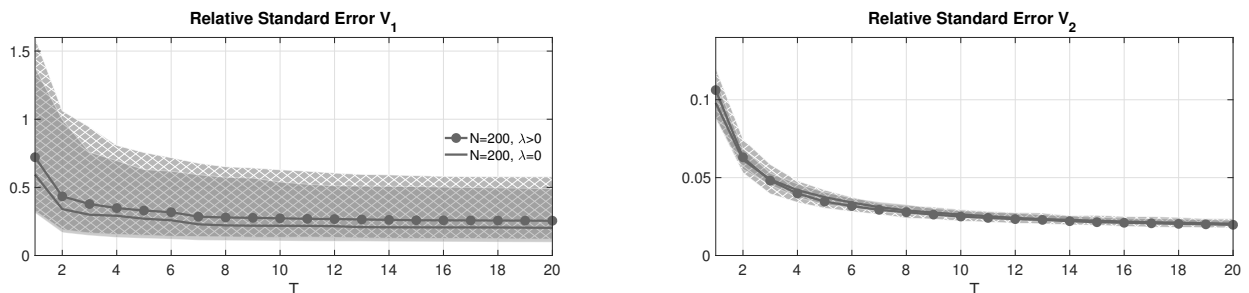


Figure 4: *Relative Standard Errors for Factors: Effect of Penalization.* Standard errors are computed for the factor realization from the last week of the first year in the sample. Standard errors are divided by the absolute value of the true value of the factor realization. The straight line in each of the plots corresponds to the median and the shaded area to the inter-quantile range, both computed from 100 draws of the process  $(V_{1,t}, V_{2,t})$ .

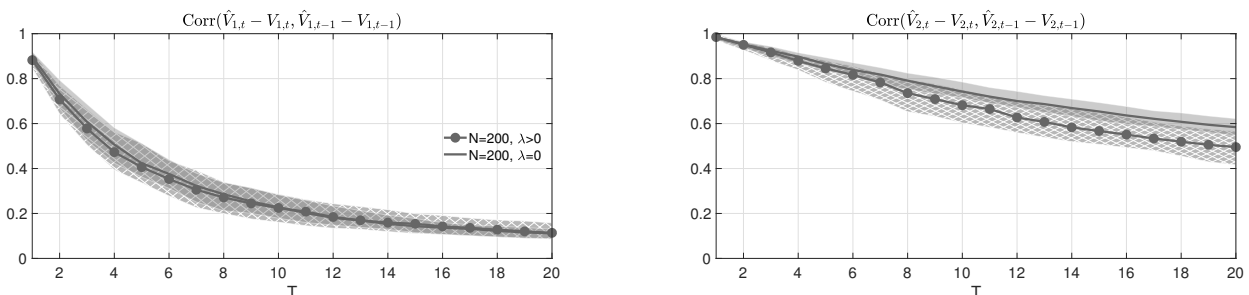


Figure 5: *Persistence of the Error in Factor Extraction: Effect of Penalization.* The plots display correlation between the errors in extracting factors over the last two weeks in the first year of the sample. The straight line in each of the plots corresponds to the median and the shaded area to the inter-quantile range, both computed from 100 draws of the process  $(V_{1,t}, V_{2,t})$ .

## 7 Conclusion

In this paper, we develop unified inference theory for parametric nonlinear factor models based on a panel of noisy observations augmented with direct, albeit imperfect, measures of a known transformation of the factors. The panel has a large cross-sectional dimension while its time span is either fixed or increasing. The estimation is done via penalized least squares, with the penalty term controlling the size of the deviation between noisy observations and model-implied values for the known transformation of the factor realizations. Even though the limit distributions of the parameter vector and the factor realizations differ substantially in the fixed and increasing time span settings, we develop a feasible inference procedure that accommodates the critical features of the asymptotic distribution in either setup. We further propose a data-driven method for the weighting of the penalty term in the objective function, reflecting the relative information content of the noisy factor signals.

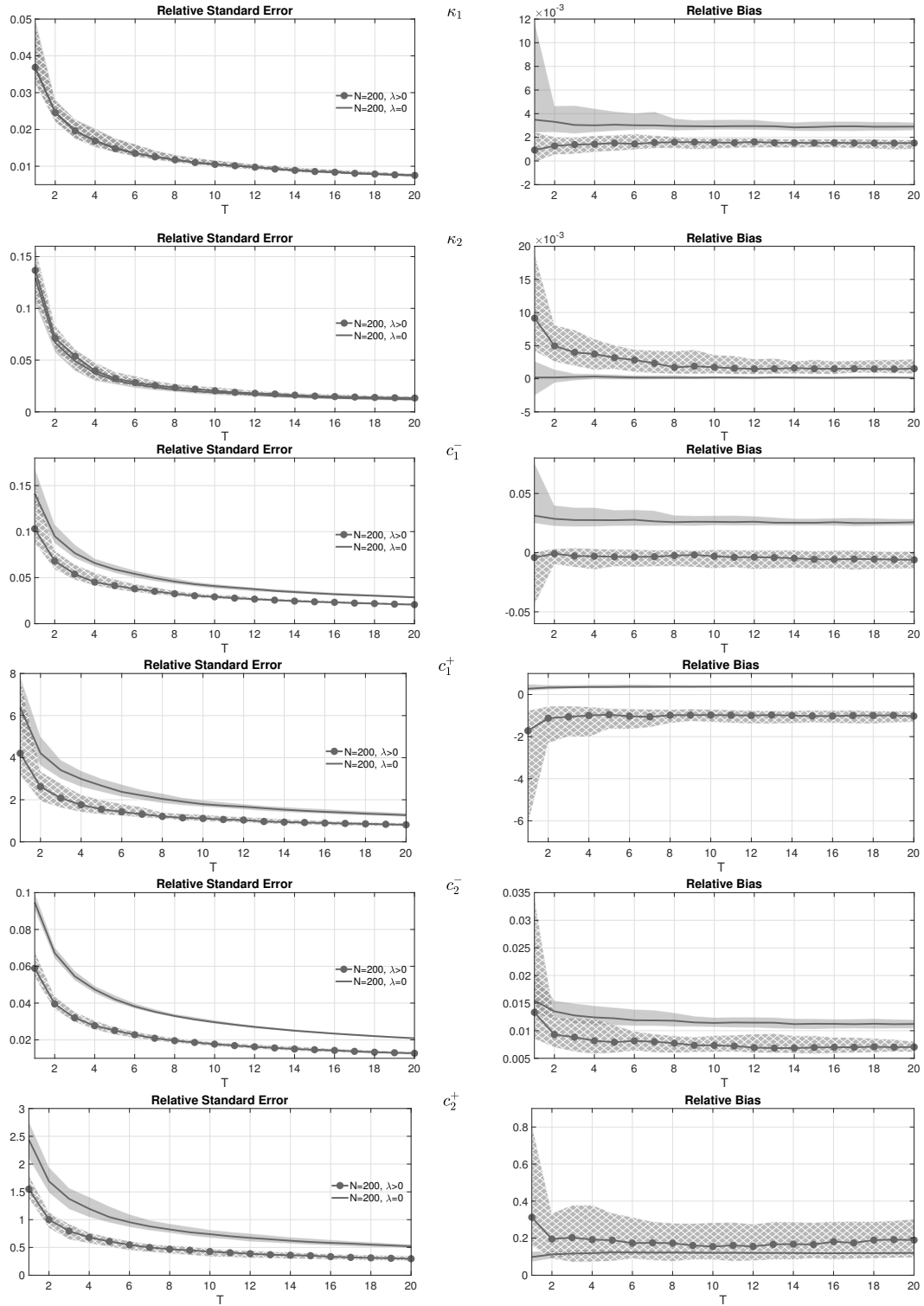


Figure 6: *Relative Standard Error and Bias for Parameters: Effect of Penalization.* Standard errors and biases are divided by the absolute value of the true value of the parameter reported in Table 1. The straight line in each of the plots corresponds to the median and the shaded area to the inter-quantile range, both computed from 100 draws of the process  $(V_{1,t}, V_{2,t})$ .

Numerical analysis demonstrates the practical relevance of our unified inference approach, as it avoids exclusive reliance on either the fixed or long time span asymptotics in any given application. In the scenarios explored, to a varying degree, we identify features of the limiting distribution that are captured best by the fixed span asymptotics (e.g., the intertemporal persistence of the error in factor recovery) and others that fall much closer to the long span asymptotics (e.g., the relative importance of bias in the parameter estimates). Our approach allows the data to determine the relative weight that is afforded the two asymptotic schemes for different aspects of the inference.

## References

- Andersen, T. G., D. Dobrev, and E. Schaumburg (2012). Robust Volatility Estimation Using Nearest Neighbor Truncation. *Journal of Econometrics* 169, 75–93.
- Andersen, T. G., N. Fusari, and V. Todorov (2015a). Parametric Inference and Dynamic State Recovery from Option Panels. *Econometrica* 83, 1081–1145.
- Andersen, T. G., N. Fusari, and V. Todorov (2015b). The Risk Premia Embedded in Index Options. *Journal of Financial Economics* 117, 558–584.
- Andrews, D. (2005). Cross-Section Regression with Common Shocks. *Econometrica* 71, 1551–1585.
- Arellano, M. and S. Bonhomme (2009). Robust Priors in Nonlinear Panel Data Models. *Econometrica* 77, 489–536.
- Arellano, M. and J. Hahn (2007). Understanding Bias in Nonlinear Panel Models: Some Recent Developments. In R. Blundell, W. Newey, and T. Persson (Eds.), *Advances in Economics and Econometrics, Ninth World Congress*, Volume 3, pp. 381–409. Cambridge University Press.
- Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica* 71, 135–171.
- Bai, J. and S. Ng (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica* 70, 191–221.
- Barndorff-Nielsen, O. E. and N. Shephard (2004). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics* 2, 1–37.
- Barndorff-Nielsen, O. E. and N. Shephard (2006). Econometrics of Testing for Jumps in Financial Economics using Bipower Variation. *Journal of Financial Econometrics* 4, 1–30.
- Bates, D. S. (2003). Empirical Option Pricing: A Retrospection. *Journal of Econometrics* 116, 387–404.
- Bernanke, B., J. Boivin, and P. Elias (2005). Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *Quarterly Journal of Economics* 120, 387–422.
- Christoffersen, P. and K. Jacobs (2004). The Importance of the Loss Function in Option Valuation. *Journal of Financial Economics* 72, 291–318.
- Duffie, D. (2001). *Dynamic Asset Pricing Theory* (3rd ed.). Princeton University Press.
- Duffie, D., J. Pan, and K. Singleton (2000). Transform Analysis and Asset Pricing for Affine Jump-Diffusions. *Econometrica* 68, 1343–1376.

- Eraker, B. (2004). Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices. *Journal of Finance* 59, 1367–1403.
- Fernandez-Val, I. and J. Lee (2013). Panel Data Models with Nonadditive Unobserved Heterogeneity: Estimation and Inference. *Quantitative Economics* 4, 453–481.
- Fernandez-Val, I. and M. Weidner (2016). Individual and Time Effects in Nonlinear Panel Models with Large N,T. *Journal of Econometrics* 196, 291–312.
- Gagliardini, P. and C. Gourieroux (2014). Efficiency in Large Dynamic Panel Models with Common Factors. *Econometric Theory* 30, 961–1020.
- Gagliardini, P., C. Gourieroux, and E. Renault (2011). Efficient Derivative Pricing by the Extended Method of Moments. *Econometrica* 79, 1181–1232.
- Garcia, R., E. Ghysels, and E. Renault (2010). The Econometrics of Option Pricing. In Y. Aït-Sahalia and L. Hansen (Eds.), *Handbook of Financial Econometrics*, Volume 1, pp. 479–552. Oxford and Amsterdam: North-Holland.
- Hahn, J. and G. Kuersteiner (2002). Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when Both N and T are Large. *Econometrica* 70, 1639–1657.
- Hahn, J. and G. Kuersteiner (2011). Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects. *Econometric Theory* 27, 1152–1191.
- Hahn, J. and W. Newey (2004). Jackknife and Analytical Bias Reduction for Nonlinear Panel Models. *Econometrica* 72, 1295–1319.
- Jacod, J. and P. Protter (2012). *Discretization of Processes*. Berlin: Springer-Verlag.
- Jacod, J. and A. N. Shiryaev (2003). *Limit Theorems For Stochastic Processes* (2nd ed.). Berlin: Springer-Verlag.
- Kuersteiner, G. and I. Prucha (2013). Limit Theory for Panel Data Models with Cross Sectional Dependence and Sequential Exogeneity. *Journal of Econometrics* 174, 107–126.
- Lancaster, T. (2000). The Incidental Parameter Problem Since 1948. *Journal of Econometrics* 95, 391–413.
- Mancini, C. (2001). Disentangling the Jumps of the Diffusion in a Geometric Brownian Motion. *Giornale dell’Istituto Italiano degli Attuari LXIV*, 19–47.
- Neyman, J. and E. Scott (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica* 16, 1–32.
- Pan, J. (2002). The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study. *Journal of Financial Economics* 63, 3–50.
- Pastorello, S., V. Patilea, and E. Renault (2003). Iterative and Recursive Estimation in Structural Nonadaptive Models. *Journal of Business and Economic Statistics* 21, 449–482.
- Robinson, P. (1987). Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form. *Econometrica* 55, 875–891.
- White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica* 50, 1–25.

## A Appendix: Definitions, Assumptions and Proofs

In the proofs, we will use the following function:

$$f_d(x) = 1 + x^d, \quad x \in \mathbb{R}_+, \quad d \geq 0. \quad (36)$$

The matrix norm being used throughout is the Frobenius (or Euclidean) norm, that is, for an  $m \times n$  dimensional matrix  $\mathbf{A}$ , we denote  $\|\mathbf{A}\| = \sqrt{\sum_{i,j} a_{i,j}^2} = \sqrt{\text{Tr}(\mathbf{A}\mathbf{A}')}.$  Let us also denote with  $\mathcal{Q}$  the set in which  $(\mathcal{T}_t, \underline{k}_t, \bar{k}_t)$  takes values, and we further set  $\bar{\mathcal{T}} = \sup_{t \in \mathbb{R}_+} \mathcal{T}_t.$  Moreover,  $K$  is a generic constant, which may take different values in different places, and  $\mathbb{E}_0(\cdot)$  and  $\mathbb{V}_0(\cdot)$  denote  $\mathbb{E}(\cdot | \mathcal{F}_T^{(0)})$  and  $\mathbb{V}(\cdot | \mathcal{F}_T^{(0)})$ , respectively. Note that (stochastic) orders sometimes refer to scalars, vectors, and sometimes to matrices. Finally, we will in the proofs of the main asymptotic results refer to various auxiliary lemmas, which are collected and established in Section A.7.

### A.1 Assumptions

**Assumption 1.** *The process  $X$  in (1) satisfies:*

- (i) For every  $p > 0$ , we have  $\sup_{t \in \mathbb{R}_+} \mathbb{E}(|\alpha_t|^p) < \infty$ ,  $\sup_{t \in \mathbb{R}_+} \mathbb{E}(|V_t|^p) < \infty$  and  $\sup_{t \in \mathbb{R}_+} \mathbb{E}(|a_t|^p) < \infty$ , and further  $\int_{x > -1} (|\log(1+x)|^r \vee |\log(1+x)|^p) \nu^{\mathbb{P}}(dx) < \infty$ , for some  $r \in [0, 1)$  and every  $p \geq r$ .
- (ii) For every  $t \in \mathbb{R}_+$  and  $s \in \mathbb{R}_+$  and  $p \geq 2$ , we have  $|\mathbb{E}_t(V_{t+s} - V_t)| + \mathbb{E}_t|V_{t+s} - V_t|^p \leq K_{t,s}^p |t - s|$  with  $\sup_{t \in \mathbb{R}_+} \mathbb{E}|K_{t,s}^p|^q < \infty$  for  $q > 0$  arbitrarily high.

**Assumption 2.** (i) As  $N \rightarrow \infty$  and  $T \rightarrow \infty$  with  $T/N \rightarrow \gamma$  for some finite  $\gamma \geq 0$ , we have

$$\sup_{t \in [1, T], \tau \in \mathcal{T}_t} \left| \frac{N_t^\tau}{N_t} - \pi_t^\tau \right| \xrightarrow{\mathbb{P}} 0, \quad \sqrt{N} \sup_{t \in [1, T]} \left| \frac{N_t}{N} - \varsigma_t \right| \xrightarrow{\mathbb{P}} 0, \quad (37)$$

where  $(\pi_t^\tau)_{\tau \in \mathcal{T}_t}$  and  $\varsigma_t$  are  $\mathcal{F}_t^{(0)}$ -adapted with  $\inf_{t \in \mathbb{R}_+, \tau \in \mathcal{T}_t} \pi_t^\tau > 0$  and  $\sup_{t \in \mathbb{R}_+, \tau \in \mathcal{T}_t} \pi_t^\tau < \infty$  as well as  $\inf_{t \in \mathbb{R}_+} \varsigma_t > 0$  and  $\sup_{t \in \mathbb{R}_+} \varsigma_t < \infty$ .

(ii) Moreover, for the grids of strike prices, with  $i_k = \min \{i \geq 2 : k_{t,\tau}(i) \geq k\}$ , for  $k \in [\underline{k}(t, \tau), \bar{k}(t, \tau)]$ , we have

$$\sup_{t \in \mathbb{R}_+} \sup_{\tau \in \mathcal{T}_t} \sup_{k \in [\underline{k}(t, \tau), \bar{k}(t, \tau)]} \sqrt{N_t^\tau} |N_t^\tau \Delta_{t,\tau}(i_k) - \psi_{t,\tau}(k)| \xrightarrow{\mathbb{P}} 0, \quad (38)$$

where  $(\psi_{t,\tau}(k))_{k \in [\underline{k}(t, \tau), \bar{k}(t, \tau)], \tau \in \mathcal{T}_t}$  is some  $\mathcal{F}_t^{(0)}$ -adapted process with

$$\inf_{t \in \mathbb{R}_+, \tau \in \mathcal{T}_t, k \in [\underline{k}(t, \tau), \bar{k}(t, \tau)]} \psi_{t,\tau}(k) > 0, \quad \text{and} \quad \sup_{t \in \mathbb{R}_+, \tau \in \mathcal{T}_t, k \in [\underline{k}(t, \tau), \bar{k}(t, \tau)]} \psi_{t,\tau}(k) < \infty,$$

and, for its derivative, we have  $\sup_{t \in \mathbb{R}_+, \tau \in \mathcal{T}_t, k \in [\underline{k}(t, \tau), \bar{k}(t, \tau)]} \psi'_{t,\tau}(k) < \infty$ .

(iii) Finally, we have  $k_n/N - \varrho = O(1/\sqrt{N})$ .

**Assumption 3.** For some  $d \geq 0$ , we have:

(i) For every  $\epsilon > 0$ , there exists  $\delta > 0$ , such that for  $\mathbf{S} \in \mathcal{S}$ ,

$$\inf_{(\mathcal{T}_t, \underline{k}_t, \bar{k}_t) \in \mathcal{Q}} \inf_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \geq \epsilon \cup \|\mathbf{Z} - \mathbf{S}\| \geq \epsilon} \sum_{\tau \in \mathcal{T}_t} \pi_t^\tau \times \int_{\underline{k}(t, \tau)}^{\bar{k}(t, \tau)} \frac{1}{\psi_{t, \tau}(k)} (\kappa(k, \tau, \mathbf{Z}, \boldsymbol{\theta}) - \kappa(k, \tau, \mathbf{S}, \boldsymbol{\theta}_0))^2 dk > \delta f_d(\|\mathbf{S}\|). \quad (39)$$

(ii) For every  $\delta > 0$ , there exist  $\epsilon > 0$ , such that

$$\inf_{(\mathcal{T}_t, \underline{k}_t, \bar{k}_t) \in \mathcal{Q}} \inf_{\tau \in \mathcal{T}_t, k \in [\underline{k}(t, \tau), \bar{k}(t, \tau)]} (\kappa(k, \tau, \mathbf{Z}, \boldsymbol{\theta}) - \kappa(k, \tau, \mathbf{S}, \boldsymbol{\theta}_0))^2 \leq \delta f_d(\|\mathbf{S}\|) \implies \|\mathbf{S} - \mathbf{Z}\| < \epsilon. \quad (40)$$

**Assumption 4.** For  $\mathbf{S}, \mathbf{Z} \in \mathcal{S}$ ,  $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}$ ,  $k, k' \in \mathbb{R}$ , and  $\tau \leq \bar{\mathcal{T}}$ , we have for some  $d \geq 0$  and  $K > 0$ :

$$|\kappa(k, \tau, \mathbf{S}, \boldsymbol{\theta})|^2 \leq K f_d(\|\mathbf{S}\|), \quad |\kappa(k, \tau, \mathbf{S}, \boldsymbol{\theta}) - \kappa(k', \tau, \mathbf{S}, \boldsymbol{\theta})|^2 \leq K f_d(\|\mathbf{S}\|) |k - k'|, \quad (41)$$

$$|\kappa(k, \tau, \mathbf{S}, \boldsymbol{\theta}) - \kappa(k, \tau, \mathbf{S}, \boldsymbol{\theta}')|^2 \leq K f_d(\|\mathbf{S}\|) \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|, \quad (42)$$

$$|\kappa(k, \tau, \mathbf{S}, \boldsymbol{\theta}) - \kappa(k, \tau, \mathbf{Z}, \boldsymbol{\theta})|^2 \leq K(1 + \|\mathbf{S}\|^d + \|\mathbf{Z}\|^d) \|\mathbf{S} - \mathbf{Z}\|. \quad (43)$$

**Assumption 5.** For the error process,  $\epsilon_{t, k, \tau}$ , we have,

(i)  $\mathbb{E}(\epsilon_{t, k, \tau} | \mathcal{F}^{(0)}) = 0$ ,

(ii)  $\epsilon_{t, k, \tau}$  and  $\epsilon_{t', k', \tau'}$  are independent conditional on  $\mathcal{F}^{(0)}$ , whenever  $(t, k, \tau) \neq (t', k', \tau')$ ,

(iii)  $\mathbb{E}(|\epsilon_{t, k, \tau}|^q | \mathcal{F}^{(0)}) = \phi_{t, k, \tau}^{(q)}$ , where  $(\phi_{t, k, \tau}^{(q)})_{k \in [\underline{k}(t, \tau), \bar{k}(t, \tau)], \tau \in \mathcal{T}_t}$  is  $\mathcal{F}_t^{(0)}$ -adapted with

$$\sup_{t \in \mathbb{R}_+, \tau \in \mathcal{T}_t, k \in [\underline{k}(t, \tau), \bar{k}(t, \tau)]} \phi_{t, k, \tau}^{(q)} < \infty \quad \text{for every finite } q > 0.$$

In addition, the derivative of  $\phi_{t, k, \tau} \equiv \phi_{t, k, \tau}^{(2)}$  in its second argument satisfies

$$\sup_{t \in \mathbb{R}_+, \tau \in \mathcal{T}_t, k \in [\underline{k}(t, \tau), \bar{k}(t, \tau)]} \left| \frac{\partial}{\partial k} \phi_{t, k, \tau} \right| < \infty.$$

**Assumption 6.** The process  $\mathbf{S}_t$  is stationary and ergodic with  $\mathbb{E}\|\mathbf{S}_t\|^q < \infty$  for every finite  $q > 0$ . In addition,  $\boldsymbol{\mathcal{X}} = \text{plim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \boldsymbol{\mathcal{X}}_t$  exists, for  $\boldsymbol{\mathcal{X}}_t$  being one of the matrices  $\boldsymbol{\mathcal{I}}_{U, t}$ ,  $\boldsymbol{\Omega}_{U, t}$ ,  $\boldsymbol{\mathcal{I}}_{U, t}^w$  and  $\boldsymbol{\Omega}_{U, t}^w$  defined in Sections 3.2 and A.6.

**Assumption 7.** The functions  $\kappa(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$  and  $\xi_1(\mathbf{Z})$  are four times continuously differentiable in their arguments. Moreover, there exist constants  $K > 0$  and  $d > 0$  such that for  $\mathbf{S} \in \mathcal{S}$ ,  $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ ,  $k \in \mathbb{R}$ , and  $\tau \leq \bar{\mathcal{T}}$ ,

$$\|\mathbf{G}(k, \tau, \mathbf{S}, \boldsymbol{\theta})\| + \|\mathbf{G}^{(S)}(k, \tau, \mathbf{S}, \boldsymbol{\theta})\| + \|\mathbf{G}^{(SS)}(k, \tau, \mathbf{S}, \boldsymbol{\theta})\| \leq K f_d(\|\mathbf{S}\|), \quad (44)$$

$$\|\mathbf{G}^{(SSS)}(k, \tau, \mathbf{S}, \boldsymbol{\theta})\| + \|\mathbf{G}^{(\theta)}(k, \tau, \mathbf{S}, \boldsymbol{\theta})\| + \left\| \frac{\partial}{\partial k} \mathbf{G}(k, \tau, \mathbf{S}, \boldsymbol{\theta}) \right\| \leq K f_d(\|\mathbf{S}\|), \quad (45)$$

$$|\xi_1(\mathbf{S})| + \|\mathbf{C}(\mathbf{S})\| + \|\mathbf{C}^{(S)}(\mathbf{S})\| + \|\mathbf{C}^{(SS)}(\mathbf{S})\| \leq K f_d(\|\mathbf{S}\|), \quad (46)$$

and the bounds in (44)-(45) continue to hold with  $\mathbf{G}$  replaced by  $\mathbf{D}$ .

**Assumption 8.** Denote the eigenvalues of  $\mathcal{I}_{S,t} + \mathcal{P}_{S,t}$  with  $0 \leq \lambda_{t,1} \leq \lambda_{t,2} \dots \leq \lambda_{t,p}$ . Then, we have  $\lambda_{t,1} > K f_d(\|\mathbf{S}_t\|)$ , for some  $K > 0$  and  $d > 0$ . In addition,  $\mathcal{I}_{\theta,t}$  and  $\mathcal{I}_{U,t}$  are positive definite for all  $t$ .

**Discussion of the Assumptions.** Assumption 1 (henceforth, A1) imposes relatively weak restrictions on the  $\mathbb{P}$ -dynamics of  $X$ , which is satisfied by many existing asset pricing models. In particular, this assumption accommodates general forms of finite variation jumps in  $X$ , allowing for, e.g., compound Poisson processes that are commonly used in empirical option pricing applications. A2 imposes conditions on the regularity of the option data sampling. The asymptotics in the paper are of infill type in the sense that the mesh of the strike grid for each day and tenor shrinks to zero, and the identification condition (A3) is given with respect to the limiting case of observing continuum of options over the strike ranges for each day and tenor pairs. The requirements on the regularity of the sampling scheme in A2 are stronger than what is needed in the fixed span case, see Andersen et al. (2015a), since, in the current setting, the rates of convergence of the parameter vector can be much faster when the time span increases jointly with  $N$  (the average size of the cross-section). For example, if we have  $\psi_{t,\tau}(k) = 1$ , then the deviations from an asymptotically equidistant grid should only be  $o_p(1/\sqrt{N})$ . The conditions in A3-A5 are used to establish consistency of the PLS estimator, while Assumptions A5-A8 are needed to show asymptotic normality. As discussed in the main text, the assumption restricting the dependence between the observation errors in A5 can be relaxed to only require zero correlation between certain (cross-)products of them. In the case where the domain of the factors is bounded, the identification condition in A3 as well as the lower bound assumption on the eigenvalues of  $\mathcal{I}_{S,t} + \mathcal{P}_{S,t}$  in A8 are analogous to the corresponding assumptions in Hahn and Kuersteiner (2011). A4 and A7 are satisfied when  $\kappa(k, \tau, \mathbf{S}, \boldsymbol{\theta})$  is four times continuously differentiable in its arguments with derivatives having polynomial growth in the state vector  $\mathbf{S}$ . The convergence in probability conditions in Assumption A6 are satisfied whenever  $\psi_{t,\tau}$ ,  $\pi_t^\tau$ ,  $\varsigma_t$ ,  $\underline{k}(t, \tau)$  and  $\bar{k}(t, \tau)$  are functions of  $\mathbf{S}_t$ , do not have any other source of randomness, and are free of any other dependence on time, provided that  $\mathbb{E}|\boldsymbol{\chi}_t| < \infty$ , with  $\boldsymbol{\chi}_t$  being one of the matrices  $\mathcal{I}_{U,t}$ ,  $\boldsymbol{\Omega}_{U,t}$ ,  $\mathcal{I}_{U,t}^w$  and  $\boldsymbol{\Omega}_{U,t}^w$ .

As in Section 3.2, we will use the shorthand notation  $\epsilon_{j,t}$ ,  $\phi_{j,t}^{(q)}$ , and  $\phi_{j,t}$  for every pair  $(k_j, \tau_j)$ .

## A.2 Proof of Theorem 1

After removing the part of the objective function that does not depend on the parameter and state vectors, we have

$$(\hat{\mathbf{S}}'_1, \dots, \hat{\mathbf{S}}'_T, \hat{\boldsymbol{\theta}}') = \underset{\mathbf{Z} \in \mathcal{S}^T, \boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \left( \mathcal{L}^{(1)}(\mathbf{Z}, \boldsymbol{\theta}) + \mathcal{L}^{(2)}(\mathbf{Z}, \boldsymbol{\theta}) + \mathcal{L}^{(3)}(\mathbf{Z}, \boldsymbol{\theta}) \right),$$

where  $\mathcal{L}^{(i)}(\mathbf{Z}, \boldsymbol{\theta}) = \sum_{t=1}^T \mathcal{L}_t^{(i)}(\mathbf{Z}_t, \boldsymbol{\theta})$  for  $i = 1, 2, 3$ , with

$$\begin{aligned}\mathcal{L}_t^{(1)}(\mathbf{Z}_t, \boldsymbol{\theta}) &= \frac{1}{N_t} \sum_{j=1}^{N_t} (\kappa(k_j, \tau_j, \mathbf{S}_t, \boldsymbol{\theta}_0) - \kappa(k_j, \tau_j, \mathbf{Z}_t, \boldsymbol{\theta}))^2, \\ \mathcal{L}_t^{(2)}(\mathbf{Z}_t, \boldsymbol{\theta}) &= \frac{2}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} (\kappa(k_j, \tau_j, \mathbf{S}_t, \boldsymbol{\theta}_0) - \kappa(k_j, \tau_j, \mathbf{Z}_t, \boldsymbol{\theta})), \\ \mathcal{L}_t^{(3)}(\mathbf{Z}_t, \boldsymbol{\theta}) &= \lambda \frac{k_n}{N_t} [2(\hat{V}_t^n - V_t)(V_t - \xi_1(\mathbf{Z}_t)) + (V_t - \xi_1(\mathbf{Z}_t))^2],\end{aligned}$$

and we note that  $\mathcal{L}_t^{(3)}(\mathbf{Z}_t, \boldsymbol{\theta})$  does not depend on  $\boldsymbol{\theta}$ . We start by introducing some sets, which we will use throughout the proof. First, define

$$\Omega_T^{(1)} = \left\{ \omega : \left| \frac{1}{N_t} \sum_{j=1}^{N_t} (\epsilon_{j,t}^2 - \phi_{j,t}) \right| < \frac{1}{2} \frac{1}{N_t} \sum_{j=1}^{N_t} \phi_{j,t}, \quad \text{for } t = 1, \dots, T \right\}.$$

We then have

$$\mathbb{P}((\Omega_T^{(1)})^c) \leq \sum_{t=1}^T \mathbb{P} \left( \left| \frac{1}{N_t} \sum_{j=1}^{N_t} (\epsilon_{j,t}^2 - \phi_{j,t}) \right| \geq \frac{1}{2} \frac{1}{N_t} \sum_{j=1}^{N_t} \phi_{j,t} \right),$$

for which, using Burkholder-Davis-Gundy inequality and our integrability assumptions for  $\phi_{j,t}$  in A5, it follows that  $\mathbb{P}((\Omega_T^{(1)})^c) \rightarrow 0$ . Similarly, for

$$\Omega_T^{(2)} = \left\{ \omega : \left| \frac{1}{N_t} \sum_{j=1}^{N_t} (|\epsilon_{j,t}| - \phi_{j,t}^{(1)}) \right| < \frac{1}{2} \frac{1}{N_t} \sum_{j=1}^{N_t} \phi_{j,t}^{(1)}, \quad \text{for } t = 1, \dots, T \right\},$$

we have  $\mathbb{P}((\Omega_T^{(2)})^c) \rightarrow 0$ . Next, define

$$\Omega_T^{(3)}(\iota) = \left\{ \omega : \sup_{t=1, \dots, T} \|\mathbf{S}_t\| < T^\iota \right\}.$$

Then, using the integrability assumptions for  $\mathbf{S}_t$  in A6, we have  $\mathbb{P}((\Omega_T^{(3)}(\iota))^c) \rightarrow 0$ , for every fixed  $\iota > 0$ . Finally, we define

$$\Omega_T^{(4)}(\delta) = \left\{ \omega : \sup_{t=1, \dots, T} \lambda \frac{k_n}{N} |\hat{V}_t^n - V_t|^2 < \delta \right\}, \quad \delta > 0.$$

Using part (b) of Lemma 8, we readily have  $\mathbb{P}((\Omega_T^{(4)}(\delta))^c) \rightarrow 0$ , for every fixed  $\delta > 0$ .

Pick a constant  $\iota \in (0, 1/10)$  and denote  $\bar{\phi} = \sup_{t \in \mathbb{R}_+, \tau \in \mathcal{T}_t, k \in [\underline{k}(t, \tau), \bar{k}(t, \tau)]} \phi_{t, k, \tau}$ . From A3(ii), there exist  $\epsilon > 0$  such that if  $\|\mathbf{S}_t - \mathbf{Z}\| \geq \epsilon$ , then

$$\mathcal{L}_t^{(1)}(\mathbf{Z}, \boldsymbol{\theta}) > \left( \frac{15}{4} \bar{\phi} \vee \frac{6\bar{\phi}}{\iota^2} \vee 10K \sup_{\boldsymbol{\theta} \in \Theta} \|\boldsymbol{\theta}\| f_d(\|\mathbf{S}_t\|) \right),$$

where  $K$  is the constant showing in the second bound of A4 (for the first two lower bounds, note that  $\frac{15}{4}\bar{\phi}$  and  $\frac{6\bar{\phi}}{\iota^2}$  are finite constants, and, for the last bound, recall that  $\Theta$  is a compact set).



We will first show that on a set with probability approaching one, we have  $\sup_{t=1,\dots,T} \|\hat{\mathbf{S}}_t - \mathbf{S}_t\| < \epsilon$ . Without further mentioning, in doing this, we will work on the set  $\Omega_T^{(1)} \cap \Omega_T^{(4)}(6\bar{\phi}/\iota)$ , whose probability approaches one. Suppose that for some  $t = 1, \dots, T$ , we have  $\|\hat{\mathbf{S}}_t - \mathbf{S}_t\| \geq \epsilon$ . Then, since by the Cauchy-Schwartz inequality,  $\mathcal{L}_t^{(2)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) \leq 2\sqrt{\mathcal{L}_t^{(1)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}})}\sqrt{\frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t}^2}$ , and since  $\mathcal{L}_t^{(1)}(\mathbf{Z}, \boldsymbol{\theta}) > \frac{6\bar{\phi}}{\iota^2}$  for  $\|\hat{\mathbf{S}}_t - \mathbf{Z}\| \geq \epsilon$  and in restriction to the set  $\Omega_T^{(1)}$ , this shows that

$$\mathcal{L}_t^{(1)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) + \mathcal{L}_t^{(2)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) > (1 - \iota)\mathcal{L}_t^{(1)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}).$$

Moreover, using the algebraic inequality  $2|xy| \leq x^2 + y^2$  and the definition of the set  $\Omega_T^{(4)}(\delta)$ , we have  $\mathcal{L}_t^{(3)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) \geq -\frac{6\bar{\phi}}{\iota}$ . Altogether, we get  $\mathcal{L}_t^{(1)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) + \mathcal{L}_t^{(2)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) + \mathcal{L}_t^{(3)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) > (1 - 2\iota)\mathcal{L}_t^{(1)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}})$ .

Now, we will evaluate  $\mathcal{L}_t^{(1)}(\mathbf{S}_t, \hat{\boldsymbol{\theta}}) + \mathcal{L}_t^{(2)}(\mathbf{S}_t, \hat{\boldsymbol{\theta}}) + \mathcal{L}_t^{(3)}(\mathbf{S}_t, \hat{\boldsymbol{\theta}})$ . First, using the second bound of A4, we have

$$|\mathcal{L}_t^{(1)}(\mathbf{S}_t, \hat{\boldsymbol{\theta}})| = |\mathcal{L}_t^{(1)}(\mathbf{S}_t, \hat{\boldsymbol{\theta}}) - \mathcal{L}_t^{(1)}(\mathbf{S}_t, \boldsymbol{\theta}_0)| \leq 2K \sup_{\boldsymbol{\theta} \in \Theta} \|\boldsymbol{\theta}\| f_d(\|\mathbf{S}_t\|),$$

and, as a result, taking into account the lower bound of  $\mathcal{L}_t^{(1)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}})$  (for  $\|\hat{\mathbf{S}}_t - \mathbf{S}_t\| \geq \epsilon$ ), we get  $\mathcal{L}_t^{(1)}(\mathbf{S}_t, \hat{\boldsymbol{\theta}}) \leq \frac{1}{5}\mathcal{L}_t^{(1)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}})$ . Next, we have  $\mathcal{L}_t^{(2)}(\mathbf{S}_t, \hat{\boldsymbol{\theta}}) \leq \mathcal{L}_t^{(1)}(\mathbf{S}_t, \hat{\boldsymbol{\theta}}) + \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t}^2$ , and, hence, using the definition of  $\Omega_T^{(1)}$ , the triangle inequality, and the lower bound of  $\mathcal{L}_t^{(1)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}})$ , we get

$$\mathcal{L}_t^{(1)}(\mathbf{S}_t, \hat{\boldsymbol{\theta}}) + \mathcal{L}_t^{(2)}(\mathbf{S}_t, \hat{\boldsymbol{\theta}}) \leq \frac{2}{5}\mathcal{L}_t^{(1)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) + \frac{3}{2}\bar{\phi} \leq \frac{4}{5}\mathcal{L}_t^{(1)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}).$$

Since  $\mathcal{L}_t^{(3)}(\mathbf{S}_t, \hat{\boldsymbol{\theta}}) = 0$  and  $\iota < 1/10$ , we have

$$\mathcal{L}_t^{(1)}(\mathbf{S}_t, \hat{\boldsymbol{\theta}}) + \mathcal{L}_t^{(2)}(\mathbf{S}_t, \hat{\boldsymbol{\theta}}) + \mathcal{L}_t^{(3)}(\mathbf{S}_t, \hat{\boldsymbol{\theta}}) < \mathcal{L}_t^{(1)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) + \mathcal{L}_t^{(2)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) + \mathcal{L}_t^{(3)}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}),$$

which is a contradiction since  $\hat{\mathbf{S}}_t = \hat{\mathbf{S}}_t(\hat{\boldsymbol{\theta}})$  and  $\hat{\mathbf{S}}(\boldsymbol{\theta}) = \operatorname{argmin}_{\mathbf{Z} \in \mathcal{S}^T} \mathcal{L}(\mathbf{Z}, \boldsymbol{\theta})$ . Therefore, with probability approaching one, we must have  $\sup_{t=1,\dots,T} \|\hat{\mathbf{S}}_t - \mathbf{S}_t\| < \epsilon$ . As a result,  $(\hat{\mathbf{S}}'_1, \dots, \hat{\mathbf{S}}'_T, \hat{\boldsymbol{\theta}}')$  is asymptotically equivalent to the solution of the optimization in which each  $\mathbf{Z}_t$  satisfies  $\|\mathbf{S}_t - \mathbf{Z}_t\| < \epsilon$ . For specificity, we denote the latter with  $(\tilde{\mathbf{S}}'_1, \dots, \tilde{\mathbf{S}}'_T, \tilde{\boldsymbol{\theta}}')$  and show next that this solution is consistent.

We start by defining the set

$$\Omega_T^{(5)}(\delta) = \left\{ \omega : \sup_{t=1,\dots,T} \sup_{\|\mathbf{S}_t - \mathbf{Z}_t\| \leq \epsilon} \sup_{\boldsymbol{\theta} \in \Theta} \left| \mathcal{L}_t^{(1)}(\mathbf{Z}_t, \boldsymbol{\theta}) - \sum_{\tau \in \mathcal{T}_t} \frac{N_t^\tau}{N_t} \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} (\kappa(k, \tau, \mathbf{Z}_t, \boldsymbol{\theta}) - \kappa(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0))^2 dk \right| < \delta \right\},$$

and, using A2 and A4, we have Riemann sum convergence and hence  $\mathbb{P}(\Omega_T^{(3)}(\frac{1}{3\delta}) \cap (\Omega_T^{(5)}(\delta))^c) \rightarrow 0$ , for every fixed  $\delta > 0$ .

We next split the set  $\mathbf{Z}_t : \|\mathbf{Z}_t - \mathbf{S}_t\| < \epsilon$  into subsets  $\mathcal{S}_t^i$  satisfying  $\|\mathbf{Z}_t - \mathbf{Z}_t^i\| < \epsilon/T^\iota$  for  $\mathbf{Z}_t, \mathbf{Z}_t^i \in \mathcal{S}_t^i$ . The asymptotic order of the number of these sets,  $N_{\mathcal{S}_t}$ , is  $O(T^{p\iota})$ . Similarly, we split  $\Theta$  (recall this

set is compact) into subsets  $\Theta^i$  satisfying  $\|\boldsymbol{\theta} - \boldsymbol{\theta}'\| < \epsilon/T^\nu$  for  $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta^i$ . The asymptotic order of the number of these sets,  $N_\Theta$ , is  $O(T^{q\iota})$ . Then, using boundedness of  $\phi_{j,t}^{(1)}$ , the bounds in A4, as well as the integrability assumption for  $\mathbf{S}_t$  in A6, we have with

$$\Omega_T^{(6)}(\delta) = \left\{ \omega : \sup_{t=1, \dots, T} \sup_{i=1, \dots, N_\Theta} \sup_{j=1, \dots, N_{\mathbf{S}_t}} \sup_{\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta^i, \mathbf{Z}_t, \mathbf{Z}'_t \in \mathbf{S}_t^j} |\mathcal{L}_t^{(2)}(\mathbf{Z}_t, \boldsymbol{\theta}) - \mathcal{L}_t^{(2)}(\mathbf{Z}'_t, \boldsymbol{\theta}')| < \delta \right\}$$

that  $\mathbb{P}(\Omega_T^{(2)} \cap (\Omega_T^{(6)}(\delta))^c) \rightarrow 0$ , for every fixed  $\delta > 0$ .

Next, for every set  $\mathbf{S}_t^i$ , we pick a point in it and denote this point by  $\mathbf{Z}_t^i$ . Similarly, let  $\boldsymbol{\theta}^i$  denote an arbitrary point that belongs to the set  $\Theta^i$ . With this notation, we introduce

$$\Omega_T^{(7)}(\delta) = \left\{ \omega : \sup_{t=1, \dots, T} \sup_{i=1, \dots, N_\Theta} \sup_{j=1, \dots, N_{\mathbf{S}_t}} |\mathcal{L}_t^{(2)}(\mathbf{Z}_t^j, \boldsymbol{\theta}^i)| < \delta \right\}, \quad \delta > 0.$$

We then have

$$\mathbb{P}\left(\left(\Omega_T^{(7)}(\delta)\right)^c\right) \leq \sum_{t=1}^T \sum_{i=1}^{N_\Theta} \sum_{j=1}^{N_{\mathbf{S}_t}} \mathbb{P}\left(|\mathcal{L}_t^{(2)}(\mathbf{Z}_t^j, \boldsymbol{\theta}^i)| > \delta\right).$$

By applying the Burkholder-Davis-Gundy inequality, using the integrability assumptions for  $\phi_{j,t}^{(k)}$  and arbitrary  $k$  as well as for  $\mathbf{S}_t$  (in A5 and A6), and the fact that every  $\mathbf{Z}_t^j$  has a distance of at most  $\epsilon$  from the true value of the state vector  $\mathbf{S}_t$ , we have

$$\mathbb{P}\left(|\mathcal{L}_t^{(2)}(\mathbf{Z}_t^j, \boldsymbol{\theta}^i)| > \delta\right) \leq K/N^\xi, \quad \text{for some arbitrarily big } \xi > 0,$$

where  $K$  depends on  $\delta$  and  $\xi$ . Then, taking into account the asymptotic orders of  $N_\Theta$  and  $N_{\mathbf{S}_t}$ , we therefore have  $\mathbb{P}\left(\left(\Omega_T^{(7)}(\delta)\right)^c\right) \rightarrow 0$ , for every fixed  $\delta > 0$ .

With the results above, we are ready to establish the consistency of  $(\tilde{\mathbf{S}}'_1, \dots, \tilde{\mathbf{S}}'_T, \tilde{\boldsymbol{\theta}}')$ . Fix some  $\zeta > 0$  arbitrarily small. From A2 and A3(i), we know that on a set with probability approaching one, there is a corresponding  $\delta > 0$ , such that

$$\inf_{(\mathcal{T}_t, \underline{\mathbf{k}}_t, \bar{\mathbf{k}}_t) \in \mathcal{Q}} \inf_{\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \geq \zeta \cup \|\mathbf{Z} - \mathbf{S}\| \geq \zeta} \sum_{\tau \in \mathcal{T}_t} \frac{N_t^\tau}{N_t} \int_{\underline{\mathbf{k}}(t, \tau)}^{\bar{\mathbf{k}}(t, \tau)} \frac{1}{\psi_{t, \tau}(k)} (\kappa(k, \tau, \mathbf{Z}, \boldsymbol{\theta}) - \kappa(k, \tau, \mathbf{S}, \boldsymbol{\theta}_0))^2 dk > \delta f_d(\|\mathbf{S}\|).$$

Let  $\iota \in (0, 1/4)$ , then for some

$$\omega \in \Omega_T^{(2)} \cap \Omega_T^{(3)}\left(\frac{1}{3d}\right) \cap \Omega_T^{(4)}(\delta\iota) \cap \Omega_T^{(5)}(\delta\iota) \cap \Omega_T^{(6)}(\delta\iota) \cap \Omega_T^{(7)}(\delta\iota),$$

if  $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| > \zeta$ , we have  $\mathcal{L}^{(1)}(\mathbf{Z}, \boldsymbol{\theta}) + \mathcal{L}^{(2)}(\mathbf{Z}, \boldsymbol{\theta}) + \mathcal{L}^{(3)}(\mathbf{Z}, \boldsymbol{\theta}) > T\delta(1 - 4\iota) > 0$  by applying the triangle inequality repeatedly over  $t = 1, \dots, T$  while  $\mathcal{L}^{(1)}(\mathbf{S}, \boldsymbol{\theta}_0) + \mathcal{L}^{(2)}(\mathbf{S}, \boldsymbol{\theta}_0) + \mathcal{L}^{(3)}(\mathbf{S}, \boldsymbol{\theta}_0) = 0$ . This shows that  $\tilde{\boldsymbol{\theta}}$  is consistent for  $\boldsymbol{\theta}_0$ .

Now, suppose for some  $t = 1, \dots, T$ , we have  $\|\tilde{\mathbf{S}}_t - \mathbf{S}_t\| > \zeta$ . Then, exactly as before, for

$$\omega \in \Omega_T^{(2)} \cap \Omega_T^{(3)}\left(\frac{1}{3d}\right) \cap \Omega_T^{(4)}(\delta\iota) \cap \Omega_T^{(5)}(\delta\iota) \cap \Omega_T^{(6)}(\delta\iota) \cap \Omega_T^{(7)}(\delta\iota),$$

where  $\iota \in (0, 1/4)$ , we have  $\mathcal{L}_t^{(1)}(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) + \mathcal{L}_t^{(2)}(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) + \mathcal{L}_t^{(3)}(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) > \delta f_d(\|\mathbf{S}_t\|)(1 - 4\iota)$ . On the other hand, given the consistency of  $\tilde{\boldsymbol{\theta}}$  and the second bound in A4, on a set with probability approaching one, we have  $\mathcal{L}_t^{(1)}(\mathbf{S}_t, \tilde{\boldsymbol{\theta}}) + \mathcal{L}_t^{(2)}(\mathbf{S}_t, \tilde{\boldsymbol{\theta}}) + \mathcal{L}_t^{(3)}(\mathbf{S}_t, \tilde{\boldsymbol{\theta}}) \leq 2\mathcal{L}_t^{(1)}(\mathbf{S}_t, \tilde{\boldsymbol{\theta}}) + \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t}^2 < \frac{1}{2} \delta f_d(\|\mathbf{S}_t\|)(1 - 3\iota)$ . This contradicts the definition of  $\tilde{\mathbf{S}}_t$ , and, as a result, we must have  $\|\tilde{\mathbf{S}}_t - \mathbf{S}_t\| \leq \zeta$  for any  $t = 1, \dots, T$ . Therefore, we have consistency of  $(\tilde{\mathbf{S}}'_1, \dots, \tilde{\mathbf{S}}'_T, \tilde{\boldsymbol{\theta}}')$ .  $\square$

### A.3 Definitions and Proof of Theorem 2

Before proceeding to the proof of Theorem 2, we start by providing expressions for the bias terms.

#### A.3.1 Definitions

The six bias terms in Theorem 2 are explicitly defined as:

$$\mathcal{B}_{T,N,1}^{(1)} = \frac{1}{TN} \sum_{t=1}^T \frac{N}{N_t} \frac{1}{N_t} \sum_{j=1}^{N_t} \phi_{j,t} \left( \mathbf{U}_{j,t,0}^{(S)} (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{G}_{j,t,0} \right),$$

$$\mathcal{B}_{T,N,2}^{(1)} = \frac{1}{TN} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{U}_{j,t,0}^{(S)} (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} (\boldsymbol{\Omega}_{S,t} + \boldsymbol{\Psi}_{S,t}) (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{G}_{j,t,0},$$

$$\begin{aligned} \mathcal{B}_{T,N,3}^{(1)} = & \frac{1}{2TN} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{U}_{j,t,0} \times \left\{ \frac{N}{N_t^2} \sum_{\ell=1}^{N_t} \phi_{\ell,t} \left( \mathbf{G}'_{\ell,t,0} (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{G}_{j,t,0}^{(S)} (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{G}_{\ell,t,0} \right) \right. \\ & \left. + 2\lambda^2 V_t^2 \frac{k_n N}{N_t^2} \left( \mathbf{C}'_{t,0} (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{G}'_{j,t,0}^{(S)} (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{C}_{t,0} \right) \right\}, \end{aligned}$$

$$\mathcal{B}_{T,N,1}^{(2)} = \frac{1}{TN} \sum_{t=1}^T 2\lambda^2 V_t^2 \frac{k_n N}{N_t^2} \mathbf{R}'_{t,0} \mathbf{C}_{t,0}^{(S)} (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{C}_{t,0},$$

$$\mathcal{B}_{T,N,2}^{(2)} = \frac{\lambda}{TN} \sum_{t=1}^T \frac{k_n}{N_t} \mathbf{R}'_{t,0} \mathbf{C}_{t,0}^{(S)} (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} (\boldsymbol{\Omega}_{S,t} + \boldsymbol{\Psi}_{S,t}) (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{C}_{t,0},$$

$$\begin{aligned} \mathcal{B}_{T,N,3}^{(2)} = & \frac{\lambda}{2TN} \sum_{t=1}^T \frac{k_n}{N_t} \mathbf{R}'_{t,0} \mathbf{C}_{t,0} \times \left\{ \frac{N}{N_t^2} \sum_{\ell=1}^{N_t} \phi_{\ell,t} \left( \mathbf{G}'_{\ell,t,0} (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{C}_{t,0}^{(S)} (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{G}_{\ell,t,0} \right) \right. \\ & \left. + 2\lambda^2 V_t^2 \frac{k_n N}{N_t^2} \left( \mathbf{C}'_{t,0} (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{C}_{t,0}^{(S)} (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{C}_{t,0} \right) \right\}. \end{aligned}$$

### A.3.2 Proof of Theorem 2

With probability approaching one for  $\boldsymbol{\theta}$  in a sufficiently small neighborhood of  $\boldsymbol{\theta}_0$  and given the proof of Theorem 1,  $\{\hat{\mathbf{S}}_t(\boldsymbol{\theta})\}_{t=1,\dots,T}$  solve the first order condition

$$\begin{aligned} \mathbf{M}_t(\boldsymbol{\theta}) &\equiv \frac{1}{N_t} \sum_{j=1}^{N_t} \left( \tilde{\kappa}_{j,t} - \kappa(k_j, \tau_j, \hat{\mathbf{S}}_t(\boldsymbol{\theta}), \boldsymbol{\theta}) \right) \frac{\partial}{\partial \mathbf{S}} \kappa(k_j, \tau_j, \hat{\mathbf{S}}_t(\boldsymbol{\theta}), \boldsymbol{\theta}) \\ &\quad + \lambda \frac{k_n}{N_t} \times \frac{\partial}{\partial \mathbf{S}} \xi_1(\hat{\mathbf{S}}_t(\boldsymbol{\theta})) \left( \hat{V}_t^n - \xi_1(\hat{\mathbf{S}}_t(\boldsymbol{\theta})) \right) = \mathbf{0}, \quad t = 1, \dots, T. \end{aligned} \quad (47)$$

Henceforth, we use the shorthand convention  $\hat{\kappa}_{j,t}(\boldsymbol{\theta}) = \kappa_j(\hat{\mathbf{S}}_t(\boldsymbol{\theta}), \boldsymbol{\theta})$ . With this in hand, the first-order condition for  $\hat{\boldsymbol{\theta}}$  is given by

$$\sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \left( \tilde{\kappa}_{j,t} - \hat{\kappa}_{j,t}(\hat{\boldsymbol{\theta}}) \right) \frac{\partial}{\partial \boldsymbol{\theta}} \hat{\kappa}_{j,t}(\hat{\boldsymbol{\theta}}) + \lambda \frac{k_n}{N_t} \sum_{t=1}^T \frac{\partial}{\partial \boldsymbol{\theta}} \xi_1(\hat{\mathbf{S}}_t(\hat{\boldsymbol{\theta}})) \left( \hat{V}_t^n - \xi_1(\hat{\mathbf{S}}_t(\hat{\boldsymbol{\theta}})) \right) = \mathbf{0}. \quad (48)$$

A first-order Taylor expansion around the true parameter vector yields  $\tilde{\mathbf{H}}_{T,N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \mathbf{A}$ , where

$$\mathbf{A} = \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \left( \tilde{\kappa}_{j,t} - \hat{\kappa}_{j,t}(\boldsymbol{\theta}_0) \right) \frac{\partial}{\partial \boldsymbol{\theta}} \hat{\kappa}_{j,t}(\boldsymbol{\theta}_0) + \lambda \frac{k_n}{N_t} \frac{1}{T} \sum_{t=1}^T \frac{\partial}{\partial \boldsymbol{\theta}} \xi_1(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \left( \hat{V}_t^n - \xi_1(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \right),$$

and the  $q \times q$  Hessian matrix is given by

$$\begin{aligned} \tilde{\mathbf{H}}_{N,T,n} &= \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \frac{\partial}{\partial \boldsymbol{\theta}} \hat{\kappa}_{j,t}(\tilde{\boldsymbol{\theta}}) \frac{\partial}{\partial \boldsymbol{\theta}'} \hat{\kappa}_{j,t}(\tilde{\boldsymbol{\theta}}) - \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \left( \tilde{\kappa}_{j,t} - \hat{\kappa}_{j,t}(\tilde{\boldsymbol{\theta}}) \right) \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \hat{\kappa}_{j,t}(\tilde{\boldsymbol{\theta}}) \\ &\quad + \lambda \frac{k_n}{N_t} \frac{1}{T} \sum_{t=1}^T \frac{\partial}{\partial \boldsymbol{\theta}} \xi_1(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})) \frac{\partial}{\partial \boldsymbol{\theta}'} \xi_1(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})) - \lambda \frac{k_n}{N_t} \frac{1}{T} \sum_{t=1}^T \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \xi_1(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})) \left( \hat{V}_t^n - \xi_1(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})) \right) \\ &\equiv \tilde{\mathbf{H}}_{N,T,n}^{(1)} - \tilde{\mathbf{H}}_{N,T,n}^{(2)} + \tilde{\mathbf{H}}_{N,T,n}^{(3)} - \tilde{\mathbf{H}}_{N,T,n}^{(4)}, \end{aligned}$$

for some  $\tilde{\boldsymbol{\theta}} \in (\boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}})$ . Next, write the derivatives in the  $q \times 1$  vector  $\mathbf{A}$  as

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}} \hat{\kappa}_{j,t}(\boldsymbol{\theta}_0) &= \mathbf{U}_{j,t}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) + \left( \frac{\partial \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \mathbf{G}_j(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0), \\ \frac{\partial}{\partial \boldsymbol{\theta}} \xi_1(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)) &= \mathbf{R}'_{t,0} \mathbf{C}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)) + \left( \frac{\partial \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \mathbf{C}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)), \end{aligned} \quad (49)$$

which are used to make the decomposition  $\mathbf{A} \equiv \mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3 + \mathbf{A}_4$ , where

$$\begin{aligned} \mathbf{A}_1 &\equiv \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{U}_{j,t}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0), \quad \mathbf{A}_2 \equiv \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} (\kappa_j(\mathbf{S}_t, \boldsymbol{\theta}_0) - \hat{\kappa}_{j,t}(\boldsymbol{\theta}_0)) \mathbf{U}_{j,t}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0), \\ \mathbf{A}_3 &\equiv \lambda \frac{k_n}{N_t} \frac{1}{T} \sum_{t=1}^T \left( \mathbf{R}'_{t,0} \mathbf{C}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \right) \left( \hat{V}_t^n - \xi_1(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \right), \end{aligned}$$

$$\begin{aligned} \mathbf{A}_4 &\equiv \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \left\{ \epsilon_{j,t} + \left( \kappa_j(\mathbf{S}_t, \boldsymbol{\theta}_0) - \kappa_j(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) \right) \right\} \left( \frac{\partial \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \mathbf{G}_j(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) \\ &\quad + \lambda \frac{k_n}{N_t T} \sum_{t=1}^T \left( \frac{\partial \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \mathbf{C}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \left( \hat{V}_t^n - \xi_1(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \right). \end{aligned}$$

To decompose  $\mathbf{A}_1$  even further, use a third-order Taylor expansion to write

$$\begin{aligned} \mathbf{U}_{j,t}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) &= \mathbf{U}_{j,t,0} + \mathbf{U}_{j,t,0}^{(S)}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) + \frac{1}{2} \mathbf{U}_{j,t,0}^{(SS)} \times \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \quad (50) \\ &\quad + \frac{1}{6} \left\{ \frac{\partial}{\partial \mathbf{S}'} \mathbf{U}_{j,t}^{(SS)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) \right\} \times \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\}, \end{aligned}$$

for some  $\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0) \in (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \mathbf{S}_t)$  and  $t = 1, \dots, T$ , utilizing that the function  $\kappa(\cdot)$  is four times continuously differentiable in its arguments by A7. Hence, inserting (50) into  $\mathbf{A}_1$ ,

$$\sqrt{TN} \mathbf{A}_1 = \sqrt{TN} \mathbf{A}_{11} + \sqrt{TN} \mathbf{A}_{12} + \sqrt{TN} \mathbf{A}_{13} \quad (51)$$

where the right-hand-side terms are defined as

$$\begin{aligned} \mathbf{A}_{11} &= \frac{1}{TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{U}_{j,t,0}, \quad \mathbf{A}_{12} = \frac{1}{TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{U}_{j,t,0}^{(S)}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t), \\ \mathbf{A}_{13} &= \frac{1}{2TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{U}_{j,t,0}^{(SS)} \times \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \\ &\quad + \frac{1}{6TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \left\{ \frac{\partial \mathbf{U}_{j,t}^{(SS)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)}{\partial \mathbf{S}'} \right\} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\}. \end{aligned}$$

Moreover, for  $\mathbf{A}_2$ , we make an additional third-order Taylor expansion in the state vector,

$$\begin{aligned} \kappa_j(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) &= \kappa_j(\mathbf{S}_t, \boldsymbol{\theta}_0) + \mathbf{G}'_{j,t,0}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) + (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' \mathbf{G}_{j,t,0}^{(S)}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)/2 \\ &\quad + (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' \left\{ \frac{\partial}{\partial \mathbf{S}'} \mathbf{G}_j^{(S)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) \right\} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} / 6, \end{aligned}$$

where, again,  $\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0) \in (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \mathbf{S}_t)$ , and we use it together with (50) to make a similar decomposition,

$$\sqrt{TN} \mathbf{A}_2 = -\sqrt{TN} \mathbf{A}_{21} - \sqrt{TN} \mathbf{A}_{22} - \sqrt{TN} \mathbf{A}_{23}, \quad (52)$$

with the right-hand-side terms defined as

$$\begin{aligned} \mathbf{A}_{21} &= \frac{1}{TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \left\{ \mathbf{G}'_{j,t,0}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \mathbf{U}_{j,t,0}, \\ \mathbf{A}_{22} &= \frac{1}{TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \left\{ \mathbf{G}'_{j,t,0}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \mathbf{U}_{j,t,0}^{(S)}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{2TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' \mathbf{G}_{j,t,0}^{(S)} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \mathbf{U}_{j,t,0}, \\
\mathbf{A}_{23} & = \frac{1}{2TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \left\{ \mathbf{G}_j(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)' (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \mathbf{U}_{j,t,0}^{(SS)} \times \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \\
& + \frac{1}{6TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \left\{ \mathbf{G}_j(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)' (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \times \left\{ \frac{\partial \mathbf{U}_{j,t}^{(SS)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)}{\partial \mathbf{S}'} \right\} \\
& \quad \times \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \\
& + \frac{1}{6TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \mathbf{U}_{j,t,0} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' \left\{ \frac{\partial \mathbf{G}_j^{(S)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)}{\partial \mathbf{S}'} \right\} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \\
& + \frac{1}{2TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' \mathbf{G}_{j,t,0}^{(S)} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \mathbf{U}_{j,t,0}^{(S)} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \\
& + \frac{1}{6TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \left\{ \frac{\partial}{\partial \mathbf{S}'} \mathbf{G}_j^{(S)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) \right\} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \right\} \\
& \quad \times \mathbf{U}_{j,t,0}^{(S)} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t),
\end{aligned}$$

where a corresponding first-order Taylor expansion,

$$\kappa_j(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) = \kappa_j(\mathbf{S}_t, \boldsymbol{\theta}_0) + \mathbf{G}_j(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)' (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t),$$

have been applied for some terms in  $\mathbf{A}_{23}$ .

For  $\mathbf{A}_3$ , we apply a third-order Taylor expansion and the mean-value theorem to decompose

$$\begin{aligned}
\xi_1(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)) & = \xi_1(\mathbf{S}_t) + \mathbf{C}'(\mathbf{S}_t)' (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) + (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' \mathbf{C}^{(S)}(\mathbf{S}_t) (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) / 2 \\
& \quad + (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' \left\{ \frac{\partial}{\partial \mathbf{S}'} \mathbf{C}^{(S)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \right\} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} / 6,
\end{aligned}$$

as well as a second-order Taylor expansion of  $\mathbf{C}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0))$ ,

$$\begin{aligned}
\mathbf{C}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)) & = \mathbf{C}_{t,0} + \mathbf{C}_{t,0}^{(S)} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \\
& \quad + \left\{ \frac{\partial}{\partial \mathbf{S}'} \mathbf{C}^{(S)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \right\} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} / 2,
\end{aligned}$$

which leads to the following decomposition

$$\sqrt{TN} \mathbf{A}_3 = \sqrt{TN} \mathbf{A}_{31} - \sqrt{TN} \mathbf{A}_{32} + \sqrt{TN} \mathbf{A}_{33} - \sqrt{TN} \mathbf{A}_{34}, \quad (53)$$

where the right-hand-side terms are defined as

$$\mathbf{A}_{31} = \frac{\lambda}{T} \sum_{t=1}^T \frac{k_n}{N_t} \mathbf{R}'_{t,0} \mathbf{C}_{t,0} (\hat{\mathbf{V}}_t^n - \mathbf{V}_t), \quad \mathbf{A}_{32} = \frac{\lambda}{T} \sum_{t=1}^T \frac{k_n}{N_t} \mathbf{R}'_{t,0} \mathbf{C}_{t,0} (\mathbf{C}'_{t,0} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)),$$

$$\begin{aligned}
\mathbf{A}_{33} &= \frac{\lambda}{T} \sum_{t=1}^T \frac{k_n}{N_t} \mathbf{R}'_{t,0} \mathbf{C}_{t,0}^{(S)} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \left( \hat{V}_t^n - \xi_1(\mathbf{S}_t) \right) \\
&\quad - \frac{\lambda}{T} \sum_{t=1}^T \frac{k_n}{N_t} \mathbf{R}'_{t,0} \mathbf{C}_{t,0}^{(S)} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \left( \mathbf{C}'_{t,0} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right) \\
&\quad - \frac{\lambda}{2T} \sum_{t=1}^T \frac{k_n}{N_t} \mathbf{R}'_{t,0} \mathbf{C}_{t,0} \left( (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' \mathbf{C}_{t,0}^{(S)} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right), \\
\mathbf{A}_{34} &= \frac{\lambda}{6T} \sum_{t=1}^T \frac{k_n}{N_t} \mathbf{R}'_{t,0} \mathbf{C}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \\
&\quad \times \left( (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' \left\{ \frac{\partial \mathbf{C}^{(S)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0))}{\partial \mathbf{S}'} \right\} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \right) \\
&\quad - \frac{\lambda}{2T} \sum_{t=1}^T \frac{k_n}{N_t} \left( \mathbf{R}'_{t,0} \left\{ \frac{\partial \mathbf{C}^{(S)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0))}{\partial \mathbf{S}'} \right\} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \right) \left( \hat{V}_t^n - \xi_1(\mathbf{S}_t) \right) \\
&\quad + \frac{\lambda}{2T} \sum_{t=1}^T \frac{k_n}{N_t} \left( \mathbf{R}'_{t,0} \left\{ \frac{\partial \mathbf{C}^{(S)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0))}{\partial \mathbf{S}'} \right\} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \right) \left( \mathbf{C}'_{t,0} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right) \\
&\quad + \frac{\lambda}{2T} \sum_{t=1}^T \frac{k_n}{N_t} \mathbf{R}'_{t,0} \left( \mathbf{C}_{t,0}^{(S)} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right) \left( (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' \mathbf{C}_{t,0}^{(S)} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right) \\
&\quad + \frac{\lambda}{4T} \sum_{t=1}^T \frac{k_n}{N_t} \left( \mathbf{R}'_{t,0} \left\{ \frac{\partial \mathbf{C}^{(S)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0))}{\partial \mathbf{S}'} \right\} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \right) \\
&\quad \times \left( (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' \mathbf{C}_{t,0}^{(S)} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right).
\end{aligned}$$

We note that the Taylor expansion of the gradient  $\mathbf{C}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0))$  has not been inserted into the first term in  $\mathbf{A}_{34}$  involving  $\partial \mathbf{C}^{(S)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0))/\partial \mathbf{S}'$ , since this will be shown below to be asymptotically negligible.

Now, we add  $\mathbf{A}_{21}$  and  $\mathbf{A}_{32}$  and define the resulting sum as

$$\mathbf{A}_5 \equiv \mathbf{A}_{21} + \mathbf{A}_{32} = \frac{1}{T} \sum_{t=1}^T \left\{ \frac{1}{N_t} \sum_{j=1}^{N_t} (\mathbf{D}_{j,t,0} + \mathbf{R}'_{t,0} \mathbf{G}_{j,t,0}) \mathbf{G}'_{j,t,0} + \lambda \frac{k_n}{N_t} \mathbf{R}'_{t,0} \mathbf{C}_{t,0} \mathbf{C}'_{t,0} \right\} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t).$$

With the above decompositions, we are ready to proceed with the proof, which is organized as a sequence of lemmas. We first show the convergence in probability of the Hessian  $\tilde{\mathbf{H}}_{N,T,n}$  in Lemma 2. Second, we have a CLT for  $\mathbf{A}_{11} + \mathbf{A}_{31}$  by an application of Lemma 7 in conjunction with the ergodicity assumption for  $\mathbf{S}_t$  and the convergence in probability results in A6. Third, we show in Lemma 3 that  $\mathbf{A}_{12} - \mathbf{A}_{22} + \mathbf{A}_{33}$  converges to the asymptotic bias (when appropriately scaled). In Lemmas 4-6, we show the asymptotic negligibility of  $\mathbf{A}_{13}$ ,  $\mathbf{A}_{23}$ ,  $\mathbf{A}_{34}$ ,  $\mathbf{A}_4$  and  $\mathbf{A}_5$ . Finally, we apply Lemma 13 to show joint convergence of the parameter and state vectors.  $\square$

In the lemmas below, we will make use of the following additional notation:

$$\begin{aligned}
\mathbf{F}_{t,N,n} &\equiv \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{G}_{j,t,0} + \lambda \frac{k_n}{N_t} \mathbf{C}_{t,0} \left( \hat{V}_t^n - V_t \right) \equiv \mathbf{F}_{t,N,n}^{(1)} + \mathbf{F}_{t,N,n}^{(2)}, \\
\hat{\mathcal{I}}_{S,t}(\mathbf{Z}, \boldsymbol{\theta}) &\equiv \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{G}_j(\mathbf{Z}, \boldsymbol{\theta}) \mathbf{G}_j(\mathbf{Z}, \boldsymbol{\theta})' - \frac{1}{N_t} \sum_{j=1}^{N_t} (\tilde{\kappa}_{j,t} - \kappa_j(\mathbf{Z}, \boldsymbol{\theta})) \mathbf{G}_j^{(S)}(\mathbf{Z}, \boldsymbol{\theta}), \\
\hat{\mathcal{P}}_{S,t}(\mathbf{Z}) &\equiv \lambda \frac{k_n}{N_t} \mathbf{C}(\mathbf{Z}) \mathbf{C}(\mathbf{Z})' - \lambda \frac{k_n}{N_t} \mathbf{C}^{(S)}(\mathbf{Z}) \left( \hat{V}_t^n - \xi_1(\mathbf{Z}) \right), \\
\hat{\mathcal{I}}_{S,\theta,t}(\mathbf{Z}, \boldsymbol{\theta}) &\equiv \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{G}_j(\mathbf{Z}, \boldsymbol{\theta}) \mathbf{D}_j(\mathbf{Z}, \boldsymbol{\theta})' - \frac{1}{N_t} \sum_{j=1}^{N_t} (\tilde{\kappa}_{j,t} - \kappa_j(\mathbf{Z}, \boldsymbol{\theta})) \mathbf{G}_j^{(\theta)}(\mathbf{Z}, \boldsymbol{\theta}).
\end{aligned}$$

With this notation, we use a first-order Taylor expansion to write

$$\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t = \left( \hat{\mathcal{I}}_{S,t}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) + \hat{\mathcal{P}}_{S,t}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \right)^{-1} \mathbf{F}_{t,N,n}, \quad t = 1, \dots, T,$$

for some  $\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0) \in (\mathbf{S}_t, \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0))$ , provided that the inverse above exists. Finally, we will also use the following simplifying notation

$$\mathcal{L}_{S,t} = \mathcal{I}_{S,t} + \mathcal{P}_{S,t}, \quad \tilde{\mathcal{L}}_{S,t} = \hat{\mathcal{I}}_{S,t}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) + \hat{\mathcal{P}}_{S,t}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0)). \quad (54)$$

**Lemma 2.** *Under the conditions of Theorem 2, we have*

$$\| \tilde{\mathbf{H}}_{N,T,n} - \mathcal{I}_U - \mathcal{P}_\theta \| \xrightarrow{\mathbb{P}} 0.$$

*Proof.* The probability limit for the Hessian is established by considering each term from the decomposition  $\tilde{\mathbf{H}}_{N,T,n} = \tilde{\mathbf{H}}_{N,T,n}^{(1)} - \tilde{\mathbf{H}}_{N,T,n}^{(2)} + \tilde{\mathbf{H}}_{N,T,n}^{(3)} - \tilde{\mathbf{H}}_{N,T,n}^{(4)}$ . First, for  $\tilde{\mathbf{H}}_{N,T,n}^{(1)}$ , we have

$$\frac{\partial}{\partial \boldsymbol{\theta}} \hat{\kappa}_{j,t}(\tilde{\boldsymbol{\theta}}) = \mathbf{U}_{j,t}(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}}), \tilde{\boldsymbol{\theta}}) + \left( \frac{\partial \hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \mathbf{G}_j(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}}), \tilde{\boldsymbol{\theta}}),$$

similarly to (49). Now, using A7, Theorem 1 and Lemma 9 (c), we have

$$\tilde{\mathbf{H}}_{N,T,n}^{(1)} - \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{U}_{j,t,0} \mathbf{U}_{j,t,0}' = o_p(1).$$

Moreover, applying A2 and A7, it follows that

$$\frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{U}_{j,t,0} \mathbf{U}_{j,t,0}' - \frac{1}{T} \sum_{t=1}^T \mathcal{I}_{U,t} = o_p(1),$$

and, from here, using the convergence in probability results of A6, we finally have

$$\tilde{\mathbf{H}}_{N,T,n}^{(1)} = \mathcal{I}_U + o_p(1). \quad (55)$$



Next, for  $\tilde{\mathbf{H}}_{N,T,n}^{(2)}$ , we make the decomposition,

$$\begin{aligned}\tilde{\mathbf{H}}_{N,T,n}^{(2)} &= \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \left( \kappa_j(\mathbf{S}_t, \boldsymbol{\theta}_0) - \hat{\kappa}_{j,t}(\tilde{\boldsymbol{\theta}}) \right) \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \hat{\kappa}_{j,t}(\tilde{\boldsymbol{\theta}}) \\ &\quad + \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \left( \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \hat{\kappa}_{j,t}(\tilde{\boldsymbol{\theta}}) - \overline{\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \hat{\kappa}_{j,t}(\boldsymbol{\theta}_0)} \right) \\ &\quad + \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \overline{\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \hat{\kappa}_{j,t}(\boldsymbol{\theta}_0)} \equiv \tilde{\mathbf{H}}_{N,T,n,1}^{(2)} + \tilde{\mathbf{H}}_{N,T,n,2}^{(2)} + \tilde{\mathbf{H}}_{N,T,n,3}^{(2)},\end{aligned}$$

where  $\overline{\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \hat{\kappa}_{j,t}(\boldsymbol{\theta}_0)}$  denotes the derivative  $\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \hat{\kappa}_{j,t}(\boldsymbol{\theta}_0)$  in which  $\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)$  is replaced with  $\mathbf{S}_t$ . This term contains a second-order derivative of  $\hat{\mathbf{S}}(\boldsymbol{\theta})$ , which can be evaluated using implicit function differentiation. Then, using A7 and Lemmas 9 (a) and (c), we have

$$\left\| \overline{\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \hat{\kappa}_{j,t}(\boldsymbol{\theta}_0)} \right\| \leq K f_d(\|\mathbf{S}_t\|), \quad \left\| \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \hat{\kappa}_{j,t}(\tilde{\boldsymbol{\theta}}) - \overline{\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \hat{\kappa}_{j,t}(\boldsymbol{\theta}_0)} \right\| \leq K \hat{\eta}_{t,N,n} f_d(\|\mathbf{S}_t\|),$$

where the nonnegative process  $\hat{\eta}_{t,N,n}$  satisfies  $\sup_{t=1,\dots,T} \hat{\eta}_{t,N,n} = o_p(1)$ . From here, using A6 and A7 as well as the uniform consistency  $\sup_{t=1,\dots,T} \|\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}}) - \mathbf{S}_t\| = o_p(1)$  and  $\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = o_p(1)$ , we get  $\|\tilde{\mathbf{H}}_{N,T,n,1}^{(2)}\| = o_p(1)$ . Similarly, using the second bound above and since  $\frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} |\epsilon_{j,t}| f_d(\|\mathbf{S}_t\|)$  is  $O_p(1)$  (by applying A5 and A6), we have  $\|\tilde{\mathbf{H}}_{N,T,n,2}^{(2)}\| = o_p(1)$ . Finally, for the third term,  $\tilde{\mathbf{H}}_{N,T,n,3}^{(2)}$ , we may use part (a) of Lemma 8 and readily conclude that  $\|\tilde{\mathbf{H}}_{N,T,n,3}^{(2)}\| = o_p(1)$ .

Next, for  $\tilde{\mathbf{H}}_{N,T,n}^{(3)}$ , we write

$$\frac{\partial}{\partial \boldsymbol{\theta}} \xi_1(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})) = \mathbf{R}'_{t,0} \mathbf{C}(\mathbf{S}_t) + \mathbf{R}'_{t,0} (\mathbf{C}(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})) - \mathbf{C}(\mathbf{S}_t)) + \left( \frac{\partial \hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \mathbf{C}(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})).$$

Then, using arguments similar to those for the term  $\tilde{\mathbf{H}}_{N,T,n}^{(1)}$ , we have

$$\tilde{\mathbf{H}}_{N,T,n}^{(3)} - \frac{1}{T} \sum_{t=1}^T \mathcal{P}_{\theta,t} = o_p(1), \quad \frac{1}{T} \sum_{t=1}^T \mathcal{P}_{\theta,t} = \mathcal{P}_{\theta} + o_p(1),$$

and, as a result,  $\tilde{\mathbf{H}}_{N,T,n}^{(3)} = \mathcal{P}_{\theta} + o_p(1)$ .

For the last term,  $\tilde{\mathbf{H}}_{N,T,n}^{(4)}$ , make the decomposition,

$$\begin{aligned}\tilde{\mathbf{H}}_{N,T,n}^{(4)} &= \lambda \frac{k_n}{N_t} \frac{1}{T} \sum_{t=1}^T \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \xi_1(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})) \left( \hat{V}_t^n - \xi_1(\mathbf{S}_t) \right) \\ &\quad + \lambda \frac{k_n}{N_t} \frac{1}{T} \sum_{t=1}^T \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \xi_1(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})) \left( \xi_1(\mathbf{S}_t) - \xi_1(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})) \right) \equiv \tilde{\mathbf{H}}_{N,T,n,1}^{(4)} + \tilde{\mathbf{H}}_{N,T,n,2}^{(4)}.\end{aligned}$$

We then use A7, the uniform consistency  $\sup_{t=1,\dots,T} \|\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}}) - \mathbf{S}_t\| = o_p(1)$  and  $\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = o_p(1)$  as well as part (b) of Lemma 8, to conclude  $\|\tilde{\mathbf{H}}_{N,T,n,1}^{(4)}\| = o_p(1)$  and  $\|\tilde{\mathbf{H}}_{N,T,n,2}^{(4)}\| = o_p(1)$ .  $\square$

**Lemma 3.** *Under the conditions of Theorem 2, we have*

$$\sqrt{TN} \|\mathbf{A}_{12} - \mathbf{A}_{22} + \mathbf{A}_{33} - (\mathcal{I}_U + \mathcal{P}_\theta) \mathbf{B}_{T,N}\| \xrightarrow{\mathbb{P}} 0.$$

*Proof.* First, recall that

$$(\mathcal{I}_U + \mathcal{P}_\theta) \mathbf{B}_{T,N} = \sum_{j=1}^2 \left( \mathbf{B}_{T,N,1}^{(j)} - \mathbf{B}_{T,N,2}^{(j)} - \mathbf{B}_{T,N,3}^{(j)} \right).$$

Then, using a first-order Taylor expansion, we may write

$$\begin{aligned} \mathbf{A}_{12} &= \frac{1}{TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{U}_{j,t,0}^{(S)} \mathcal{L}_{S,t}^{-1} \mathbf{F}_{t,N,n} \\ &\quad + \frac{1}{TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{U}_{j,t,0}^{(S)} \left( \tilde{\mathcal{L}}_{S,t}^{-1} - \mathcal{L}_{S,t}^{-1} \right) \mathbf{F}_{t,N,n} \equiv \mathbf{A}_{121} + \mathbf{A}_{122}. \end{aligned}$$

From part (a) of Lemma 9,  $\|\tilde{\mathcal{L}}_{S,t}^{-1} - \mathcal{L}_{S,t}^{-1}\| \leq K \hat{\eta}_{t,N,n} \times f_d(\|\mathbf{S}_t\|)$  where the nonnegative process  $\hat{\eta}_{t,N,n}$  satisfies  $\sup_{t=1,\dots,T} \hat{\eta}_{t,N,n} = o_p(1)$ . Therefore, for the second term in the decomposition, using the triangle inequality as well as Lemma 9 (a) and Lemma 10, we have

$$\sqrt{TN} \|\mathbf{A}_{122}\| \leq \frac{K}{\sqrt{TN}} \sup_{t=1,\dots,T} |\hat{\eta}_{t,N,n}| \sum_{t=1}^T \left\| \frac{1}{\sqrt{N_t}} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{U}_{j,t,0}^{(S)} \right\| \times \|\sqrt{N} \mathbf{F}_{t,N,n}\| \times f_d(\|\mathbf{S}_t\|) \xrightarrow{\mathbb{P}} 0.$$

For the first term,  $\mathbf{A}_{121}$ , we write

$$\begin{aligned} \mathbf{A}_{121} &= \frac{1}{TN} \sum_{t=1}^T \frac{N}{N_t^2} \sum_{j=1}^{N_t} \phi_{j,t} \mathbf{U}_{j,t,0}^{(S)} \mathcal{L}_{S,t}^{-1} \mathbf{G}_{j,t,0} + \frac{1}{TN} \sum_{t=1}^T \frac{N}{N_t^2} \sum_{j=1}^{N_t} (\epsilon_{j,t}^2 - \phi_{j,t}) \mathbf{U}_{j,t,0}^{(S)} \mathcal{L}_{S,t}^{-1} \mathbf{G}_{j,t,0} \\ &\quad + \frac{1}{TN} \sum_{t=1}^T \frac{N}{N_t^2} \sum_{j,\ell=1; j \neq \ell}^{N_t} \epsilon_{j,t} \epsilon_{\ell,t} \mathbf{U}_{j,t,0}^{(S)} \mathcal{L}_{S,t}^{-1} \mathbf{G}_{\ell,t,0} \\ &\quad + \frac{\lambda}{TN} \sum_{t=1}^T \frac{k_n}{N_t} \frac{N}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{U}_{j,t,0}^{(S)} \times \left\{ \mathcal{L}_{S,t}^{-1} \mathbf{C}_{t,0} \left( \hat{V}_t^n - \xi_1(\mathbf{S}_t) \right) \right\} \\ &\equiv \mathbf{B}_{T,N,1}^{(1)} + \mathbf{A}_{1211} + \mathbf{A}_{1212} + \mathbf{A}_{1213}. \end{aligned}$$

Now, since  $\mathbf{U}_{j,t,0}^{(S)}$ ,  $\mathcal{I}_{S,t}$ ,  $\mathcal{P}_{S,t}$  and  $\mathbf{G}_{\ell,t,0}$  are all adapted to  $\mathcal{F}_t^{(0)}$ , and since the noise satisfies the regularity conditions in A5, we have  $\mathbb{E}_0(\mathbf{A}_{1211}) = \mathbf{0}$  and  $\mathbb{E}_0(\mathbf{A}_{1212}) = \mathbf{0}$ . Moreover, using, again, A5 and A6 as well as successive conditioning, we can write

$$\mathbb{V}_0 \left( \sqrt{TN} \mathbf{A}_{1211} \right) = \frac{1}{TN} \sum_{t=1}^T \left( \frac{N}{N_t} \right)^2 \frac{1}{N_t^2} \sum_{j=1}^{N_t} \mathbb{E}_0 \left( (\epsilon_{j,t}^2 - \phi_{j,t})^2 \right) \left( \mathbf{U}_{j,t,0}^{(S)} \mathcal{L}_{S,t}^{-1} \mathbf{G}_{j,t,0} \right) \left( \mathbf{U}_{j,t,0}^{(S)} \mathcal{L}_{S,t}^{-1} \mathbf{G}_{j,t,0} \right)'$$

$$\leq \frac{K}{TN} \sum_{t=1}^T \left(\frac{N}{N_t}\right)^2 \frac{1}{N_t^2} \sum_{j=1}^{N_t} (\phi_{j,t}^{(4)} + \phi_{j,t}^2) \left(\mathbf{U}_{j,t,0}^{(S)} \boldsymbol{\mathcal{L}}_{S,t}^{-1} \mathbf{G}_{j,t,0}\right) \left(\mathbf{U}_{j,t,0}^{(S)} \boldsymbol{\mathcal{L}}_{S,t}^{-1} \mathbf{G}_{j,t,0}\right)' = O_p(N^{-2}),$$

where the last equality follows from the assumed integrability properties of the  $(\mathcal{F}^{(0)})$ -conditional moments of the observation error and the state vector. Similar arguments provide

$$\begin{aligned} \mathbb{V}_0\left(\sqrt{TN} \mathbf{A}_{1212}\right) &= \frac{1}{TN} \sum_{t=1}^T \left(\frac{N}{N_t}\right)^2 \frac{1}{N_t^2} \sum_{j,\ell=1; j \neq \ell}^{N_t} \mathbb{E}_0\left(\epsilon_{j,t}^2 \epsilon_{\ell,t}^2\right) \left(\mathbf{U}_{j,t,0}^{(S)} \boldsymbol{\mathcal{L}}_{S,t}^{-1} \mathbf{G}_{\ell,t,0}\right) \left(\mathbf{U}_{\ell,t,0}^{(S)} \boldsymbol{\mathcal{L}}_{S,t}^{-1} \mathbf{G}_{j,t,0}\right)' \\ &= \frac{1}{TN} \sum_{t=1}^T \left(\frac{N}{N_t}\right)^2 \frac{1}{N_t^2} \sum_{j,\ell=1; j \neq \ell}^{N_t} \phi_{j,t} \phi_{\ell,t} \left(\mathbf{U}_{j,t,0}^{(S)} \boldsymbol{\mathcal{L}}_{S,t}^{-1} \mathbf{G}_{\ell,t,0}\right) \left(\mathbf{U}_{\ell,t,0}^{(S)} \boldsymbol{\mathcal{L}}_{S,t}^{-1} \mathbf{G}_{j,t,0}\right)' = O_p(N^{-1}). \end{aligned}$$

These conditional moment results imply the following stochastic bounds,  $\|\mathbf{A}_{1211}\| = O_p(N^{-1})$  and  $\|\mathbf{A}_{1212}\| = O_p(N^{-1/2})$ . For  $\mathbf{A}_{1213}$ , we can use successive conditioning and A5 for the noise to get

$$\mathbb{V}_0\left(\sqrt{TN} \mathbf{A}_{1213}\right) \leq K \mathbb{E}_0 \left( \frac{N}{T} \sum_{t=1}^T \left\| \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{U}_{j,t,0}^{(S)} \right\|^2 \times \left\| \frac{k_n}{N} \boldsymbol{\mathcal{L}}_{S,t}^{-1} \mathbf{C}_{t,0} \left( \hat{V}_t^n - \xi_1(\mathbf{S}_t) \right) \right\|^2 \right).$$

Now, we apply Lemma 10 with  $r_1 = 0$ ,  $r_2 = 2$ ,  $r_3 = 2$  (taking into account that  $k_n^2 \Delta_n \rightarrow 0$ ), A6 and A7, to get  $\mathbb{V}_0(\sqrt{TN} \mathbf{A}_{1213}) = O_p(1/N)$  and hence  $\sqrt{TN} \|\mathbf{A}_{1213}\| \rightarrow \mathbf{0}$ . Altogether,

$$\sqrt{TN} \left\| \mathbf{A}_{12} - \mathbf{B}_{T,N,1}^{(1)} \right\| \xrightarrow{\mathbb{P}} 0.$$

Turning to the second main term,  $\mathbf{A}_{22}$ , we define the two terms in its definition as  $\mathbf{A}_{221}$  and  $\mathbf{A}_{222}$  (i.e., we have  $\mathbf{A}_{22} \equiv \mathbf{A}_{221} + \mathbf{A}_{222}$ ), and we further decompose  $\mathbf{A}_{221}$  using addition and subtraction as

$$\begin{aligned} \mathbf{A}_{221} &= \frac{1}{TN} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{U}_{j,t,0}^{(S)} \left\{ \boldsymbol{\mathcal{L}}_{S,t}^{-1} (\boldsymbol{\Omega}_{S,t} + \boldsymbol{\Psi}_{S,t}) \boldsymbol{\mathcal{L}}_{S,t}^{-1} \right\} \mathbf{G}_{j,t,0} \\ &\quad + \frac{1}{TN} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{U}_{j,t,0}^{(S)} \left\{ N(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' - \boldsymbol{\mathcal{L}}_{S,t}^{-1} (\boldsymbol{\Omega}_{S,t} + \boldsymbol{\Psi}_{S,t}) \boldsymbol{\mathcal{L}}_{S,t}^{-1} \right\} \mathbf{G}_{j,t,0} \\ &\equiv \mathbf{B}_{T,N,2}^{(1)} + \mathbf{A}_{2211}. \end{aligned}$$

We have  $\|(TN)^{1/2} \mathbf{A}_{2211}\| = o_p(\sqrt{T/N})$  by making use of the integrability assumptions for the state vector in A6 and A7, and then applying Lemma 12. For  $\mathbf{A}_{222}$ , apply a first-order Taylor expansion as well as further addition and subtraction to make the decomposition,

$$\begin{aligned} \mathbf{A}_{222} &= \frac{1}{2TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \left\{ \mathbf{F}'_{t,N,n} \boldsymbol{\mathcal{L}}_{S,t}^{-1} \mathbf{G}_{j,t,0}^{(S)} \boldsymbol{\mathcal{L}}_{S,t}^{-1} \mathbf{F}_{t,Nn} \right\} \mathbf{U}_{j,t,0} \\ &\quad + \frac{1}{2TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \left\{ \mathbf{F}'_{t,N,n} \left( \tilde{\boldsymbol{\mathcal{L}}}_{S,t}^{-1} - \boldsymbol{\mathcal{L}}_{S,t}^{-1} \right) \mathbf{G}_{j,t,0}^{(S)} \boldsymbol{\mathcal{L}}_{S,t}^{-1} \mathbf{F}_{t,Nn} \right\} \mathbf{U}_{j,t,0} \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{2TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \left\{ \mathbf{F}'_{t,N,n} \tilde{\mathcal{L}}_{S,t}^{-1} \mathbf{G}_{j,t,0}^{(S)} \left( \tilde{\mathcal{L}}_{S,t}^{-1} - \mathcal{L}_{S,t}^{-1} \right) \mathbf{F}_{t,N,n} \right\} \mathbf{U}_{j,t,0} \\
& \equiv \mathbf{A}_{2221} + \mathbf{A}_{2222} + \mathbf{A}_{2223},
\end{aligned}$$

whose second term,  $\mathbf{A}_{2222}$ , may be bounded by the triangle inequality and Lemma 9 (a) as

$$\sqrt{TN} \|\mathbf{A}_{2222}\| \leq \frac{K}{\sqrt{TN}} \sup_{t=1, \dots, T} |\hat{\eta}_{t,N,n}| \sum_{t=1}^T \left\| \sqrt{N} \mathbf{F}_{t,N,n} \right\|^2 f_d(\|\mathbf{S}_t\|) \left\{ \frac{1}{N_t} \sum_{j=1}^{N_t} \left\| \mathbf{G}_{j,t,0}^{(S)} \mathcal{L}_{S,t}^{-1} \right\| \times \|\mathbf{U}_{j,t,0}\| \right\}.$$

Then, using part (a) of Lemma 9 and Lemma 10, we have  $\sqrt{TN} \|\mathbf{A}_{2222}\| \xrightarrow{\mathbb{P}} 0$ , and similar arguments may be used to show  $\sqrt{TN} \|\mathbf{A}_{2223}\| \xrightarrow{\mathbb{P}} 0$ . Returning to the first term in the decomposition of  $\mathbf{A}_{222}$ , that is, to  $\mathbf{A}_{2221}$ , we make a further decomposition of its term inside the braces,

$$\begin{aligned}
& (\mathcal{L}_{S,t}^{-1} \mathbf{F}_{t,N,n})' \mathbf{G}_{j,t,0}^{(S)} (\mathcal{L}_{S,t}^{-1} \mathbf{F}_{t,N,n}) = \frac{1}{N_t^2} \sum_{\ell=1}^{N_t} \phi_{\ell,t} \left( \mathcal{L}_{S,t}^{-1} \mathbf{G}_{\ell,t,0} \right)' \mathbf{G}_{j,t,0}^{(S)} \left( \mathcal{L}_{S,t}^{-1} \mathbf{G}_{\ell,t,0} \right) \\
& \quad + 2\lambda^2 V_t^2 \frac{k_n}{N_t^2} \left( \mathcal{L}_{S,t}^{-1} \mathbf{C}_{t,0} \right)' \mathbf{G}_{j,t,0}^{(S)} \left( \mathcal{L}_{S,t}^{-1} \mathbf{C}_{t,0} \right) \\
& \quad + \frac{1}{N_t^2} \sum_{\ell=1}^{N_t} (\epsilon_{\ell,t}^2 - \phi_{\ell,t}) \left( \mathcal{L}_{S,t}^{-1} \mathbf{G}_{\ell,t,0} \right)' \mathbf{G}_{j,t,0}^{(S)} \left( \mathcal{L}_{S,t}^{-1} \mathbf{G}_{\ell,t,0} \right) \\
& \quad + \frac{1}{N_t^2} \sum_{\ell, g=1; \ell \neq g}^{N_t} \epsilon_{\ell,t} \epsilon_{g,t} \left( \mathcal{L}_{S,t}^{-1} \mathbf{G}_{\ell,t,0} \right)' \mathbf{G}_{j,t,0}^{(S)} \left( \mathcal{L}_{S,t}^{-1} \mathbf{G}_{g,t,0} \right) \\
& \quad + \lambda^2 \frac{k_n}{N_t^2} \left\{ k_n \left( \hat{V}_t^n - \xi_1(\mathbf{S}_t) \right)^2 - 2V_t^2 \right\} \left( \mathcal{L}_{S,t}^{-1} \mathbf{C}_{t,0} \right)' \mathbf{G}_{j,t,0}^{(S)} \left( \mathcal{L}_{S,t}^{-1} \mathbf{C}_{t,0} \right) \\
& \quad + 2\lambda \frac{k_n}{N_t^2} \left( \hat{V}_t^n - \xi_1(\mathbf{S}_t) \right) \sum_{\ell=1}^{N_t} \epsilon_{\ell,t} \left( \mathcal{L}_{S,t}^{-1} \mathbf{G}_{\ell,t,0} \right)' \mathbf{G}_{j,t,0}^{(S)} \left( \mathcal{L}_{S,t}^{-1} \mathbf{C}_{t,0} \right) \\
& \equiv \tilde{A}_{22211} + \tilde{A}_{22212} + \tilde{A}_{22213} + \tilde{A}_{22214} + \tilde{A}_{22215} + \tilde{A}_{22216},
\end{aligned}$$

where, for brevity, we have dropped the dependence on  $t$  and  $j$  from the notation used for the six terms on the right-hand-side. By definition, we have

$$\frac{1}{2TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \left\{ \tilde{A}_{22211} + \tilde{A}_{22212} \right\} \mathbf{U}_{j,t,0} = \mathbf{B}_{T,N,3}^{(1)}.$$

Moreover, note that the corresponding terms for  $\tilde{A}_{22213}$ ,  $\tilde{A}_{22214}$ , and  $\tilde{A}_{22216}$  exactly mirror the structure of the terms  $\mathbf{A}_{1211}$ ,  $\mathbf{A}_{1212}$ , and  $\mathbf{A}_{1213}$ , respectively. Hence, by applying arguments that are similar to the analyses of those terms, we get

$$\left\| \frac{1}{2\sqrt{TN}} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \left\{ \tilde{A}_{22213} + \tilde{A}_{22214} + \tilde{A}_{22216} \right\} \mathbf{U}_{j,t,0} \right\| \xrightarrow{\mathbb{P}} 0.$$

For the term  $\tilde{\mathbf{A}}_{22215}$ , we use the integrability assumptions in A6 and A7 in conjunction with Lemma 12 to show

$$\left\| \frac{1}{2\sqrt{TN}} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \tilde{\mathbf{A}}_{22215} \mathbf{U}_{j,t,0} \right\| \xrightarrow{\mathbb{P}} 0.$$

Combining all the results for the terms in the decomposition of  $\mathbf{A}_{22}$ , we have

$$\sqrt{TN} \left\| \mathbf{A}_{22} - \mathbf{B}_{T,N,2}^{(1)} - \mathbf{B}_{T,N,3}^{(1)} \right\| \xrightarrow{\mathbb{P}} 0.$$

For the last main term,  $\mathbf{A}_{33}$ , we write

$$\begin{aligned} \mathbf{A}_{33} &= \frac{\lambda}{T} \sum_{t=1}^T \frac{k_n}{N_t} \mathbf{R}'_{t,0} \mathbf{C}_{t,0}^{(S)} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{s}_t) \left( \hat{\mathbf{V}}_t^n - \xi_1(\mathbf{s}_t) \right) \\ &\quad - \frac{\lambda}{T} \sum_{t=1}^T \frac{k_n}{N_t} \mathbf{R}'_{t,0} \mathbf{C}_{t,0}^{(S)} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{s}_t) \left( \mathbf{C}'_{t,0} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{s}_t) \right) \\ &\quad - \frac{\lambda}{2T} \sum_{t=1}^T \frac{k_n}{N_t} \mathbf{R}'_{t,0} \mathbf{C}_{t,0} \left( (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{s}_t)' \mathbf{C}_{t,0}^{(S)} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{s}_t) \right) \equiv \mathbf{A}_{331} - \mathbf{A}_{332} - \mathbf{A}_{333}. \end{aligned}$$

Then, as the analyses of  $\mathbf{A}_{331}$ ,  $\mathbf{A}_{332}$ , and  $\mathbf{A}_{333}$  exactly mirror those for the terms  $\mathbf{A}_{12}$ ,  $\mathbf{A}_{221}$ , and  $\mathbf{A}_{222}$ , respectively, we have by similar arguments,

$$\sqrt{TN} \left\| \mathbf{A}_{331} - \mathbf{B}_{T,N,1}^{(2)} \right\| \xrightarrow{\mathbb{P}} 0, \quad \sqrt{TN} \left\| \mathbf{A}_{332} - \mathbf{B}_{T,N,2}^{(2)} \right\| \xrightarrow{\mathbb{P}} 0, \quad \sqrt{TN} \left\| \mathbf{A}_{333} - \mathbf{B}_{T,N,3}^{(2)} \right\| \xrightarrow{\mathbb{P}} 0,$$

concluding the proof.  $\square$

**Lemma 4.** *Under the conditions of Theorem 2, we have*

$$\sqrt{TN} (\|\mathbf{A}_{13}\| + \|\mathbf{A}_{23}\| + \|\mathbf{A}_{34}\|) \xrightarrow{\mathbb{P}} 0.$$

*Proof.* First, decompose  $\mathbf{A}_{13}$  as

$$\begin{aligned} \mathbf{A}_{13} &= \frac{1}{2TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{U}_{j,t,0}^{(SS)} \times \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{s}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{s}_t) \right\} \\ &\quad + \frac{1}{6TN} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \left\{ \frac{\partial \mathbf{U}_{j,t,0}^{(SS)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)}{\partial \mathbf{S}'} \right\} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{s}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{s}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{s}_t) \right\} \\ &\equiv \mathbf{A}_{131} + \mathbf{A}_{132}. \end{aligned}$$

Now, for the first these terms,  $\mathbf{A}_{131}$ , we invoke Lemma 10 with  $r_1 = 2$ ,  $r_2 = 1$ ,  $r_3 = 0$  (noting that  $k_n \Delta_n \rightarrow 0$ ), to establish the bound  $\|\mathbf{A}_{131}\| = O_p(N^{-3/2})$ .

Next, using the triangle inequality and addition and subtraction,  $\mathbf{A}_{132}$  may be bounded as

$$\sqrt{TN} \|\mathbf{A}_{132}\| \leq \frac{\sqrt{N}}{6\sqrt{T}} \sum_{t=1}^T \left\| \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \frac{\partial \mathbf{U}_{j,t,0}^{(SS)}}{\partial \mathbf{S}'} \right\| \times \left\| (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{s}_t) \right\|^3$$

$$\begin{aligned}
& + \frac{\sqrt{N}}{6\sqrt{T}} \sum_{t=1}^T \left\| \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \left( \frac{\partial \mathbf{U}_{j,t}^{(SS)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)}{\partial \mathbf{S}'} - \frac{\partial \mathbf{U}_{j,t,0}^{(SS)}}{\partial \mathbf{S}'} \right) \right\| \times \left\| (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\|^3 \\
& \equiv \mathbf{A}_{1321} + \mathbf{A}_{1322}.
\end{aligned}$$

As above, we use Lemma 10 with  $r_1 = 3$ ,  $r_2 = 1$ ,  $r_3 = 0$  to establish the bound  $\mathbf{A}_{1321} = O_p(\sqrt{T}(k_n^{5/2} \Delta_n \vee 1)/N^{3/2})$ . For  $\mathbf{A}_{1322}$ , we work on a set on which  $\sup_{t=1, \dots, T} \|\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t\| \leq \epsilon$  for some  $\epsilon > 0$ . This set has probability approaching one since  $\sup_{t=1, \dots, T} \|\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t\| \xrightarrow{\mathbb{P}} 0$  and, as a result, focusing on it is no restriction. Using A7, we have on this set

$$\left\| \frac{\partial \mathbf{U}_{j,t}^{(SS)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)}{\partial \mathbf{S}'} - \frac{\partial \mathbf{U}_{j,t,0}^{(SS)}}{\partial \mathbf{S}'} \right\| \leq K f_d(\|\mathbf{S}_t\|) \|\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t\|.$$

Hence, we may bound  $\mathbf{A}_{1322}$  by  $K \sup_{t=1, \dots, T} \|\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t\| \times (\mathbf{A}_{13221} + \mathbf{A}_{13222})$ , where

$$\begin{aligned}
\mathbf{A}_{13221} &= \frac{\sqrt{N}}{\sqrt{T}} \sum_{t=1}^T \left( \frac{1}{N_t} \sum_{j=1}^{N_t} \phi_{j,t}^{(1)} f_d(\|\mathbf{S}_t\|) \right) \times \left\| (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\|^3, \\
\mathbf{A}_{13222} &= \frac{\sqrt{N}}{\sqrt{T}} \sum_{t=1}^T \left| \frac{1}{N_t} \sum_{j=1}^{N_t} (|\epsilon_{j,t}| - \phi_{j,t}^{(1)}) f_d(\|\mathbf{S}_t\|) \right| \times \left\| (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\|^3.
\end{aligned}$$

Now, for  $\mathbf{A}_{13221}$ , we use Lemma 10 with  $r_1 = 3$ ,  $r_2 = r_3 = 0$ , to show  $\mathbf{A}_{1321} = O_p(\sqrt{T}(k_n^{5/2} \Delta_n \vee 1)/N)$ . For the second term  $\mathbf{A}_{13222}$ , we can apply an analogous result to the one in Lemma 10 in which  $\epsilon_{j,t}$  is replaced by  $|\epsilon_{j,t}| - \phi_{j,t}^{(1)}$ , with  $r_1 = 3$ ,  $r_2 = 1$ ,  $r_3 = 0$  to get  $\mathbf{A}_{13222} = O_p(\sqrt{T}(k_n^{5/2} \Delta_n \vee 1)/N^{3/2})$ . This plus the fact that  $\sup_{t=1, \dots, T} \|\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t\| \xrightarrow{\mathbb{P}} 0$  as well as  $\sqrt{T} k_n^{3/2} \Delta_n \rightarrow 0$  and  $k_n/N = O(1)$ , implies  $\mathbf{A}_{1322} = o_p(1)$ . Thus, altogether,  $\sqrt{TN} \|\mathbf{A}_{132}\| = o_p(1)$  and hence also  $\sqrt{TN} \|\mathbf{A}_{13}\| = o_p(1)$ .

Similar arguments are used to derive the bounds  $\sqrt{TN} \|\mathbf{A}_{23}\| = O_p(\sqrt{T}(k_n^3 \Delta_n \vee 1)/N^{3/2})$  and  $\sqrt{TN} \|\mathbf{A}_{34}\| = O_p(\sqrt{T}(k_n^{5/2} \Delta_n \vee 1)/N)$ , and, since  $\sqrt{T} k_n^{3/2} \Delta_n \rightarrow 0$  and  $k_n/N = O(1)$ , we may combine results to show  $\sqrt{TN} \|\mathbf{A}_{23}\| + \sqrt{TN} \|\mathbf{A}_{34}\| = o_p(1)$ . This concludes the proof.  $\square$

**Lemma 5.** *Under the conditions of Theorem 2, we have*

$$\sqrt{TN} \|\mathbf{A}_4\| \xrightarrow{\mathbb{P}} 0.$$

*Proof.* First, define

$$\begin{aligned}
\mathbf{A}_4 &\equiv \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \left( \frac{\partial \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \mathbf{G}_j(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) \\
&+ \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \left( \kappa_j(\mathbf{S}_t, \boldsymbol{\theta}_0) - \kappa_j(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) \right) \left( \frac{\partial \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \mathbf{G}_j(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)
\end{aligned}$$

$$\begin{aligned}
& + \frac{\lambda}{T} \sum_{t=1}^T \frac{k_n}{N_t} \left( \frac{\partial \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \mathbf{C}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \left( \hat{V}_t^n - \xi_1(\mathbf{S}_t) \right) \\
& + \frac{\lambda}{T} \sum_{t=1}^T \frac{k_n}{N_t} \left( \frac{\partial \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \mathbf{C}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \left( \xi_1(\mathbf{S}_t) - \xi_1(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \right) \\
& \equiv \hat{\mathbf{A}}_{41} + \hat{\mathbf{A}}_{42} + \hat{\mathbf{A}}_{43} + \hat{\mathbf{A}}_{44}.
\end{aligned}$$

For  $\hat{\mathbf{A}}_{41}$ , we apply a second-order Taylor expansion,

$$\begin{aligned}
\mathbf{G}_j(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) &= \mathbf{G}_{j,t,0} + \mathbf{G}_{j,t,0}^{(S)}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \\
&+ \left\{ \frac{\partial \mathbf{G}_j^{(S)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)}{\partial \mathbf{S}'} \right\} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\}, \tag{56}
\end{aligned}$$

where  $\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0) \in (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \mathbf{S}_t)$ , and make the decomposition,

$$\begin{aligned}
\hat{\mathbf{A}}_{41} &= \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \left( \frac{\partial \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \mathbf{G}_{j,t,0} \\
&+ \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \left( \frac{\partial \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \mathbf{G}_{j,t,0}^{(S)}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \\
&+ \frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \left( \frac{\partial \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \left\{ \frac{\partial \mathbf{G}_j^{(S)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)}{\partial \mathbf{S}'} \right\} \left\{ (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \\
&\equiv \mathbf{A}_{41} + \hat{\mathbf{A}}_{411} + \hat{\mathbf{A}}_{412}.
\end{aligned}$$

Then, for  $\hat{\mathbf{A}}_{411}$ , we use the triangle inequality as well as part (c) of Lemma 9 to bound

$$\sqrt{TN} \left\| \hat{\mathbf{A}}_{411} \right\| \leq K \sup_{t=1, \dots, T} |\hat{\eta}_{t,n}| \times \frac{\sqrt{N}}{\sqrt{T}} \sum_{t=1}^T f_d(\|\mathbf{S}_t\|) \left\| \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{G}_{j,t,0}^{(S)} \right\| \times \left\| \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t \right\|,$$

where  $\hat{\eta}_{t,n}$  is the variable in Lemma 9, for which we have  $\sup_{t=1, \dots, T} |\hat{\eta}_{t,n}| = o_p(1)$ . Now, we apply Lemma 10 with  $r_1 = r_2 = 1$ ,  $r_3 = 0$  and conclude  $\sqrt{TN} \left\| \hat{\mathbf{A}}_{411} \right\| = o_p(\sqrt{T}/\sqrt{N})$ . For  $\hat{\mathbf{A}}_{412}$ , we may restrict attention to the set on which  $\sup_{t=1, \dots, T} \|\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t\| \leq \epsilon$  for some  $\epsilon > 0$ . The probability of this set is approaching one and, hence, this is no restriction. On this set, using A7, we have

$$\left\| \frac{\partial \mathbf{G}_j^{(S)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)}{\partial \mathbf{S}'} - \frac{\partial \mathbf{G}_j^{(S)}(\mathbf{S}_t, \boldsymbol{\theta}_0)}{\partial \mathbf{S}'} \right\| \leq K f_d(\|\mathbf{S}_t\|) \|\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t\|.$$

By applying this bound as well as part (c) of Lemma 9 and Lemma 10 with  $r_1 = 2$ ,  $r_2 = 1$ ,  $r_3 = 0$ , we readily get  $\sqrt{TN} \left\| \hat{\mathbf{A}}_{412} \right\| = o_p(\sqrt{T}/N)$ . Moreover, by collecting the above results, we have  $\sqrt{TN} \left\| \hat{\mathbf{A}}_{41} - \mathbf{A}_{41} \right\| \xrightarrow{\mathbb{P}} 0$ . The corresponding results,

$$\sqrt{TN} \left\| \hat{\mathbf{A}}_{42} - \mathbf{A}_{42} \right\| + \sqrt{TN} \left\| \hat{\mathbf{A}}_{43} - \mathbf{A}_{43} \right\| + \sqrt{TN} \left\| \hat{\mathbf{A}}_{44} - \mathbf{A}_{44} \right\| = o_p(\sqrt{T}/\sqrt{N}),$$

where  $\mathbf{A}_{42}$ ,  $\mathbf{A}_{43}$ , and  $\mathbf{A}_{44}$  are defined as

$$\begin{aligned}\mathbf{A}_{42} &\equiv -\frac{1}{T} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \left( \mathbf{G}'_{j,t,0} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right) \left( \frac{\partial \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \mathbf{G}_{j,t,0}, \\ \mathbf{A}_{43} &\equiv \frac{\lambda}{T} \sum_{t=1}^T \frac{k_n}{N_t} \left( \frac{\partial \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \mathbf{C}_{t,0} \left( \hat{V}_t^n - \xi_1(\mathbf{S}_t) \right), \\ \mathbf{A}_{44} &\equiv \frac{\lambda}{T} \sum_{t=1}^T \frac{k_n}{N_t} \left( \frac{\partial \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right)' \mathbf{C}_{t,0} \left( (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' \mathbf{C}_{t,0} \right),\end{aligned}$$

respectively, are shown in exactly the same way. Hence, the proof will be concluded by establishing that

$$\sqrt{TN}(\mathbf{A}_{41} + \mathbf{A}_{42} + \mathbf{A}_{43} + \mathbf{A}_{44}) = o_p(1).$$

Defining  $\hat{\boldsymbol{g}}_t \equiv \partial \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) / \partial \boldsymbol{\theta}' - \mathbf{R}_{t,0}$  and using a first-order Taylor expansion, we decompose

$$\begin{aligned}\mathbf{A}_{41} + \mathbf{A}_{42} + \mathbf{A}_{43} + \mathbf{A}_{44} &= \frac{1}{T} \sum_{t=1}^T \hat{\boldsymbol{g}}_t' \left\{ \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{G}_{j,t,0} + \lambda \frac{k_n}{N_t} \mathbf{C}_{t,0} \left( \hat{V}_t^n - \xi_1(\mathbf{S}_t) \right) \right. \\ &\quad \left. - \left( \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{G}_{j,t,0} \mathbf{G}'_{j,t,0} + \lambda \frac{k_n}{N_t} \mathbf{C}_{t,0} \mathbf{C}'_{t,0} \right) (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\} \\ &= \frac{1}{T} \sum_{t=1}^T \hat{\boldsymbol{g}}_t' \left\{ \left( \tilde{\boldsymbol{\mathcal{L}}}_{S,t} - \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{G}_{j,t,0} \mathbf{G}'_{j,t,0} - \lambda \frac{k_n}{N_t} \mathbf{C}_{t,0} \mathbf{C}'_{t,0} \right) (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\}.\end{aligned}$$

Moreover, using the definition of  $\tilde{\boldsymbol{\mathcal{L}}}_{S,t}$ , we make the decomposition,

$$\begin{aligned}\tilde{\boldsymbol{\mathcal{L}}}_{S,t} - \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{G}_{j,t,0} \mathbf{G}'_{j,t,0} - \lambda \frac{k_n}{N_t} \mathbf{C}_{t,0} \mathbf{C}'_{t,0} &= \frac{1}{N_t} \sum_{j=1}^{N_t} \left( \mathbf{G}_j(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) \mathbf{G}_j(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)' - \mathbf{G}_{j,t,0} \mathbf{G}'_{j,t,0} \right) \\ &\quad + \lambda \frac{k_n}{N_t} \left( \mathbf{C}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \mathbf{C}'(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0))' - \mathbf{C}_{t,0} \mathbf{C}'_{t,0} \right) - \lambda \frac{k_n}{N_t} \mathbf{C}^{(S)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \left( \hat{V}_t^n - \xi_1(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \right) \\ &\quad - \frac{1}{N_t} \sum_{j=1}^{N_t} \left( \tilde{\kappa}_{j,t} - \kappa_j(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) \right) \mathbf{G}_j^{(S)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) \equiv \tilde{\boldsymbol{\mathcal{L}}}_{S,t,1} + \tilde{\boldsymbol{\mathcal{L}}}_{S,t,2} + \tilde{\boldsymbol{\mathcal{L}}}_{S,t,3} + \tilde{\boldsymbol{\mathcal{L}}}_{S,t,4}.\end{aligned}$$

Starting with  $\tilde{\boldsymbol{\mathcal{L}}}_{S,t,1}$ , we apply a first-order Taylor expansion and write

$$\begin{aligned}\tilde{\boldsymbol{\mathcal{L}}}_{S,t,1} &= \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{G}_{j,t,0} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' \mathbf{G}_j^{(S)}(\check{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) + \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{G}_j^{(S)}(\check{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \mathbf{G}'_{j,t,0} \\ &\quad + \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{G}_j^{(S)}(\check{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' \mathbf{G}_j^{(S)}(\check{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0),\end{aligned}$$

for some  $\check{\mathbf{S}}_t(\boldsymbol{\theta}_0) \in (\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \mathbf{S}_t)$ . From here, we may apply exactly the same steps as for the analysis of  $\hat{\mathbf{A}}_{411}$  and conclude that  $\sqrt{TN} \left\| \frac{1}{T} \sum_{t=1}^T \hat{\boldsymbol{g}}_t' \tilde{\boldsymbol{\mathcal{L}}}_{S,t,1} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\| = o_p(\sqrt{T}/\sqrt{N})$ . Similar arguments



provide

$$\sqrt{TN} \left\| \frac{1}{T} \sum_{t=1}^T \hat{\mathbf{G}}'_t(\tilde{\mathcal{L}}_{S,t,2} + \tilde{\mathcal{L}}_{S,t,3} + \tilde{\mathcal{L}}_{S,t,4})(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\| = o_p(\sqrt{T}/\sqrt{N}).$$

Hence, altogether, we have  $\sqrt{TN} \|\mathbf{A}_{41} + \mathbf{A}_{42} + \mathbf{A}_{43} + \mathbf{A}_{44}\| = o_p(1)$ , concluding the proof.  $\square$

**Lemma 6.** *Under the conditions of Theorem 2, we have*

$$\sqrt{TN} \|\mathbf{A}_5\| \xrightarrow{\mathbb{P}} 0.$$

*Proof.* We start by decomposing  $\mathbf{A}_5$  as

$$\begin{aligned} \mathbf{A}_5 &= \frac{1}{T} \sum_{t=1}^T \left\{ \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{D}_{j,t,0} \mathbf{G}'_{j,t,0} - \mathcal{I}'_{S,\theta,t} \right\} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \\ &\quad + \frac{1}{T} \sum_{t=1}^T \mathbf{R}'_{t,0} \left\{ \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{G}_{j,t,0} \mathbf{G}'_{j,t,0} - \mathcal{I}_{S,t} \right\} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \\ &\quad + \frac{\lambda}{T} \sum_{t=1}^T \left( \frac{k_n}{N_t} - \frac{\varrho}{S_t} \right) \mathbf{R}'_{t,0} \mathbf{C}_{t,0} \mathbf{C}'_{t,0} (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \\ &\quad + \frac{1}{T} \sum_{t=1}^T (\mathcal{I}'_{S,\theta,t} + \mathbf{R}'_{t,0} (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})) (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \equiv \mathbf{A}_{51} + \mathbf{A}_{52} + \mathbf{A}_{53} + \mathbf{A}_{54}. \end{aligned}$$

Then, by the definition of  $\mathbf{R}_{t,0}$ , we have  $\mathbf{A}_{54} = \mathbf{0}$ . Next, define

$$\mathbf{h}_{t,N,n}(\mathbf{Z}, \boldsymbol{\theta}) = \frac{\partial}{\partial \mathbf{Z}'} \hat{\mathcal{I}}_{S,t}(\mathbf{Z}, \boldsymbol{\theta}) + \frac{\partial}{\partial \mathbf{Z}'} \hat{\mathcal{P}}_{S,t}(\mathbf{Z}),$$

and use a second-order Taylor expansion for the equation that solves  $\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)$  to write

$$\mathbf{F}_{t,N,n} + \bar{\mathcal{L}}_{S,t}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) + \frac{1}{2} \mathbf{h}_{t,N,n}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) = \mathbf{0},$$

where, as before,  $\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0)$  is an intermediate value between  $\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)$  and  $\mathbf{S}_t$ , and

$$\bar{\mathcal{L}}_{S,t} = \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{G}_{j,t,0} \mathbf{G}'_{j,t,0} - \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{G}_{j,t,0}^{(S)} + \lambda \frac{k_n}{N_t} \mathbf{C}_{t,0} \mathbf{C}'_{t,0} - \lambda \frac{k_n}{N_t} \mathbf{C}_{t,0}^{(S)} (\hat{V}_t^n - V_t).$$

With this notation, decompose  $\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t = \hat{\mathbf{S}}_t^{(1)} + \hat{\mathbf{S}}_t^{(2)} + \hat{\mathbf{S}}_t^{(3)}$ , where  $\hat{\mathbf{S}}_t^{(1)} = \bar{\mathcal{L}}_{S,t}^{-1} \mathbf{F}_{t,N,n}$ ,

$$\hat{\mathbf{S}}_t^{(2)} = (\bar{\mathcal{L}}_{S,t}^{-1} - \mathcal{L}_{S,t}^{-1}) \mathbf{F}_{t,N,n}, \quad \hat{\mathbf{S}}_t^{(3)} = \frac{1}{2} \bar{\mathcal{L}}_{S,t}^{-1} \mathbf{h}_{t,N,n}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t).$$

By applying this decomposition, we write  $\mathbf{A}_{51} = \sum_{j=1}^3 \mathbf{A}_{51j}$ . Furthermore, using A2 and A7, we have

$$\left\| \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{D}_{j,t,0} \mathbf{G}'_{j,t,0} - \mathcal{I}'_{S,\theta,t} \right\| + \left\| \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{G}_{j,t,0} \mathbf{G}'_{j,t,0} - \mathcal{I}'_{S,t} \right\| \leq K \hat{\eta}_{t,N,n} \frac{f_d(\|\mathbf{S}_t\|)}{\sqrt{N}}.$$

Hence, by taking the above into account, applying part (b) of Lemma 9 and Lemma 10, we readily establish that  $\|\mathbf{A}_{512} + \mathbf{A}_{513}\| = o_p(1/N)$ . We now turn to  $\mathbf{A}_{511}$ , for which, using A5, part (c) of Lemma 8, and for  $n$  sufficiently high (so that  $k_n/n < 1$ ), we have

$$\|\mathbb{E}(\mathbf{F}_{t,N,n}|\mathcal{F}_t^{(0)})\| \leq K_{t,n} \left( \Delta_n^{(2-r)\varpi} \vee k_n \Delta_n \right),$$

where  $K_{t,n}$  has finite absolute moments and is adapted to  $\mathcal{F}_t^{(0)}$ . From here, we have

$$\frac{1}{T} \sum_{t=1}^T \left\{ \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{D}_{j,t,0} \mathbf{G}'_{j,t,0} - \mathcal{I}'_{S,\theta,t} \right\} \mathcal{L}_{S,t}^{-1} \mathbb{E}(\mathbf{F}_{t,N,n}|\mathcal{F}_t^{(0)}) = O_p \left( \frac{1}{\sqrt{N}} \left( \Delta_n^{(2-r)\varpi} \vee k_n \Delta_n \right) \right),$$

and, in addition, using Lemma 10 on a set with probability approaching one (that is, on the set in which we have  $\sup_{t=1,\dots,T} \hat{\eta}_{t,N,n} < 1$ ), it follows that

$$\mathbb{E} \left( \frac{1}{T} \sum_{t=1}^T \left\{ \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{D}_{j,t,0} \mathbf{G}'_{j,t,0} - \mathcal{I}'_{S,\theta,t} \right\} \mathcal{L}_{S,t}^{-1} (\mathbf{F}_{t,N,n} - \mathbb{E}(\mathbf{F}_{t,N,n}|\mathcal{F}_t^{(0)})) \right)^2 \leq \frac{K}{TN^2}.$$

Finally, taking into account that  $\sqrt{T}(k_n \Delta_n \vee \Delta_n^{(2-r)\varpi}) \rightarrow 0$ , we, thus, get  $\sqrt{TN} \|\mathbf{A}_{51}\| = o_p(1)$ . Similar arguments provide  $\sqrt{TN} \|\mathbf{A}_{52}\| + \sqrt{TN} \|\mathbf{A}_{53}\| = o_p(1)$ , concluding the proof.  $\square$

#### A.4 Proof of Theorem 3

First, by using Theorem 1 and applying a first-order Taylor expansion, we have

$$\|\hat{\mathbf{S}}_t - \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)\| \leq \left\| \frac{\partial \hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'} \right\| \times \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|,$$

and, from here, by invoking Lemma 9 (c) in conjunction with  $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = O_p(1/\sqrt{TN})$  and the fact that the function  $g$  has a bounded derivative, we additionally have

$$\frac{1}{\sqrt{T}} \sum_{t=2}^T (g(\hat{\mathbf{S}}_t, \hat{\mathbf{S}}_{t-1}) - g(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \hat{\mathbf{S}}_{t-1}(\boldsymbol{\theta}_0))) = O_p(1/\sqrt{N}).$$

Next, make the decomposition,  $\frac{1}{\sqrt{T}} \sum_{t=2}^T (g(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \hat{\mathbf{S}}_{t-1}(\boldsymbol{\theta}_0)) - g(\mathbf{S}_t, \mathbf{S}_{t-1})) = \mathbf{G}_1 + \mathbf{G}_2$ , where

$$\begin{aligned} \mathbf{G}_1 &= \frac{1}{\sqrt{T}} \sum_{t=2}^T (g(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \hat{\mathbf{S}}_{t-1}(\boldsymbol{\theta}_0)) - g(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \mathbf{S}_{t-1})), \\ \mathbf{G}_2 &= \frac{1}{\sqrt{T}} \sum_{t=2}^T (g(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \mathbf{S}_{t-1}) - g(\mathbf{S}_t, \mathbf{S}_{t-1})), \end{aligned}$$

by addition and subtraction. Let  $g^{(i)}$  and  $g^{(ii)}$  be the first and second derivatives of  $g$  with respect to the  $i$ th argument for  $i = 1, 2$ , then we first use a second-order Taylor expansion of the function  $g$  in

the second argument and the mean-value theorem to write,

$$\begin{aligned} \mathbf{G}_1 &\equiv \mathbf{G}_{11} + \mathbf{G}_{12} = \frac{1}{\sqrt{T}} \sum_{t=2}^T g^{(2)}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \mathbf{S}_{t-1})(\hat{\mathbf{S}}_{t-1}(\boldsymbol{\theta}_0) - \mathbf{S}_{t-1}) \\ &\quad + \frac{1}{2\sqrt{T}} \sum_{t=2}^T g^{(22)}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \tilde{\mathbf{S}}_{t-1}(\boldsymbol{\theta}_0)) \left\{ (\hat{\mathbf{S}}_{t-1}(\boldsymbol{\theta}_0) - \mathbf{S}_{t-1}) \otimes (\hat{\mathbf{S}}_{t-1}(\boldsymbol{\theta}_0) - \mathbf{S}_{t-1}) \right\}, \end{aligned}$$

for some  $\tilde{\mathbf{S}}_{t-1}(\boldsymbol{\theta}_0) \in (\hat{\mathbf{S}}_{t-1}(\boldsymbol{\theta}_0), \mathbf{S}_{t-1})$ . Now, for the first term in the last decomposition,  $\mathbf{G}_{11}$ , we may use another first-order Taylor expansion in the first argument of  $g$  to write,

$$\begin{aligned} \mathbf{G}_{11} &\equiv \mathbf{G}_{111} + \mathbf{G}_{112} = \frac{1}{\sqrt{T}} \sum_{t=2}^T g^{(2)}(\mathbf{S}_t, \mathbf{S}_{t-1}) \left\{ \hat{\mathbf{S}}_{t-1}(\boldsymbol{\theta}_0) - \mathbf{S}_{t-1} \right\} \\ &\quad + \frac{1}{2\sqrt{T}} \sum_{t=2}^T g^{(21)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \mathbf{S}_{t-1}) \left\{ (\hat{\mathbf{S}}_{t-1}(\boldsymbol{\theta}_0) - \mathbf{S}_{t-1}) \otimes (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t) \right\}, \end{aligned}$$

where, again,  $\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0) \in (\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \mathbf{S}_t)$ . Using the notation in (54), we can decompose  $\mathbf{G}_{111}$  into

$$\begin{aligned} \mathbf{G}_{1111} &= \frac{1}{\sqrt{T}} \sum_{t=2}^T g^{(2)}(\mathbf{S}_t, \mathbf{S}_{t-1}) \mathcal{L}_{S,t-1}^{-1} \mathbf{F}_{t-1,N,n}, \quad \text{and} \\ \mathbf{G}_{1112} &= \frac{1}{\sqrt{T}} \sum_{t=2}^T g^{(2)}(\mathbf{S}_t, \mathbf{S}_{t-1}) (\tilde{\mathcal{L}}_{S,t-1}^{-1} - \mathcal{L}_{S,t-1}^{-1}) \mathbf{F}_{t-1,N,n}. \end{aligned}$$

Then, for  $\mathbf{G}_{1111}$ , using successive conditioning in conjunction with the boundedness of  $g^{(2)}$  and Lemma 7, we have

$$\mathbb{E} \|\mathbf{G}_{111}(a)\|^2 \leq K/N,$$

and, thus,  $\mathbf{G}_{1111} = O_p(1/\sqrt{N})$ . For  $\mathbf{G}_{1112}$ , we may use Lemmas 9 and 10 (noting that  $k_n^{3/2} \Delta_n \rightarrow 0$ ) to show  $\mathbf{G}_{1112} = o_p(1)$ . Next, for  $\mathbf{G}_{112}$ , we use the boundedness of  $g^{(21)}$  and Lemma 11 to deduce

$$\|\mathbf{G}_{112}\| \leq \frac{1}{2\sqrt{T}} \sum_{t=2}^T \left\| g^{(21)}(\tilde{\mathbf{S}}_t(\boldsymbol{\theta}_0), \mathbf{S}_{t-1}) \right\| \times \left\| \hat{\mathbf{S}}_{t-1}(\boldsymbol{\theta}_0) - \mathbf{S}_{t-1} \right\| \times \left\| \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t \right\| = O_p(T^{1/2}N^{-1}).$$

The same arguments may readily be used to establish that  $\mathbf{G}_{12} = O_p(T^{1/2}/N)$ , which, combined with the above results, provides  $\|\mathbf{G}_1\| = o_p(1)$ . The second term,  $\mathbf{G}_2$ , can be analyzed in an identical manner, thereby providing the claim in (26).  $\square$

## A.5 Definitions and Proofs for Section 4

To develop and establish the feasible central limit theory in Section 4, this section first introduces the necessary notation to establish both Lemma 1 and Theorem 4 before proceeding to the proofs.

### A.5.1 Definitions

First, in analogy with  $\mathcal{I}_{\theta,t}$  and  $\mathcal{I}_{S,\theta,t}$ , defined in the main text, write

$$\begin{aligned}\boldsymbol{\Omega}_{S,\theta,t} &= \frac{1}{\zeta_t} \sum_{\tau} \pi_t^{\tau} \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \phi_{t,k,\tau} \mathbf{G}(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0) \mathbf{D}_t(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0)' dk, \\ \boldsymbol{\Omega}_{\theta,t} &= \frac{1}{\zeta_t} \sum_{\tau} \pi_t^{\tau} \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \phi_{t,k,\tau} \mathbf{D}(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0) \mathbf{D}_t(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0)' dk,\end{aligned}$$

and define the  $Tp \times q$  vector  $\mathbf{R}_0 = (\mathbf{R}_{1,0}, \mathbf{R}_{2,0}, \dots, \mathbf{R}_{T,0})'$  as well as

$$\mathcal{I} = \begin{pmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{pmatrix}, \quad \mathcal{P} = \begin{pmatrix} \mathcal{P}_{11} & \mathbf{0}_{Tp \times q} \\ \mathbf{0}_{q \times Tp} & \mathbf{0}_{q \times q} \end{pmatrix}, \quad \boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}, \quad \boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}_{11} & \mathbf{0}_{Tp \times q} \\ \mathbf{0}_{q \times Tp} & \mathbf{0}_{q \times q} \end{pmatrix}$$

where  $\mathcal{I}_{22} = \mathcal{I}_{\theta} \equiv T^{-1} \sum_{t=1}^T \mathcal{I}_{\theta,t}$ ,  $\boldsymbol{\Omega}_{22} = \boldsymbol{\Omega}_{\theta} \equiv T^{-1} \sum_{t=1}^T \boldsymbol{\Omega}_{\theta,t}$ ,

$$\begin{aligned}\mathcal{I}_{11} &\equiv \text{diag}(\mathcal{I}_{S,1}, \mathcal{I}_{S,2}, \dots, \mathcal{I}_{S,T}), & \boldsymbol{\Omega}_{11} &\equiv \text{diag}(\boldsymbol{\Omega}_{S,1}, \boldsymbol{\Omega}_{S,2}, \dots, \boldsymbol{\Omega}_{S,T}), \\ \mathcal{P}_{11} &\equiv \text{diag}(\mathcal{P}_{S,1}, \mathcal{P}_{S,2}, \dots, \mathcal{P}_{S,T}), & \boldsymbol{\Psi}_{11} &\equiv \text{diag}(\boldsymbol{\Psi}_{S,1}, \boldsymbol{\Psi}_{S,2}, \dots, \boldsymbol{\Psi}_{S,T}), \\ \mathcal{I}_{12} = \mathcal{I}'_{21} &\equiv (\mathcal{I}_{S,\theta,1}, \mathcal{I}_{S,\theta,2}, \dots, \mathcal{I}_{S,\theta,T})' / \sqrt{T}, & \boldsymbol{\Omega}_{12} = \boldsymbol{\Omega}'_{21} &\equiv (\boldsymbol{\Omega}_{S,\theta,1}, \boldsymbol{\Omega}_{S,\theta,2}, \dots, \boldsymbol{\Omega}_{S,\theta,T})' / \sqrt{T}.\end{aligned}$$

This simplified notation is used to prove Lemma 1.

Next, for the feasible central limit theory, we first denote  $\bar{N} = \frac{1}{T} \sum_{t=1}^T N_t$ ,  $\bar{\varrho} = k_n / \bar{N}$ , as well as  $\bar{\zeta}_t = N_t / \bar{N}$ . With this notation, write  $\hat{V}_t = \xi_1(\hat{\mathbf{S}}_t)$ ,  $\hat{\mathbf{C}}_t = \mathbf{C}(\hat{\mathbf{S}}_t)$ , and  $\hat{\mathbf{C}}_t^{(S)} = \mathbf{C}^{(S)}(\hat{\mathbf{S}}_t)$ . We then define

$$\hat{\mathcal{P}}_{S,t} = \frac{\lambda \bar{\varrho}}{\bar{\zeta}_t} \times \hat{\mathbf{C}}_t \hat{\mathbf{C}}_t', \quad \text{and} \quad \hat{\boldsymbol{\Psi}}_{S,t} = 2 \frac{\lambda^2 \bar{\varrho}}{\bar{\zeta}_t^2} \hat{V}_t^2 \hat{\mathbf{C}}_t \hat{\mathbf{C}}_t'.$$

Similarly, we denote the gradient vector estimators by  $\hat{\mathbf{D}}_{j,t} \equiv \mathbf{D}(k_j, \tau_j, \hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}})$ ,  $\hat{\mathbf{G}}_{j,t} \equiv \mathbf{G}(k_j, \tau_j, \hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}})$ ,  $\hat{\mathbf{G}}_{j,t}^{(S)} \equiv \mathbf{G}^{(S)}(k_j, \tau_j, \hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}})$ , and  $\hat{\mathbf{D}}_{j,t}^{(S)} \equiv \mathbf{D}^{(S)}(k_j, \tau_j, \hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}})$ , as well as the Hessian-type estimators

$$\hat{\mathcal{I}}_{S,t} \equiv \frac{1}{N_t} \sum_{j=1}^{N_t} \hat{\mathbf{G}}_{j,t} \hat{\mathbf{G}}_{j,t}', \quad \hat{\mathcal{I}}_{S,\theta,t} \equiv \frac{1}{N_t} \sum_{j=1}^{N_t} \hat{\mathbf{G}}_{j,t} \hat{\mathbf{D}}_{j,t}', \quad \hat{\boldsymbol{\Omega}}_{S,t} \equiv \frac{1}{\bar{\zeta}_t N_t} \sum_{j=1}^{N_t} \left( \tilde{\kappa}_{j,t} - \kappa_j(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) \right)^2 \hat{\mathbf{G}}_{j,t} \hat{\mathbf{G}}_{j,t}'$$

which, in turn, are used to define

$$\hat{\mathbf{R}}_t \equiv -(\hat{\mathcal{I}}_{S,t} + \hat{\mathcal{P}}_{S,t})^{-1} \hat{\mathcal{I}}_{S,\theta,t}, \quad \hat{\mathbf{U}}_{j,t} \equiv \hat{\mathbf{D}}_{j,t} + \hat{\mathbf{R}}_t' \hat{\mathbf{G}}_{j,t}, \quad \hat{\mathbf{U}}_{j,t}^{(S)} \equiv \hat{\mathbf{D}}_{j,t}^{(S)} + \hat{\mathbf{R}}_t' \hat{\mathbf{G}}_{j,t}^{(S)},$$

together with  $\hat{\mathcal{P}}_{\theta,t} = \hat{\mathbf{R}}_t' \hat{\mathcal{P}}_{S,t} \hat{\mathbf{R}}_t$ ,  $\hat{\boldsymbol{\Psi}}_{\theta,t} = \hat{\mathbf{R}}_t' \hat{\boldsymbol{\Psi}}_{S,t} \hat{\mathbf{R}}_t$ ,

$$\hat{\mathcal{I}}_{U,t} \equiv \frac{1}{N_t} \sum_{j=1}^{N_t} \hat{\mathbf{U}}_{j,t} \hat{\mathbf{U}}_{j,t}' \quad \text{and} \quad \hat{\boldsymbol{\Omega}}_{U,t} \equiv \frac{1}{\bar{\zeta}_t N_t} \sum_{j=1}^{N_t} \left( \tilde{\kappa}_{j,t} - \kappa_j(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) \right)^2 \hat{\mathbf{U}}_{j,t} \hat{\mathbf{U}}_{j,t}'.$$

The asymptotic covariance matrix terms are, then, estimated as

$$\hat{\boldsymbol{\chi}} = T^{-1} \sum_{t=1}^T \hat{\boldsymbol{\chi}}_t \quad \text{for} \quad \hat{\boldsymbol{\chi}}_t = \{\hat{\boldsymbol{\mathcal{I}}}_{U,t}, \hat{\boldsymbol{\mathcal{P}}}_{\theta,t}, \hat{\boldsymbol{\Omega}}_{U,t}, \hat{\boldsymbol{\Psi}}_{\theta,t}\}$$

and, similarly, the asymptotic bias terms as

$$\begin{aligned} \hat{\boldsymbol{B}}_{T,N} &= \left( \hat{\boldsymbol{\mathcal{I}}}_U + \hat{\boldsymbol{\mathcal{P}}}_\theta \right)^{-1} \times \sum_{j=1}^2 \left( \hat{\boldsymbol{B}}_{T,N,1}^{(j)} - \hat{\boldsymbol{B}}_{T,N,2}^{(j)} - \hat{\boldsymbol{B}}_{T,N,3}^{(j)} \right), \quad \text{where} \\ \hat{\boldsymbol{B}}_{T,N,1}^{(1)} &= \frac{1}{T\bar{N}} \sum_{t=1}^T \frac{\bar{N}}{N_t} \frac{1}{N_t} \sum_{j=1}^{N_t} \left( \tilde{\kappa}_{j,t} - \kappa_j(\hat{\boldsymbol{S}}_t, \hat{\boldsymbol{\theta}}) \right)^2 \left( \hat{\boldsymbol{U}}_{j,t}^{(S)} \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \hat{\boldsymbol{G}}_{j,t} \right), \\ \hat{\boldsymbol{B}}_{T,N,2}^{(1)} &= \frac{1}{T\bar{N}} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \hat{\boldsymbol{U}}_{j,t}^{(S)} \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \left( \hat{\boldsymbol{\Omega}}_{S,t} + \hat{\boldsymbol{\Psi}}_{S,t} \right) \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \hat{\boldsymbol{G}}_{j,t}, \\ \hat{\boldsymbol{B}}_{T,N,3}^{(1)} &= \frac{1}{2T\bar{N}} \sum_{t=1}^T \frac{1}{N_t} \sum_{j=1}^{N_t} \hat{\boldsymbol{U}}_{j,t} \times \\ &\quad \left\{ \frac{\bar{N}}{N_t^2} \sum_{\ell=1}^{N_t} \left( \tilde{\kappa}_{\ell,t} - \kappa_\ell(\hat{\boldsymbol{S}}_t, \hat{\boldsymbol{\theta}}) \right)^2 \left( \hat{\boldsymbol{G}}'_{\ell,t} \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \hat{\boldsymbol{G}}_{j,t}^{(S)} \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \hat{\boldsymbol{G}}_{\ell,t} \right) \right. \\ &\quad \left. + 2\lambda \hat{V}_t^2 \frac{k_n \bar{N}}{N_t^2} \left( \hat{\boldsymbol{C}}'_t \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \hat{\boldsymbol{G}}_{j,t}^{(S)} \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \hat{\boldsymbol{C}}_t \right) \right\}, \\ \hat{\boldsymbol{B}}_{T,N,1}^{(2)} &= \frac{1}{T\bar{N}} \sum_{t=1}^T 2\lambda^2 \hat{V}_t^2 \frac{k_n \bar{N}}{N_t^2} \hat{\boldsymbol{R}}_t \hat{\boldsymbol{C}}_t^{(S)} \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \hat{\boldsymbol{C}}_t, \\ \hat{\boldsymbol{B}}_{T,N,2}^{(2)} &= \frac{\lambda}{T\bar{N}} \sum_{t=1}^T \frac{k_n}{N_t} \hat{\boldsymbol{R}}'_t \hat{\boldsymbol{C}}_t^{(S)} \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \left( \hat{\boldsymbol{\Omega}}_{S,t} + \hat{\boldsymbol{\Psi}}_{S,t} \right) \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \hat{\boldsymbol{C}}_t, \\ \hat{\boldsymbol{B}}_{T,N,3}^{(2)} &= \frac{\lambda}{2T\bar{N}} \sum_{t=1}^T \frac{k_n}{N_t} \hat{\boldsymbol{R}}'_t \hat{\boldsymbol{C}}_t \times \\ &\quad \left\{ \frac{\bar{N}}{N_t^2} \sum_{\ell=1}^{N_t} \left( \tilde{\kappa}_{\ell,t} - \kappa_\ell(\hat{\boldsymbol{S}}_t, \hat{\boldsymbol{\theta}}) \right)^2 \left( \hat{\boldsymbol{G}}'_{\ell,t} \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \hat{\boldsymbol{C}}_t^{(S)} \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \hat{\boldsymbol{G}}_{\ell,t} \right) \right. \\ &\quad \left. + 2\lambda \hat{V}_t^2 \frac{k_n \bar{N}}{N_t^2} \left( \hat{\boldsymbol{C}}'_t \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \hat{\boldsymbol{C}}_t^{(S)} \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \hat{\boldsymbol{C}}_t \right) \right\}. \end{aligned}$$

Finally, let us define

$$\hat{\boldsymbol{\Omega}}_{S,U,t} \equiv \frac{1}{\bar{\zeta}_t N_t} \sum_{j=1}^{N_t} \left( \tilde{\kappa}_{j,t} - \kappa_j(\hat{\boldsymbol{S}}_t, \hat{\boldsymbol{\theta}}) \right)^2 \hat{\boldsymbol{G}}_{j,t} \hat{\boldsymbol{U}}'_{j,t}$$

such that the covariance matrix for the state vector at time  $t = 1, \dots, T$ , may be estimated as

$$\begin{aligned} \widehat{\mathbf{AVAR}}_{S_t} &= \widehat{\mathbf{AVAR}}_{S_{t,1}} + \widehat{\mathbf{AVAR}}_{S_{t,2}} + \widehat{\mathbf{AVAR}}_{S_{t,3}} + \widehat{\mathbf{AVAR}}_{S_{t,4}}, \quad \text{where} \\ \widehat{\mathbf{AVAR}}_{S_{t,1}} &= \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \left( \hat{\boldsymbol{\Omega}}_{S,t} + \hat{\boldsymbol{\Psi}}_{S,t} \right) \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1}, \end{aligned}$$

$$\begin{aligned}
\widehat{\mathbf{AVAR}}_{S_t,2} &= \hat{\mathbf{R}}_t (\hat{\mathcal{I}}_U + \hat{\mathcal{P}}_\theta)^{-1} (\hat{\Omega}_U + \hat{\Psi}_\theta) (\hat{\mathcal{I}}_U + \hat{\mathcal{P}}_\theta)^{-1} \hat{\mathbf{R}}_t' / T, \\
\widehat{\mathbf{AVAR}}_{S_t,3} &= (\hat{\mathcal{I}}_{S,t} + \hat{\mathcal{P}}_{S,t})^{-1} (\hat{\Omega}_{S,U,t} + \hat{\Psi}_{S,t} \hat{\mathbf{R}}_t) (\hat{\mathcal{I}}_U + \hat{\mathcal{P}}_\theta)^{-1} \hat{\mathbf{R}}_t' / T, \\
\widehat{\mathbf{AVAR}}_{S_t,4} &= \hat{\mathbf{R}}_t (\hat{\mathcal{I}}_U + \hat{\mathcal{P}}_\theta)^{-1} (\hat{\Omega}'_{S,U,t} + \hat{\mathbf{R}}_t' \hat{\Psi}_{S,t}) (\hat{\mathcal{I}}_{S,t} + \hat{\mathcal{P}}_{S,t})^{-1} / T.
\end{aligned}$$

### A.5.2 Proof of Lemma 1

First, define  $\mathcal{L} = \mathcal{I} + \mathcal{P}$ . Then, by the joint stable central limit theory in (Andersen et al., 2015a, Theorems 2 and 6), the vector  $((\hat{\mathbf{S}}_1 - \mathbf{S}_1)', (\hat{\mathbf{S}}_1 - \mathbf{S}_2)', \dots, (\hat{\mathbf{S}}_T - \mathbf{S}_T)', (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)')'$  converges stably in law to a mean zero mixed Gaussian distribution with (random) asymptotic covariance matrix

$$\mathbf{AVAR} \equiv \mathcal{L}^{-1} (\boldsymbol{\Omega} + \boldsymbol{\Psi}) \mathcal{L}^{-1},$$

when appropriately scaled by  $\sqrt{N}$  for the factor realizations and by  $\sqrt{NT}$  for the parameter vector. Next, by rules for the inverse of partitioned matrices, we can show that

$$\mathcal{L}^{-1} = \begin{pmatrix} \mathcal{L}_{11}^{-1} + \mathcal{L}_{11}^{-1} \mathcal{L}_{12} (\mathcal{L}_{22} - \mathcal{L}'_{21} \mathcal{L}_{11}^{-1} \mathcal{L}_{12})^{-1} \mathcal{L}'_{21} \mathcal{L}_{11}^{-1} & -\mathcal{L}_{11}^{-1} \mathcal{L}_{12} (\mathcal{L}_{22} - \mathcal{L}'_{21} \mathcal{L}_{11}^{-1} \mathcal{L}_{12})^{-1} \\ -(\mathcal{L}_{22} - \mathcal{L}'_{21} \mathcal{L}_{11}^{-1} \mathcal{L}_{12})^{-1} \mathcal{L}_{21} \mathcal{L}_{11}^{-1} & (\mathcal{L}_{22} - \mathcal{L}'_{21} \mathcal{L}_{11}^{-1} \mathcal{L}_{12})^{-1} \end{pmatrix},$$

since  $\mathcal{L}_{22} - \mathcal{L}'_{21} \mathcal{L}_{11}^{-1} \mathcal{L}_{12}$  is non-singular by A6 and A8. By definition,  $-\mathcal{L}_{11}^{-1} \mathcal{L}_{12} \equiv \mathbf{R}_0 / \sqrt{T}$ . Furthermore, we may write

$$\mathcal{L}_{22} - \mathcal{L}'_{21} \mathcal{L}_{11}^{-1} \mathcal{L}_{12} = \frac{1}{T} \sum_{t=1}^T \mathcal{I}_{\theta,t} - \mathcal{I}'_{S,\theta,t} (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathcal{I}_{S,\theta,t} = \frac{1}{T} \sum_{t=1}^T \mathcal{I}_{U,t} + \mathcal{P}_{\theta,t} \equiv \mathcal{I}_U + \mathcal{P}_\theta,$$

using simple algebra for the second equality. This implies

$$\mathcal{L}^{-1} = \begin{pmatrix} \mathcal{L}_{11}^{-1} + \mathbf{R}_0 (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \mathbf{R}_0' / T & \mathbf{R}_0 (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} / \sqrt{T} \\ (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \mathbf{R}_0' / \sqrt{T} & (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \end{pmatrix}.$$

Next, let us compute the  $q \times q$  partial asymptotic covariance matrix for  $\sqrt{TN}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ , which we denote by  $\mathbf{AVAR}_\theta$ . Using matrix multiplication and simple algebra,

$$\begin{aligned}
\mathbf{AVAR}_\theta &= (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \mathbf{R}_0' (\boldsymbol{\Omega}_{11} + \boldsymbol{\Psi}_{11}) \mathbf{R}_0 (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} / T + (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \mathbf{R}_0' \boldsymbol{\Omega}_{12} (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} / \sqrt{T} \\
&\quad + (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \boldsymbol{\Omega}_{21} \mathbf{R}_0 (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} / \sqrt{T} + (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \boldsymbol{\Omega}_{22} (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \\
&\equiv (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} (\boldsymbol{\Omega}_U + \boldsymbol{\Psi}_\theta) (\mathcal{I}_U + \mathcal{P}_\theta)^{-1}.
\end{aligned}$$

Similarly, for the partial asymptotic covariance matrix of  $\sqrt{N}((\hat{\mathbf{S}}_1 - \mathbf{S}_1)', (\hat{\mathbf{S}}_1 - \mathbf{S}_2)', \dots, (\hat{\mathbf{S}}_T - \mathbf{S}_T)')'$ , denoted by  $\mathbf{AVAR}_S$ , it follows that

$$\mathbf{AVAR}_S = \left( \mathcal{L}_{11}^{-1} + \mathbf{R}_0 (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \mathbf{R}_0' / T \right) (\boldsymbol{\Omega}_{11} + \boldsymbol{\Psi}_{11}) \left( \mathcal{L}_{11}^{-1} + \mathbf{R}_0 (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \mathbf{R}_0' / T \right)$$

$$\begin{aligned}
& + \mathbf{R}_0 (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \Omega_{21} \left( \mathcal{L}_{11}^{-1} + \mathbf{R}_0 (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \mathbf{R}'_0/T \right) / \sqrt{T} \\
& + \left( \mathcal{L}_{11}^{-1} + \mathbf{R}_0 (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \mathbf{R}'_0/T \right) \Omega_{12} (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \mathbf{R}'_0 / \sqrt{T} \\
& + \mathbf{R}_0 (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \Omega_{22} (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \mathbf{R}'_0/T \\
& = \mathcal{L}_{11}^{-1} (\Omega_{11} + \Psi_{11}) \mathcal{L}_{11}^{-1} + \mathcal{L}_{11}^{-1} \left( (\Omega_{11} \mathbf{R}_0 + \Omega_{12} \sqrt{T}) + \Psi_{11} \mathbf{R}_0 \right) (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \mathbf{R}'_0/T \\
& \quad + \mathbf{R}_0 (\mathcal{I}_U + \mathcal{P}_\theta)^{-1} \left( (\mathbf{R}'_0 \Omega_{11} + \Omega_{21} \sqrt{T}) + \mathbf{R}'_0 \Psi_{11} \right) \mathcal{L}_{11}^{-1}/T + \mathbf{R}_0 \times \mathbf{AVAR}_\theta \times \mathbf{R}'_0/T,
\end{aligned}$$

implying that we may write the asymptotic covariance matrix for  $\sqrt{N}(\hat{\mathbf{S}}_t - \mathbf{S}_t)$ ,  $t = 1, \dots, T$ , as

$$\mathbf{AVAR}_{S_t} = \mathbf{AVAR}_{S_{t,1}} + \mathbf{AVAR}_{S_{t,2}} + \mathbf{AVAR}_{S_{t,3}} + \mathbf{AVAR}_{S_{t,4}},$$

where  $\mathbf{AVAR}_{S_{t,j}}$ , for  $j = 1, 2, 3, 4$ , are defined in the lemma. Since  $\mathcal{L}_{11}$  is block-diagonal, this readily provides the representation of  $\mathbf{AVAR}_S$ , where  $\mathbf{ACOV}_S$  is defined from the off-diagonal block elements of matrix  $\mathbf{R}_0 \times \mathbf{AVAR}_\theta \times \mathbf{R}'_0/T$ , concluding the proof.  $\square$

### A.5.3 Proof of Theorem 4

Let us first consider the fixed  $T$  case for  $\hat{\boldsymbol{\theta}}$ , that is, the case when  $\gamma = 0$ . Here, we may readily invoke Lemma 1 and (Andersen et al., 2015a, Theorems 3 and 6) to show  $\hat{\boldsymbol{\mathcal{X}}} \xrightarrow{\mathbb{P}} \boldsymbol{\mathcal{X}}$  for  $\boldsymbol{\mathcal{X}} = \{\mathcal{I}_U, \mathcal{P}_\theta, \Omega_U, \Psi_\theta\}$  as  $N \rightarrow \infty$ . Next, for the bias correction, we need to show that  $\sqrt{T\bar{N}}\hat{\boldsymbol{\mathcal{B}}}_{T,N} = o_p(1)$ . However, since asymptotic negligibility of the various components of the bias correction term follows analogously, we will only show that  $\sqrt{T\bar{N}}\hat{\boldsymbol{\mathcal{B}}}_{T,N,1}^{(1)} = o_p(1)$ . For the latter, it suffices to prove

$$\frac{1}{N_t} \sum_{j=1}^{N_t} \left( \epsilon_{j,t} + \left( \kappa_j(\mathbf{S}_t, \boldsymbol{\theta}_0) - \kappa_j(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) \right) \right)^2 \left( \hat{\mathbf{U}}_{j,t}^{(S)} \left( \hat{\boldsymbol{\mathcal{I}}}_{S,t} + \hat{\boldsymbol{\mathcal{P}}}_{S,t} \right)^{-1} \hat{\mathbf{G}}_{j,t} \right) = O_p(1) \quad (57)$$

From (Andersen et al., 2015a, Theorem 1) we have  $\|\hat{\mathbf{S}}_t - \mathbf{S}_t\| \xrightarrow{\mathbb{P}} 0$ ,  $\forall t = 1, \dots, T$ , and  $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \xrightarrow{\mathbb{P}} 0$ . Therefore, we may work on a set where  $\|\hat{\mathbf{S}}_t - \mathbf{S}_t\| \leq \epsilon_1$  and  $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| \leq \epsilon_2$ , for some arbitrarily small  $\epsilon_1, \epsilon_2 > 0$ , which has probability approaching one. On this set, we have

$$\left| \kappa_j(\mathbf{S}_t, \boldsymbol{\theta}_0) - \kappa_j(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) \right|^2 \leq K f_d(\|\mathbf{S}_t\|) \|\hat{\mathbf{S}}_t - \mathbf{S}_t\|, \quad \left\| \hat{\mathbf{U}}_{j,t}^{(S)} - \mathbf{U}_{j,t}^{(S)} \right\|^2 \leq K f_d(\|\mathbf{S}_t\|) \|\hat{\mathbf{S}}_t - \mathbf{S}_t\|, \quad (58)$$

and, similarly,  $\|\hat{\mathbf{G}}_{j,t} - \mathbf{G}_{j,t}\|^2 \leq K f_d(\|\mathbf{S}_t\|) \|\hat{\mathbf{S}}_t - \mathbf{S}_t\|$  using A4 and A7. When we combine these inequalities with A5, A6, and A8, the consistency results for  $\hat{\boldsymbol{\mathcal{I}}}_{S,t}$  and  $\hat{\boldsymbol{\mathcal{P}}}_{S,t}$  in (Andersen et al., 2015a, Theorems 3 and 6), as well as the Cauchy-Schwarz inequality, this establishes (57), and the desired result  $\sqrt{T\bar{N}}\hat{\boldsymbol{\mathcal{B}}}_{T,N,1}^{(1)} = o_p(1)$  follows by  $T/N \rightarrow 0$ . Since the remaining bias correction terms are  $o_p(1)$  by similar arguments, the fixed  $T$  result follows by the stable central limit result in Lemma 1.

Next, for the large  $T$  case, that is, when  $\gamma > 0$ , we will show that

$$\sqrt{T\bar{N}} \left( \boldsymbol{\mathcal{B}}_{T,N,1}^{(1)} - \hat{\boldsymbol{\mathcal{B}}}_{T,N,1}^{(1)} \right) = o_p(1),$$

since consistency of the asymptotic covariance estimators,  $\hat{\mathcal{I}}_U$ ,  $\hat{\mathcal{P}}_\theta$ ,  $\hat{\Omega}_\theta$ , and  $\hat{\Psi}_\theta$  as well as of the remaining bias correction terms follow analogously. We make the following decomposition,

$$\begin{aligned} \sqrt{T\bar{N}} \left( \mathcal{B}_{T,N,1}^{(1)} - \hat{\mathcal{B}}_{T,N,1}^{(1)} \right) &\equiv \bar{\mathcal{B}}_{T,N,11}^{(1)} + \bar{\mathcal{B}}_{T,N,12}^{(1)}, \quad \text{where} \\ \bar{\mathcal{B}}_{T,N,11}^{(1)} &\equiv \frac{1}{\sqrt{T\bar{N}}} \sum_{t=1}^T \frac{\bar{N}}{N_t^2} \sum_{j=1}^{N_t} \left( \phi_{j,t} - \left( \tilde{\kappa}_{j,t} - \kappa_j(\hat{\mathcal{S}}, \hat{\theta}) \right)^2 \right) \left( \mathbf{U}_{j,t,0}^{(S)} \mathcal{L}_{S,t}^{-1} \mathbf{G}_{j,t,0} \right), \\ \bar{\mathcal{B}}_{T,N,12}^{(1)} &\equiv \frac{1}{\sqrt{T\bar{N}}} \sum_{t=1}^T \frac{\bar{N}}{N_t^2} \sum_{j=1}^{N_t} \left( \tilde{\kappa}_{j,t} - \kappa_j(\hat{\mathcal{S}}, \hat{\theta}) \right)^2 \left( \hat{\mathbf{U}}_{j,t,0}^{(S)} \hat{\mathcal{L}}_{S,t}^{-1} \hat{\mathbf{G}}_{j,t,0} - \mathbf{U}_{j,t,0}^{(S)} \mathcal{L}_{S,t}^{-1} \mathbf{G}_{j,t,0} \right), \end{aligned}$$

with  $\hat{\mathcal{L}}_{S,t} \equiv \hat{\mathcal{I}}_{S,t} + \hat{\mathcal{P}}_{S,t}$ . First, by applying the condition A2 and the continuous mapping theorem, we readily have  $\sup_{t \in \mathbb{R}_+} |\bar{N}/N_t - 1/\zeta_t| \xrightarrow{\mathbb{P}} 0$ . Next, we expand

$$\phi_{j,t} - \left( \tilde{\kappa}_{j,t} - \kappa_j(\hat{\mathcal{S}}, \hat{\theta}) \right)^2 = \left( \phi_{j,t} - \epsilon_{j,t}^2 \right) - \left( \kappa_j(\mathcal{S}, \theta) - \kappa_j(\hat{\mathcal{S}}, \hat{\theta}) \right)^2 + 2\epsilon_{j,t} \left( \kappa_j(\mathcal{S}, \theta) - \kappa_j(\hat{\mathcal{S}}, \hat{\theta}) \right), \quad (59)$$

and decompose  $\bar{\mathcal{B}}_{T,N,11}^{(1)} = \sum_{i=1}^3 \bar{\mathcal{B}}_{T,N,11i}^{(1)}$ , corresponding to the contribution from the three right-hand-side terms above. Since  $\bar{\mathcal{B}}_{T,N,111}^{(1)}$  is identical to  $\mathbf{A}_{1211}$  in the proof of Lemma 3,  $\|\bar{\mathcal{B}}_{T,N,111}^{(1)}\| \leq O_p(N^{-1})$  follows by the same arguments. Moreover, and similarly to the fixed  $T$  case, since  $\sup_{t=1,\dots,T} \|\hat{\mathcal{S}}_t - \mathcal{S}_t\| \xrightarrow{\mathbb{P}} 0$  and  $\hat{\theta}_t \xrightarrow{\mathbb{P}} \theta_0$  by Theorem 1, we may readily work on a set where  $\|\hat{\mathcal{S}}_t - \mathcal{S}_t\| \leq \epsilon_1$  and  $\|\hat{\theta}_t - \theta_0\| \leq \epsilon_2$ , for some arbitrarily small and positive  $\epsilon_1$  and  $\epsilon_2$ , which has probability approaching one. Hence, by applying A4, A6, and A7, we have

$$\left\| \bar{\mathcal{B}}_{T,N,112}^{(1)} \right\| \leq \sup_{t=1,\dots,T} \|\hat{\mathcal{S}}_t - \mathcal{S}_t\| \frac{K}{\sqrt{T\bar{N}}} \sum_{t=1}^T f_d(\|\mathcal{S}_t\|) \frac{\bar{N}}{N_t^2} \sum_{j=1}^{N_t} \left\| \mathbf{U}_{j,t,0}^{(S)} \mathcal{L}_{S,t}^{-1} \mathbf{G}_{j,t,0} \right\| = o_p\left((T/N)^{1/2}\right).$$

Moreover, by combining these two bounds with the Cauchy-Schwarz inequality,  $\|\bar{\mathcal{B}}_{T,N,113}^{(1)}\| = o_p(1)$ . As a result, we have  $\|\bar{\mathcal{B}}_{T,N,11}^{(1)}\| = o_p(1)$ . For the second main term,  $\bar{\mathcal{B}}_{T,N,12}^{(1)}$ , we may write

$$\begin{aligned} \hat{\mathbf{U}}_{j,t,0}^{(S)} \hat{\mathcal{L}}_{S,t}^{-1} \hat{\mathbf{G}}_{j,t,0} - \mathbf{U}_{j,t,0}^{(S)} \mathcal{L}_{S,t}^{-1} \mathbf{G}_{j,t,0} &= \left( \hat{\mathbf{U}}_{j,t,0}^{(S)} - \mathbf{U}_{j,t,0}^{(S)} \right) \hat{\mathcal{L}}_{S,t}^{-1} \hat{\mathbf{G}}_{j,t,0} + \mathbf{U}_{j,t,0}^{(S)} \left( \hat{\mathcal{L}}_{S,t}^{-1} - \mathcal{L}_{S,t}^{-1} \right) \hat{\mathbf{G}}_{j,t,0} \\ &\quad + \mathbf{U}_{j,t,0}^{(S)} \mathcal{L}_{S,t}^{-1} \left( \hat{\mathbf{G}}_{j,t,0} - \mathbf{G}_{j,t,0} \right) \end{aligned}$$

by addition and subtraction. Moreover, since we work on a set where  $\|\hat{\mathcal{S}}_t - \mathcal{S}_t\| \leq \epsilon_1$  and  $\|\hat{\theta}_t - \theta_0\| \leq \epsilon_2$ , we may invoke the inequalities in (58). When these inequalities are applied in conjunction with the decomposition in (59), the regularity conditions in A5, and part (a) of Lemma 9, we have

$$\left\| \bar{\mathcal{B}}_{T,N,12}^{(1)} \right\| \leq \sup_{t=1,\dots,T} \left( \|\hat{\mathcal{S}}_t - \mathcal{S}_t\| + \hat{\eta}_{t,N,n} \right) \frac{K}{\sqrt{T\bar{N}}} \sum_{t=1}^T f_d(\|\mathcal{S}_t\|) \frac{\bar{N}}{N_t^2} \sum_{j=1}^{N_t} \left\| \mathbf{U}_{j,t,0}^{(S)} \mathcal{L}_{S,t}^{-1} \mathbf{G}_{j,t,0} \right\|$$

where the nonnegative process  $\hat{\eta}_{t,N,n}$  satisfies  $\sup_{t=1,\dots,T} \hat{\eta}_{t,N,n} = o_p(1)$ . Hence,  $\|\bar{\mathcal{B}}_{T,N,12}^{(1)}\| \leq o_p(1)$ , as above, and, by the triangle inequality, this establishes

$$\sqrt{T\bar{N}} \left\| \mathcal{B}_{T,N,1}^{(1)} - \hat{\mathcal{B}}_{T,N,1}^{(1)} \right\| = o_p(1).$$



As consistency of the remaining bias and covariance estimators follow analogously, the large  $T$  result follows by combining Theorem 2, the continuous mapping theorem and Slutsky's theorem.

Finally, the feasible central limit theory for the state vector,  $\hat{\mathbf{S}}_t$ , follows, as above, by invoking Lemma 1 and the corresponding consistency results in (Andersen et al., 2015a, Theorem 3 and 6) for the components of  $\widehat{\mathbf{AVAR}}_{S_t}$  when  $\gamma = 0$ . Furthermore, since we may readily show that

$$\widehat{\mathbf{AVAR}}_{S_t,1} \xrightarrow{\mathbb{P}} \mathbf{AVAR}_{S_t,1}, \quad \sum_{i=2}^4 \widehat{\mathbf{AVAR}}_{S_t,i} = o_p(1),$$

when  $\gamma > 0$ , using the same arguments as for  $\hat{\boldsymbol{\theta}}$ , this concludes the proof.  $\square$

## A.6 Definitions for Section 5 and Proofs of Theorem 5 and Corollary 1

This section first defines the necessary asymptotic bias and variance terms to state the central limit theorem for the weighted PLS case, and then proceeds to establish Theorem 5 and Corollary 1.

### A.6.1 Definitions for Section 5

Let us define the weighted PLS analogues of the various components in Theorem 2. First, write,

$$\mathcal{P}_{S_t}^w \equiv (1/\lambda w(2V_t^2))\mathcal{P}_{S_t}, \quad \Psi_{S_t}^w \equiv (1/\lambda w(2V_t^2))^2\Psi_{S_t}, \quad \Omega_{S_t}^w = \Omega_{S_t}/w(\phi_t)^2, \quad \boldsymbol{\chi}_t^w \equiv \boldsymbol{\chi}_t/w(\phi_t),$$

for  $\boldsymbol{\chi}_t = \{\mathcal{I}_{\theta,t}, \mathcal{I}_{S,t}, \mathcal{I}_{S,\theta,t}\}$ , and  $\mathbf{R}_{t,0}^w \equiv -(\mathcal{I}_{S,t}^w + \mathcal{P}_{S_t}^w)^{-1}\mathcal{I}_{S,\theta,t}^w$ . Then, let us define,

$$\mathbf{U}_t^w(k, \tau, \mathbf{Z}, \boldsymbol{\theta}) = \mathbf{D}(k, \tau, \mathbf{Z}, \boldsymbol{\theta}) + (\mathbf{R}_{t,0}^w)' \mathbf{G}(k, \tau, \mathbf{Z}, \boldsymbol{\theta}),$$

such that we may, correspondingly, write,

$$\begin{aligned} \mathcal{I}_{U,t}^w &= \frac{1}{w(\phi_t)} \sum_{\tau} \pi_t^{\tau} \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \mathbf{U}_t^w(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0) \mathbf{U}_t^w(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0)' dk, \\ \Omega_{U,t}^w &= \frac{1}{\varsigma_t w(\phi_t)^2} \sum_{\tau} \pi_t^{\tau} \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \phi_{t,k,\tau} \mathbf{U}_t^w(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0) \mathbf{U}_t^w(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0)' dk, \\ \Omega_{S,U,t}^w &= \frac{1}{\varsigma_t w(\phi_t)^2} \sum_{\tau} \pi_t^{\tau} \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \phi_{t,k,\tau} \mathbf{G}(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0) \mathbf{U}_t^w(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0)' dk, \end{aligned}$$

as well as the matrices  $\mathcal{P}_{\theta,t}^w = (\mathbf{R}_{t,0}^w)' \mathcal{P}_{S_t}^w \mathbf{R}_{t,0}^w$  and  $\Psi_{\theta,t}^w = (\mathbf{R}_{t,0}^w)' \Psi_{S_t}^w \mathbf{R}_{t,0}^w$ . Moreover, let

$$\boldsymbol{\chi}^w = \text{plim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \boldsymbol{\chi}_t^w, \quad \boldsymbol{\chi}^w = \{\mathcal{I}_U^w, \mathcal{P}_{\theta}^w, \Omega_U^w, \Psi_{\theta}^w\}.$$

Next, let us denote the  $q \times p$  derivative of  $\mathbf{U}_t^w(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$  with respect to  $\mathbf{Z}$  as  $\mathbf{U}_t^{(S,w)}(k, \tau, \mathbf{Z}, \boldsymbol{\theta})$ . Similar to the case without weighting, when a (higher-order) gradient is evaluated at  $(\mathbf{S}'_t, \boldsymbol{\theta}'_0)$ , we use

the notation  $\mathbf{U}_{j,t,0}^w$ . Finally, the corresponding asymptotic bias term may be defined as

$$\begin{aligned}
\mathbf{B}_{T,N}^w &= (\mathcal{I}_U^w + \mathcal{P}_\theta^w)^{-1} \times \sum_{j=1}^2 \left( \mathbf{B}_{T,N,1}^{(j,w)} - \mathbf{B}_{T,N,2}^{(j,w)} - \mathbf{B}_{T,N,3}^{(j,w)} \right), \quad \text{where} \\
\mathbf{B}_{T,N,1}^{(1,w)} &= \frac{1}{TN} \sum_{t=1}^T \frac{N}{N_t} \frac{1}{w(\phi_t)^2 N_t} \sum_{j=1}^{N_t} \phi_{j,t} \left( \mathbf{U}_{j,t,0}^{(S,w)} (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} \mathbf{G}_{j,t,0} \right), \\
\mathbf{B}_{T,N,2}^{(1,w)} &= \frac{1}{TN} \sum_{t=1}^T \frac{1}{w(\phi_t) N_t} \sum_{j=1}^{N_t} \mathbf{U}_{j,t,0}^{(S,w)} (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} (\boldsymbol{\Omega}_{S,t}^w + \boldsymbol{\Psi}_{S,t}^w) (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} \mathbf{G}_{j,t,0}, \\
\mathbf{B}_{T,N,3}^{(1,w)} &= \frac{1}{2TN} \sum_{t=1}^T \frac{1}{w(\phi_t) N_t} \sum_{j=1}^{N_t} \mathbf{U}_{j,t,0}^w \times \\
&\quad \left\{ \frac{N}{(w(\phi_t) N_t)^2} \sum_{\ell=1}^{N_t} \phi_{\ell,t} \left( \mathbf{G}'_{\ell,t,0} (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} \mathbf{G}_{j,t,0}^{(S)} (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} \mathbf{G}_{\ell,t,0} \right) \right. \\
&\quad \left. + \frac{2V_t^2}{w(2V_t^2)^2} \frac{k_n N}{N_t^2} \left( \mathbf{C}'_{t,0} (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} \mathbf{G}_{j,t,0}^{(S)} (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} \mathbf{C}_{t,0} \right) \right\}, \\
\mathbf{B}_{T,N,1}^{(2,w)} &= \frac{1}{TN} \sum_{t=1}^T \frac{2V_t^2}{w(2V_t^2)^2} \frac{k_n N}{N_t^2} (\mathbf{R}_{t,0}^w)' \mathbf{C}_{t,0}^{(S)} (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} \mathbf{C}_{t,0}, \\
\mathbf{B}_{T,N,2}^{(2,w)} &= \frac{1}{TN} \sum_{t=1}^T \frac{k_n}{w(2V_t^2) N_t} (\mathbf{R}_{t,0}^w)' \mathbf{C}_{t,0}^{(S)} (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} (\boldsymbol{\Omega}_{S,t}^w + \boldsymbol{\Psi}_{S,t}^w) (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} \mathbf{C}_{t,0}, \\
\mathbf{B}_{T,N,3}^{(2,w)} &= \frac{1}{2TN} \sum_{t=1}^T \frac{k_n}{w(2V_t^2) N_t} (\mathbf{R}_{t,0}^w)' \mathbf{C}_{t,0} \times \\
&\quad \left\{ \frac{N}{(w(\phi_t) N_t)^2} \sum_{\ell=1}^{N_t} \phi_{\ell,t} \left( \mathbf{G}'_{\ell,t,0} (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} \mathbf{C}_{t,0}^{(S)} (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} \mathbf{G}_{\ell,t,0} \right) \right. \\
&\quad \left. + \frac{2V_t^2}{w(2V_t^2)^2} \frac{k_n N}{N_t^2} \left( \mathbf{C}'_{t,0} (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} \mathbf{C}_{t,0}^{(S)} (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} \mathbf{C}_{t,0} \right) \right\}.
\end{aligned}$$

### A.6.2 Proof of Theorem 5

First, let us make the decomposition

$$\hat{\phi}_t - \frac{1}{N_t} \sum_{j=1}^{N_t} \phi_{j,t} = \hat{\phi}_t^{(1)} + \hat{\phi}_t^{(2)} + \hat{\phi}_t^{(3)}, \quad \hat{\phi}_t^{(1)} = \frac{1}{N_t} \sum_{j=1}^{N_t} (\epsilon_{j,t}^2 - \phi_{j,t}), \quad (60)$$

$$\hat{\phi}_t^{(2)} = \frac{2}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} (\kappa_{j,t}(\mathbf{S}_t, \boldsymbol{\theta}_0) - \kappa_{j,t}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}})), \quad \hat{\phi}_t^{(3)} = \frac{1}{N_t} \sum_{j=1}^{N_t} (\kappa_{j,t}(\mathbf{S}_t, \boldsymbol{\theta}_0) - \kappa_{j,t}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}))^2. \quad (61)$$

Then, using a second-order Taylor expansion, the regularity conditions in A7, the uniform consistency bounds for the state vector  $\|\hat{\mathbf{S}}_t - \mathbf{S}_t\| \leq \hat{\eta}_{t,N,n} f_d(\|\mathbf{S}_t\|) / \sqrt{N}$ , where  $\sup_{t=1, \dots, T} \hat{\eta}_{t,N,n} = O_p(1)$ , as well

as  $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p(1/\sqrt{NT})$  in conjunction with A5, we may readily conclude

$$|\hat{\phi}_t^{(2)}| + |\hat{\phi}_t^{(3)}| \leq \hat{\eta}_{t,N,n} f_d(\|\mathbf{S}_t\|)/N^{1-\iota}, \quad (62)$$

for some arbitrarily small  $\iota > 0$ . We now make the same decomposition of terms as in the proof of Theorem 2, but we divide  $\mathbf{A}_1$ ,  $\mathbf{A}_2$ , and the first part of  $\mathbf{A}_4$  by  $w(\hat{\phi}_t)$ , and for  $\mathbf{A}_3$  as well as the second part of  $\mathbf{A}_4$ , we replace  $\lambda$  with  $1/w(2(\hat{V}_t^n)^2)$ . The corresponding terms are denoted with hats.

To proceed, we will make use of the following algebraic inequality, which follows from a second-order Taylor expansion and the properties of the weight function  $w$ :

$$\left| \frac{1}{w(y)} - \frac{1}{w(x)} + \frac{w'(x)}{w(x)^2}(y-x) \right| \leq K|y-x|^2, \quad x, y \in \mathbb{R}_+. \quad (63)$$

Specifically, we will apply this inequality with  $y = \hat{\phi}_t$  and  $x = \frac{1}{N_t} \sum_{j=1}^{N_t} \phi_{j,t}$  for  $\hat{\mathbf{A}}_1$ ,  $\hat{\mathbf{A}}_2$  and the first part of  $\hat{\mathbf{A}}_4$ , and, similarly, with  $y = 2(\hat{V}_t^n)^2$  and  $x = 2V_t^2$  for  $\hat{\mathbf{A}}_3$  as well as the second part of  $\hat{\mathbf{A}}_4$ . If we denote with  $\mathbf{A}_1^w$  and  $\mathbf{A}_2^w$  the analogues of  $\hat{\mathbf{A}}_1$  and  $\hat{\mathbf{A}}_2$ , respectively, in which  $w(\hat{\phi}_t)$  is replaced with  $w(\frac{1}{N_t} \sum_{j=1}^{N_t} \phi_{j,t})$ , and, similarly, with  $\mathbf{A}_3^w$  and  $\mathbf{A}_4^w$  the analogues of  $\hat{\mathbf{A}}_3$  and  $\hat{\mathbf{A}}_4$  in which  $w(2(\hat{V}_t^n)^2)$  is also replaced with  $w(2V_t^2)$ , then we only need to show the asymptotic negligibility of  $\sqrt{NT}(\hat{\mathbf{A}}_i - \mathbf{A}_i^w)$  for  $i = 1, 2, 3, 4$ . The analysis of the four terms follows by the above bounds, by applying Lemmas 8 and 10 as well as by using arguments similar to those for Theorem 2. The only nontrivial part is the analysis of the differences  $\sqrt{NT}(\hat{\mathbf{A}}_{11} - \mathbf{A}_{11}^w)$  and  $\sqrt{NT}(\hat{\mathbf{A}}_{31} - \mathbf{A}_{31}^w)$ . In particular, using the algebraic inequality in (63), the nontrivial part of the first of them is

$$\frac{1}{\sqrt{NT}} \sum_{t=1}^T \frac{N}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{U}_{j,t,0} \frac{w'(\frac{1}{N_t} \sum_{j=1}^{N_t} \phi_{j,t})}{w^2(\frac{1}{N_t} \sum_{j=1}^{N_t} \phi_{j,t})} \hat{\phi}_t^{(1)},$$

and, by applying A5, the boundedness of  $w$  from below, the boundedness of  $w'$ , as well as the definition of  $\hat{\phi}_t^{(1)}$ , we have that its  $\mathcal{F}^{(0)}$ -conditional variance is bounded by  $\frac{K}{NT} \sum_{t=1}^T \frac{N^2}{N_t^4} \sum_{j=1}^{N_t} \|\mathbf{U}_{j,t,0}\|^2$ . This readily shows asymptotic negligibility of the term.

Next, the nontrivial component of  $\sqrt{NT}(\hat{\mathbf{A}}_{31} - \mathbf{A}_{31}^w)$  is

$$4\lambda V_t \frac{\sqrt{N}}{\sqrt{T}} \sum_{t=1}^T \frac{k_n}{N_t} \mathbf{R}'_{t,0} \mathbf{C}_{t,0} \frac{w'(2V_t^2)}{w^2(2V_t^2)} (\hat{V}_t^n - V_t)^2.$$

Then, by applying Lemma 10 in conjunction with A2 for the ratio  $N_t/N$ , we easily get that this term is of order  $O_p(1/k_n)$ , establishing asymptotic negligibility, thereby concluding the proof.  $\square$

### A.6.3 Proof of Corollary 1

First, since the support of  $\phi_t$  and the process  $V_t$  is above  $\epsilon > 0$ , we can equivalently work with weight function  $w(x) = x$ . For this choice of  $w$ , we have

$$\mathcal{P}_{S,t}^w = \frac{\rho}{2\zeta_t V_t^2} \mathbf{C}_{t,0} \mathbf{C}'_{t,0}, \quad \mathcal{I}_{U,t}^w = \frac{1}{\phi_t} \sum_{\tau} \pi_{\tau} \int_{\underline{k}(t,\tau)}^{\bar{k}(t,\tau)} \frac{1}{\psi_{t,\tau}(k)} \mathbf{U}_t^w(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0) \mathbf{U}_t^w(k, \tau, \mathbf{S}_t, \boldsymbol{\theta}_0)' dk,$$

as well as  $\mathcal{I}_{S,t}^w = \mathcal{I}_{S,t}/\phi_t$ ,  $\Psi_{S,t}^w = \mathcal{P}_{S,t}^w/\varsigma_t$ ,  $\Omega_{S,t}^w = \mathcal{I}_{S,t}^w/\varsigma_t$  and  $\Omega_{U,t}^w = \mathcal{I}_{U,t}^w/\varsigma_t$ . Then, using A2 and these definitions, we have

$$\begin{aligned}\mathcal{B}_{T,N,1}^{(1,w)} &= \frac{1}{TN} \sum_{t=1}^T \frac{1}{\varsigma_t \phi_t N_t} \sum_{j=1}^{N_t} \mathbf{U}_{j,t,0}^{(S,w)} (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} \mathbf{G}_{j,t,0} + o_p((TN)^{-1/2}), \\ \mathcal{B}_{T,N,2}^{(1,w)} &= \frac{1}{TN} \sum_{t=1}^T \frac{1}{\varsigma_t \phi_t N_t} \sum_{j=1}^{N_t} \mathbf{U}_{j,t,0}^{(S,w)} (\mathcal{I}_{S,t}^w + \mathcal{P}_{S,t}^w)^{-1} \mathbf{G}_{j,t,0},\end{aligned}$$

implying that their contribution to the asymptotic bias will cancel, and, similarly, for the corresponding terms  $\mathcal{B}_{T,N,1}^{(2,w)} - \mathcal{B}_{T,N,2}^{(2,w)}$ . Hence, the final result follows by setting  $\varsigma_t = 1$  and applying Theorem 5 as well as the definitions in Section A.6.1 with  $w(x) = x$ , concluding the proof.  $\square$

## A.7 Auxiliary Results

In what follows, we will use the simplifying notation  $W_t \equiv W_t^{\mathbb{P}}$  and write  $\mathbb{E}_i^n(\cdot)$  for  $\mathbb{E}(\cdot | \mathcal{F}_{i\Delta_n})$ .

**Lemma 7.** *Let A1, A2 and A5 hold. Assume  $N \rightarrow \infty$ ,  $T \rightarrow \infty$  with  $T/N \rightarrow \gamma$  for some finite  $\gamma \geq 0$ ,  $\Delta_n \rightarrow 0$  and  $k_n \rightarrow \infty$ . In addition, let*

$$\varpi \in \left( \frac{1}{2(2-r)}, \frac{1}{2} \right), \quad k_n \sqrt{\Delta_n} \rightarrow 0, \quad \sqrt{T k_n} (k_n \Delta_n \vee \Delta_n^{(2-r)\varpi}) \rightarrow 0. \quad (64)$$

Denote a  $p_1$ -dimensional vector  $\mathbf{M}_{j,t}$ , and a  $p_2$ -dimensional vector  $\mathbf{M}_t$ , with  $p_1, p_2 \geq 1$ , which are both adapted to  $\mathcal{F}_t^{(0)}$ , for  $j = 1, \dots, N_t$  and  $t \in \mathbb{N}$ . Assume we have

$$\frac{1}{T} \sum_{t=1}^T \frac{N}{N_t^2} \sum_{j=1}^{N_t} \phi_{j,t} \mathbf{M}_{j,t} \mathbf{M}'_{j,t} \xrightarrow{\mathbb{P}} \mathcal{M}_1, \quad \frac{1}{T} \sum_{t=1}^T 2V_t^2 \mathbf{M}_t \mathbf{M}'_t \xrightarrow{\mathbb{P}} \mathcal{M}_2, \quad (65)$$

and further  $\|\mathbf{M}_{j,t}\| \leq \overline{M}_t$  with  $\sup_t \mathbb{E}|\overline{M}_t|^{4+\iota} < \infty$ , and  $\sup_t \mathbb{E}\|\mathbf{M}_t\|^{4+\iota} < \infty$  for some  $\iota > 0$ . Then

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \left( \frac{\sqrt{N}}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{M}_{j,t} \right) \xrightarrow{\mathcal{L}-s} \mathcal{M}^{1/2} \times \mathbf{Y}, \quad \mathcal{M} = \begin{pmatrix} \mathcal{M}_1 & \mathbf{0}_{p_1 \times p_2} \\ \mathbf{0}_{p_2 \times p_1} & \mathcal{M}_2 \end{pmatrix} \quad (66)$$

where  $\mathbf{Y}$  is a  $(p_1 + p_2) \times 1$  vector of standard Gaussian random variables, independent of each other, defined on an extension of the original probability space and independent of  $\mathcal{F}$ .

*Proof.* We start with some preliminary estimates for  $\hat{V}_t^n - V_t$ . First, by application of Itô's lemma, we have

$$dx_t = \tilde{\alpha}_t dt + \sqrt{V_t} dW_t + \int_{x>-1} \log(1+x) \tilde{\mu}^{\mathbb{P}}(ds, dx), \quad (67)$$

where

$$\tilde{\alpha}_t = \alpha_t - \frac{1}{2} V_t - a_t \int_{x>-1} (\log(1+x) - x) \nu^{\mathbb{P}}(dx). \quad (68)$$

In what follows, we denote the continuous and discontinuous parts of  $x$  by  $x^c$  and  $x^d$ . We next decompose  $\hat{V}_t^n - V_t$  into

$$\hat{V}_t^n - V_t = \hat{V}_t^{n,1} + \hat{V}_t^{n,2} + \hat{V}_t^{n,3}, \quad (69)$$

$$\hat{V}_t^{n,1} = \frac{n}{k_n} \sum_{i \in I_t^n} \left( |\Delta_i^n x|^2 \mathbf{1}_{\{|\Delta_i^n x| \leq \alpha \Delta_n^\varpi\}} - V_{(i-1)\Delta_n} |\Delta_i^n W|^2 \right), \quad (70)$$

$$\hat{V}_t^{n,2} = \frac{1}{k_n} \sum_{i \in I_t^n} (V_{(i-1)\Delta_n} - V_t), \quad \hat{V}_t^{n,3} = \frac{1}{k_n} \sum_{i \in I_t^n} (n |\Delta_i^n W|^2 - 1) V_{(i-1)\Delta_n}. \quad (71)$$

For  $\hat{V}_t^{n,1}$ , we use the following algebraic inequality

$$\begin{aligned} \left| |\Delta_i^n x|^2 \mathbf{1}_{\{|\Delta_i^n x| \leq \alpha \Delta_n^\varpi\}} - |\Delta_i^n x^c|^2 \right| &\leq 2 |\Delta_i^n x^c|^2 (\mathbf{1}_{\{|\Delta_i^n x^c| \geq \frac{\alpha}{2} \Delta_n^\varpi\}} + \mathbf{1}_{\{|\Delta_i^n x^d| \geq \frac{\alpha}{2} \Delta_n^\varpi\}}) \\ &+ 2 |\Delta_i^n x^d|^2 (\mathbf{1}_{\{|\Delta_i^n x^c| \geq \alpha \Delta_n^\varpi\}} + \mathbf{1}_{\{|\Delta_i^n x^d| \leq 2\alpha \Delta_n^\varpi\}}) + 2 |\Delta_i^n x^c| |\Delta_i^n x^d| \mathbf{1}_{\{|\Delta_i^n x^d| \leq 2\alpha \Delta_n^\varpi\}}. \end{aligned} \quad (72)$$

Hence, using bounds for increments of Itô semimartingales (the Burkholder-Davis-Gundy inequality and Lemmas 2.15-2.18 in Jacod and Protter (2012)) and the Hölder inequality, we have

$$\mathbb{E}_{i-1}^n \left| |\Delta_i^n x|^2 \mathbf{1}_{\{|\Delta_i^n x| \leq \alpha \Delta_n^\varpi\}} - |\Delta_i^n x^c|^2 \right|^p \leq K_{t,n} \Delta_n^{1+(2p-r)\varpi}, \quad \forall p \geq 1. \quad (73)$$

$$|\mathbb{E}_{i-1}^n (|\Delta_i^n x^c|^2 - V_{(i-1)\Delta_n} (\Delta_i^n W)^2)| \leq K_{t,n} \Delta_n^2, \quad (74)$$

$$\mathbb{E}_{i-1}^n \left| |\Delta_i^n x^c|^2 - V_{(i-1)\Delta_n} (\Delta_i^n W)^2 \right|^p \leq K_{t,n} \Delta_n^{p+1-\iota}, \quad \forall p \geq 2 \text{ and } \forall \iota > 0, \quad (75)$$

where  $\mathbb{E}|K_{t,n}|^q < K$  for some sufficiently high  $q$ . Using these inequalities and successively applying the Burkholder-Davis-Gundy inequality, we have for some arbitrary small  $\iota > 0$ ,

$$|\mathbb{E}_t(\hat{V}_t^{n,1})| \leq K_{t,n} \Delta_n^{(2-r)\varpi}, \quad \mathbb{E}_t |\hat{V}_t^{n,1}| \leq K_{t,n} \Delta_n^{\frac{1}{2}-\iota}, \quad \mathbb{E}_t |\hat{V}_t^{n,1}|^p \leq K_{t,n} \frac{\Delta_n^{1+(2p-r)\varpi-p}}{k_n^{p-1}}, \quad \forall p \geq 2, \quad (76)$$

where, again,  $\mathbb{E}|K_{t,n}|^q < K$  for some sufficiently high  $q$ .

Turning next to  $\hat{V}_t^{n,2}$ , we may write

$$\hat{V}_t^{n,2} = \sum_{j=1}^{k_n-1} \frac{k_n-j}{k_n} (V_{(tn+j)\Delta_n} - V_{(tn+j-1)\Delta_n}). \quad (77)$$

Now, by using the inequalities

$$\mathbb{E} |\mathbb{E}_{i-1}^n (V_{i\Delta_n} - V_{(i-1)\Delta_n})|^p \leq K \Delta_n^p, \quad \mathbb{E} |V_{i\Delta_n} - V_{(i-1)\Delta_n}|^p \leq K \Delta_n^{\frac{p}{2} \wedge 1}, \quad \forall p \geq 1, \quad (78)$$

we, then, have by application of the Burkholder-Davis-Gundy inequality and Jensen's inequality that

$$|\mathbb{E}_t(\hat{V}_t^{n,2})| \leq K_{t,n} k_n \Delta_n, \quad \mathbb{E}_t |\hat{V}_t^{n,2}|^p \leq K_{t,n} (k_n \Delta_n)^{\frac{p}{2} \wedge 1}, \quad \forall p > 0, \quad (79)$$

where  $\mathbb{E}|K_{t,n}|^q < K$  for some sufficiently high  $q$ .

By combining the bounds in (76) and (79), and upon using successive conditioning and Cauchy-Schwartz inequality, we have

$$\frac{\sqrt{k_n}}{\sqrt{T}} \sum_{t=1}^T \mathbb{E}_t(\hat{V}_t^{n,1}) \mathbf{M}_t = O_p \left( \sqrt{T k_n} \Delta_n^{(2-r)\varpi} \right), \quad \frac{\sqrt{k_n}}{\sqrt{T}} \sum_{t=1}^T \mathbb{E}_t(\hat{V}_t^{n,2}) \mathbf{M}_t = O_p \left( \sqrt{T k_n} k_n \Delta_n \right), \quad (80)$$

$$\frac{\sqrt{k_n}}{\sqrt{T}} \sum_{t=1}^T [(\hat{V}_t^{n,1} - \mathbb{E}_t(\hat{V}_t^{n,1})) \mathbf{M}_t] = O_p \left( \Delta_n^{\frac{(4-r)\varpi}{2} - \frac{1}{2}} \right), \quad (81)$$

$$\frac{\sqrt{k_n}}{\sqrt{T}} \sum_{t=1}^T [(\hat{V}_t^{n,2} - \mathbb{E}_t(\hat{V}_t^{n,2})) \mathbf{M}_t] = O_p \left( k_n \sqrt{\Delta_n} \right). \quad (82)$$

Given our restrictions on  $\varpi$  and  $k_n$  in (64), the above quantities are all asymptotically negligible. Hence, we are left with the analysis of  $\sum_{t=1}^T \boldsymbol{\chi}_t$ , where we set

$$\boldsymbol{\chi}_t = \frac{1}{\sqrt{T}} \left( \frac{\sqrt{N}}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{M}_{j,t} \quad \sqrt{k_n} \hat{V}_t^{n,3} \mathbf{M}_t \right)',$$

and note that  $\boldsymbol{\chi}_t$  is  $\mathcal{F}_{t+1/2}$ -adapted, provided  $n$  is sufficiently big, and, additionally,  $\mathbb{E} \|\boldsymbol{\chi}_t\| = O_p(1)$ . We will show that  $\sum_{t=1}^T \boldsymbol{\chi}_t \xrightarrow{\mathcal{L}-s} \mathcal{M}^{1/2} \times \mathbf{Y}$ . For this, we will make use of Theorem VIII.5.42 of Jacod and Shiryaev (2003). First, however, we define  $\mathcal{F}_t^T = \mathcal{F}_{t+1/2}^{(0)} \times \mathcal{F}_t^{(1)}$ , for  $t = 0, 1, \dots, T$ , and  $\mathcal{G}_k^T = \mathcal{F}_{[kT]}^T$ , for  $k \in [0, 1]$  ( $\mathcal{F}_k^T$  should also contain subscripts  $N$  and  $n$ , but since we are considering a joint limit in which both  $T \rightarrow \infty$  as well as  $n \rightarrow \infty$  and  $N \rightarrow \infty$ , we will suppress this dependence). For  $n$  sufficiently high, we have  $\boldsymbol{\chi}_t \in \mathcal{F}_t^T$ . In addition,  $\mathcal{G}_{\alpha_T}^T \subset \mathcal{G}_{\alpha_{T+1}}^{T+1}$  for  $\alpha_T = 1/\sqrt{T}$ . Finally, we trivially have  $\mathcal{F} = \vee_T \mathcal{G}_{\alpha_T}^T$ . This implies that the nesting condition VIII.5.37 in Jacod and Shiryaev (2003) on the filtration is satisfied. Hence, using Theorem VIII.5.42 of Jacod and Shiryaev (2003) (applied for a fixed time  $t = 1$  and with  $X_1^n$  in Theorem VIII.5.42 of Jacod and Shiryaev (2003) set equal to  $\sum_{t=1}^T \boldsymbol{\chi}_t$ ), we only need to show that the following conditions hold:

$$\begin{cases} \sum_{t=2}^T \mathbb{E}(\boldsymbol{\chi}_t | \mathcal{F}_{t-1}^T) \xrightarrow{\mathbb{P}} \mathbf{0}, \\ \sum_{t=2}^T \{ \mathbb{E}(\boldsymbol{\chi}_t \boldsymbol{\chi}_t' | \mathcal{F}_{t-1}^T) - \mathbb{E}(\boldsymbol{\chi}_t | \mathcal{F}_{t-1}^T) \mathbb{E}(\boldsymbol{\chi}_t' | \mathcal{F}_{t-1}^T) \} \xrightarrow{\mathbb{P}} \mathcal{M}, \\ \sum_{t=2}^T \mathbb{E}(\|\boldsymbol{\chi}_t\|^{2+\iota} | \mathcal{F}_{t-1}^T) \xrightarrow{\mathbb{P}} 0, \quad \text{for some } \iota > 0. \end{cases} \quad (83)$$

The first of these results holds trivially as we have

$$\frac{\sqrt{N}}{N_t} \sum_{j=1}^{N_t} \mathbb{E}(\epsilon_{j,t} \mathbf{M}_{j,t} | \mathcal{F}^{(0)}) = \mathbf{0}, \quad \mathbb{E}(\hat{V}_t^{n,3} \mathbf{M}_t | \mathcal{F}_t) = \mathbf{0}, \quad (84)$$

and  $\epsilon_{j,t}$  are  $\mathcal{F}^{(0)}$ -conditionally independent.

Next, using successive conditioning, the bound  $\mathbb{E} |\mathbb{E}_s(V_t^2 - V_s^2)|^2 \leq K|t-s|^2$  for  $0 \leq s \leq t$  and the Cauchy-Schwartz inequality, we have

$$\sum_{t=2}^T \mathbb{E}_{t-1/2} (\boldsymbol{\chi}_t \boldsymbol{\chi}_t')$$

$$= \frac{1}{T} \sum_{t=2}^T \begin{pmatrix} \mathbb{E}_{t-1/2} \left( \frac{N}{N_t^2} \sum_{j=1}^{N_t} (\phi_{j,t} \mathbf{M}_{j,t} \mathbf{M}'_{j,t}) \right) & \mathbf{0}_{p_1 \times p_2} \\ \mathbf{0}_{p_2 \times p_1} & 2\mathbb{E}_{t-1/2} (V_t^2 \mathbf{M}_t \mathbf{M}'_t) \end{pmatrix} + O_p(k_n \Delta_n). \quad (85)$$

The second limit result in (83), then, follows from the convergence in probability result in (65), assumed in the lemma, and the following two bounds

$$\begin{aligned} \frac{1}{T} \sum_{t=2}^T \left[ \frac{N}{N_t^2} \sum_{j=1}^{N_t} (\phi_{j,t} \mathbf{M}_{j,t} \mathbf{M}'_{j,t}) - \mathbb{E}_{t-1/2} \left( \frac{N}{N_t^2} \sum_{j=1}^{N_t} (\phi_{j,t} \mathbf{M}_{j,t} \mathbf{M}'_{j,t}) \right) \right] &= O_p \left( \frac{1}{\sqrt{T}} \right), \\ \frac{1}{T} \sum_{t=2}^T (V_t^2 \mathbf{M}_t \mathbf{M}'_t - \mathbb{E}_{t-1/2} (V_t^2 \mathbf{M}_t \mathbf{M}'_t)) &= O_p \left( \frac{1}{\sqrt{T}} \right), \end{aligned} \quad (86)$$

which, in turn, follow from bounding their respective second moments.

We turn next to the the third limit in (83). First, since  $\epsilon_{j,t}$  are  $\mathcal{F}^{(0)}$ -conditionally independent and centered from A5, we have, by applying successive conditioning and using the integrability assumption for  $\mathbf{M}_{j,t}$  of the lemma, that

$$\mathbb{E} \left\| \frac{1}{\sqrt{N_t}} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{M}_{j,t} \right\|^4 \leq \mathbb{E} \left( \frac{K}{N_t} \sum_{j=1}^{N_t} \phi_{j,t}^{(4)} \|\mathbf{M}_{j,t}\|^4 \right) \leq K. \quad (87)$$

Moreover, by the Burkholder-Davis-Gundy inequality and inequality in means, we also have

$$\mathbb{E}_t |\sqrt{k_n} \hat{V}_t^{n,3}|^p \leq \frac{K}{k_n} \sum_{i \in I_t^n} |V_{(i-1)\Delta_n}|^p, \quad \forall p \geq 2, \quad (88)$$

and, hence,

$$\mathbb{E}_t \|\sqrt{k_n} \hat{V}_t^{n,3} \mathbf{M}_t\|^4 \leq \frac{K}{k_n} \sum_{i \in I_t^n} V_{(i-1)\Delta_n}^4 \|\mathbf{M}_t\|^4. \quad (89)$$

Combining the above two bounds, we have  $\mathbb{E} \|\chi_t\|^4 \leq K/T^2$ , from which the third limit in (83) follows. From here, by combining (80)-(82) and (83), we get the convergence result in (66).  $\square$

**Lemma 8.** *Let A1, A2 and A5 hold. Assume  $N \rightarrow \infty$ ,  $T \rightarrow \infty$  with  $T/N \rightarrow \gamma$  for some finite  $\gamma \geq 0$ ,  $\Delta_n \rightarrow 0$  and  $k_n \rightarrow \infty$ . In addition, let  $\varpi \in \left(\frac{1}{4-r}, \frac{1}{2}\right)$  and  $k_n \sqrt{\Delta_n} \rightarrow 0$ . Denote  $p_1 \times p_2$  matrices  $\mathbf{M}_{j,t}$  and  $\mathbf{M}_t$ , with  $p_1, p_2 \geq 1$ , which are adapted to  $\mathcal{F}_t^{(0)}$ , for  $j = 1, \dots, N_t$  and  $t \in \mathbb{N}$  and for which  $\|\mathbf{M}_{j,t}\| \leq \bar{\mathbf{M}}_t$  with  $\sup_t \mathbb{E} \|\bar{\mathbf{M}}_t\|^{4+\iota} < \infty$  and  $\sup_t \mathbb{E} \|\mathbf{M}_t\|^{4+\iota} < \infty$  for some  $\iota > 0$ . Then, we have,*

- (a)  $\sup_{t=1, \dots, T} \left\| N_t^{-1} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{M}_{j,t} \right\| \xrightarrow{\mathbb{P}} 0$ ;
- (b)  $\sup_{t=1, \dots, T} \left\| \lambda \frac{k_n}{N_t} \left( \hat{V}_t^n - V_t \right) \mathbf{M}_t \right\| \xrightarrow{\mathbb{P}} 0$ ;
- (c)  $|\mathbb{E}(\hat{V}_t^n - V_t)| \leq K_{t,n} (\Delta_n^{(2-r)\varpi} \vee k_n \Delta_n)$  for  $K_{t,n}$  adapted to  $\mathcal{F}_t^{(0)}$  and  $\mathbb{E} |K_{t,n}|^p < K_p$  for every  $p > 0$  and finite  $K_p$ .

*Proof. Part (a).* For arbitrary  $\epsilon > 0$ , it readily follows that

$$\mathbb{P} \left( \sup_{t=1, \dots, T} \left\| N_t^{-1} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{M}_{j,t} \right\| > \epsilon \middle| \mathcal{F}_T^{(0)} \right) \leq \sum_{t=1}^T \mathbb{P} \left( \left\| N_t^{-1} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{M}_{j,t} \right\| > \epsilon \middle| \mathcal{F}_T^{(0)} \right). \quad (90)$$

Then, by applying Chebyshev's inequality, the Burkholder-Davis-Gundy inequality, and using the fact that the random variables  $\epsilon_{j,t}$  are  $\mathcal{F}^{(0)}$ -conditionally independent, we have

$$\mathbb{P} \left( \left\| N_t^{-1} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{M}_{j,t} \right\| > \epsilon \middle| \mathcal{F}_T^{(0)} \right) \leq \frac{K}{N_t^3} \sum_{j=1}^{N_t} \phi_{j,t}^{(4)} \|\mathbf{M}_{j,t}\|^4. \quad (91)$$

From here, applying the integrability assumptions for  $\|\mathbf{M}_{j,t}\|$  in the lemma and for  $\phi_{j,t}^{(4)}$  in A5, we get the desired result.

**Part (b).** Similarly to the proof of part (a), we have for arbitrary  $\epsilon > 0$

$$\mathbb{P} \left( \sup_{t=1, \dots, T} \left\| \lambda \frac{k_n}{N_t} (\hat{V}_t^n - V_t) \mathbf{M}_t \right\| > \epsilon \right) \leq \sum_{t=1}^T \mathbb{P} \left( \lambda \frac{k_n}{N_t} |\hat{V}_t^n - V_t| \|\mathbf{M}_t\| > \epsilon \right). \quad (92)$$

Now, using the decomposition of  $\hat{V}_t^n - V_t = \hat{V}_t^{n,1} + \hat{V}_t^{n,2} + \hat{V}_t^{n,3}$ , introduced in the proof Lemma 7, we may write

$$\mathbb{P} \left( \lambda \frac{k_n}{N_t} |\hat{V}_t^n - V_t| \|\mathbf{M}_t\| > \epsilon \right) \leq \sum_{j=1}^3 \mathbb{P} \left( \lambda \frac{k_n}{N_t} |\hat{V}_t^{n,j}| \|\mathbf{M}_t\| > \frac{\epsilon}{3} \right). \quad (93)$$

Then, by invoking the bounds derived in the proof of Lemma 7, Chebyshev's inequality, and the fact that  $k_n/N = O(1)$  as well as  $T/N$  is asymptotically bounded, we have

$$\mathbb{P} \left( \lambda \frac{k_n}{N_t} |\hat{V}_t^{n,1}| \|\mathbf{M}_t\| > \frac{\epsilon}{3} \right) \leq K \left( \frac{\Delta_n^{(4-r)\varpi-1}}{T} \right), \quad (94)$$

$$\mathbb{P} \left( \lambda \frac{k_n}{N_t} |\hat{V}_t^{n,2}| \|\mathbf{M}_t\| > \frac{\epsilon}{3} \right) \leq K \frac{k_n^2 \Delta_n}{T}, \quad (95)$$

$$\mathbb{P} \left( \lambda \frac{k_n}{N_t} |\hat{V}_t^{n,3}| \|\mathbf{M}_t\| > \frac{\epsilon}{3} \right) \leq \frac{K}{T^{1+\iota/2}}, \quad \text{for some } \iota > 0, \quad (96)$$

where the constant  $K$  in the above inequalities depends on  $\epsilon$ . From here, taking into account of the fact that  $k_n \sqrt{\Delta_n} \rightarrow 0$  and  $\varpi > \frac{1}{4-r}$ , the result in part (b) follows.

**Part (c).** The result follows immediately from combining the bounds in (80) and (84) in the proof of the previous lemma.  $\square$

**Lemma 9.** *Under the conditions of Theorem 2, and for some positive-valued process  $\{\hat{\eta}_{t,N,n}\}_{t \geq 0}$  with  $\sup_{t=1, \dots, T} \hat{\eta}_{t,N,n} = o_p(1)$  as well as positive  $K$  and  $d$ :*



(a) For some  $\tilde{\mathbf{S}}_1, \tilde{\mathbf{S}}_2, \dots, \tilde{\mathbf{S}}_T$  and  $\tilde{\boldsymbol{\theta}}$  with  $\sup_{t=1, \dots, T} \|\tilde{\mathbf{S}}_t - \mathbf{S}_t\| = o_p(1)$  and  $\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = o_p(1)$ , we have

$$\left\| \left( \hat{\mathcal{I}}_{S,t}(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) + \hat{\mathcal{P}}_{S,t}(\tilde{\mathbf{S}}_t) \right)^{-1} - \mathcal{L}_{S,t}^{-1} \right\| \leq K \hat{\eta}_{t,N,n} f_d(\|\mathbf{S}_t\|).$$

(b) For  $\bar{\mathcal{L}}_{S,t} = \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{G}_{j,t,0} \mathbf{G}'_{j,t,0} - \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{G}_{j,t,0}^{(S)} + \lambda \frac{k_n}{N_t} \mathbf{C}_{t,0} \mathbf{C}'_{t,0} - \lambda \frac{k_n}{N_t} \mathbf{C}_{t,0}^{(S)} (\hat{V}_t^n - V_t)$ , we have

$$\left\| \bar{\mathcal{L}}_{S,t}^{-1} - \mathcal{L}_{S,t}^{-1} \right\| \leq K \left( \frac{\hat{\eta}_{t,N,n}}{\sqrt{N}} + \left| \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{G}_{j,t,0}^{(S)} \right| + \lambda \frac{k_n}{N_t} |\hat{V}_t^n - V_t| \right) f_d(\|\mathbf{S}_t\|).$$

(c) For any  $\tilde{\boldsymbol{\theta}}$  such that  $\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = o_p(1)$ , we have

$$\left\| \frac{\partial \hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'} - \mathbf{R}_{t,0} \right\| \leq K \hat{\eta}_{t,N,n} f_d(\|\mathbf{S}_t\|).$$

*Proof. Part (a).* Given the uniform consistency of  $(\tilde{\mathbf{S}}_1, \dots, \tilde{\mathbf{S}}_T)$ , it is no restriction to work on a set (whose probability approaches 1) on which  $\sup_{t=1, \dots, T} \|\tilde{\mathbf{S}}_t - \mathbf{S}_t\| \leq \epsilon$  for some arbitrarily small  $\epsilon > 0$ . We will assume this to be the case for the rest of the proof, without further mentioning. Also, in the proof,  $\hat{\eta}_{t,N,n}$  will be a generic positive-valued sequence with  $\sup_{t=1, \dots, T} \hat{\eta}_{t,N,n} = o_p(1)$ , and which may change from line to line. Similarly,  $K$  and  $d$  are positive numbers, which may change from line to line.

We start by defining  $\hat{\mathcal{I}}_{S,t}(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) = \hat{\mathcal{I}}_{S,t,1} + \hat{\mathcal{I}}_{S,t,2}$  and  $\hat{\mathcal{P}}_{S,t}(\tilde{\mathbf{S}}_t) = \hat{\mathcal{P}}_{S,t,1} + \hat{\mathcal{P}}_{S,t,2}$ . Then, using the triangle inequality, the difference  $\hat{\mathcal{I}}_{S,t,1} - \mathcal{I}_{S,t}$  may be bounded as

$$\begin{aligned} \left\| \hat{\mathcal{I}}_{S,t,1} - \mathcal{I}_{S,t} \right\| &\leq \left\| \frac{1}{N_t} \sum_{j=1}^{N_t} \mathbf{G}_{j,t,0} \mathbf{G}'_{j,t,0} - \mathcal{I}_{S,t} \right\| + \left\| \frac{2}{N_t} \sum_{j=1}^{N_t} \left( \mathbf{G}_j(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) - \mathbf{G}_{j,t,0} \right) \mathbf{G}'_{j,t,0} \right\| \\ &\quad + \left\| \frac{1}{N_t} \sum_{j=1}^{N_t} \left( \mathbf{G}_j(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) - \mathbf{G}_{j,t,0} \right) \left( \mathbf{G}_j(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) - \mathbf{G}_{j,t,0} \right)' \right\| \\ &\equiv \hat{\mathcal{I}}_{S,t,11} + \hat{\mathcal{I}}_{S,t,12} + \hat{\mathcal{I}}_{S,t,13}. \end{aligned} \tag{97}$$

First, for  $\hat{\mathcal{I}}_{S,t,11}$ , using A2 and A7, we have  $\|\hat{\mathcal{I}}_{S,t,11}\| \leq \frac{K}{\sqrt{N}} \hat{\eta}_{t,N,n} f_d(\|\mathbf{S}_t\|)$ . Applying, again, A7 and the uniform consistency of  $(\tilde{\mathbf{S}}_1, \dots, \tilde{\mathbf{S}}_T)$  and  $\tilde{\boldsymbol{\theta}}$ , we have  $\|\hat{\mathcal{I}}_{S,t,12}\| + \|\hat{\mathcal{I}}_{S,t,13}\| \leq K \hat{\eta}_{t,N,n} f_d(\|\mathbf{S}_t\|)$ .

Next, using the triangle inequality, we bound  $\hat{\mathcal{I}}_{S,t,2}$  as

$$\begin{aligned} \left\| \hat{\mathcal{I}}_{S,t,2} \right\| &\leq \left\| \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{G}_{j,t,0}^{(S)} \right\| + \left\| \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \left( \mathbf{G}_j^{(S)}(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) - \mathbf{G}_{j,t,0}^{(S)} \right) \right\| \\ &\quad + \left\| \frac{1}{N_t} \sum_{j=1}^{N_t} \left( \kappa_j(\mathbf{S}_t, \boldsymbol{\theta}_0) - \kappa_j(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) \right) \mathbf{G}_j^{(S)}(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) \right\| \\ &\equiv \hat{\mathcal{I}}_{S,t,21} + \hat{\mathcal{I}}_{S,t,22} + \hat{\mathcal{I}}_{S,t,23}. \end{aligned} \tag{98}$$

From Lemma 8 (a), we readily have  $\sup_{t=1,\dots,T} \|\hat{\mathcal{I}}_{S,t,21}\| = o_p(1)$ . Next, for  $\hat{\mathcal{I}}_{S,t,22}$ , we may use A4 and A7, to bound

$$\|\hat{\mathcal{I}}_{S,t,22}\| \leq \frac{K}{N_t} \sum_{j=1}^{N_t} |\epsilon_{j,t}| f_d(\|\mathbf{S}_t\|) \|\tilde{\mathbf{S}}_t - \mathbf{S}_t\|.$$

Then, by applying Burkholder-Davis-Gundy inequality as well as using the integrability conditions for a given  $\phi_{j,t}^{(k)}$  in A5 and for the state vector  $\mathbf{S}_t$  in A6, we have

$$\sup_{t=1,\dots,T} \left| \frac{1}{N_t} \sum_{j=1}^{N_t} (|\epsilon_{j,t}| - \phi_{j,t}^{(1)}) f_d(\|\mathbf{S}_t\|) \right| = o_p(1),$$

the proof of which parallels that of Lemma 8 (a). Hence, using, again, A7 and uniform consistency of the state vectors  $(\tilde{\mathbf{S}}_1, \dots, \tilde{\mathbf{S}}_T)$ , we have, altogether, that  $\|\hat{\mathcal{I}}_{S,t,22}\| \leq K \hat{\eta}_{t,N,n} f_d(\|\mathbf{S}_t\|)$ . By applying almost identical arguments, we also have  $\|\hat{\mathcal{I}}_{S,t,23}\| \leq K \hat{\eta}_{t,N,n} f_d(\|\mathbf{S}_t\|)$ .

We continue with  $\hat{\mathcal{P}}_{S,t,1} - \mathcal{P}_{S,t}$ , which, using the triangle inequality, may be bounded as

$$\begin{aligned} \left\| \hat{\mathcal{P}}_{S,t,1} - \mathcal{P}_{S,t} \right\| &\leq \left\| \lambda \left( \frac{k_n}{N_t} - \frac{\rho}{\varsigma_t} \right) \mathbf{C}_{t,0} \mathbf{C}'_{t,0} \right\| + 2 \left\| \lambda \frac{k_n}{N_t} \mathbf{C}_{t,0} \left( \mathbf{C}(\tilde{\mathbf{S}}_t) - \mathbf{C}_{t,0} \right)' \right\| \\ &\quad + \left\| \lambda \frac{k_n}{N_t} \left( \mathbf{C}(\tilde{\mathbf{S}}_t) - \mathbf{C}_{t,0} \right) \left( \mathbf{C}(\tilde{\mathbf{S}}_t) - \mathbf{C}_{t,0} \right)' \right\| \\ &\equiv \hat{\mathcal{P}}_{S,t,11} + \hat{\mathcal{P}}_{S,t,12} + \hat{\mathcal{P}}_{S,t,13}. \end{aligned} \quad (99)$$

First, using A2 and A7, we have  $\|\hat{\mathcal{P}}_{S,t,11}\| \leq \frac{K}{\sqrt{N}} \hat{\eta}_{t,N,n} f_d(\|\mathbf{S}_t\|)$ , and by applying A7 as well as the uniform consistency of  $(\tilde{\mathbf{S}}_1, \dots, \tilde{\mathbf{S}}_T)$ , we get  $\|\hat{\mathcal{P}}_{S,t,12}\| + \|\hat{\mathcal{P}}_{S,t,13}\| \leq K \hat{\eta}_{t,N,n} f_d(\|\mathbf{S}_t\|)$ .

Next, for  $\hat{\mathcal{P}}_{S,t,2}$ , we have by the triangle inequality that

$$\begin{aligned} \left\| \hat{\mathcal{P}}_{S,t,2} \right\| &\leq \lambda \frac{k_n}{N_t} \left\| \left( \mathbf{C}^{(S)}(\tilde{\mathbf{S}}_t) - \mathbf{C}_{t,0}^{(S)} \right) \left( \hat{V}_t^n - \xi_1(\tilde{\mathbf{S}}_t) \right) \right\| \\ &\quad + \lambda \frac{k_n}{N_t} \left\| \mathbf{C}_{t,0}^{(S)} \left( \hat{V}_t^n - V_t \right) \right\| + \lambda \frac{k_n}{N_t} \left\| \mathbf{C}_{t,0}^{(S)} \left( \xi_1(\mathbf{S}_t) - \xi_1(\tilde{\mathbf{S}}_t) \right) \right\| \\ &\equiv \hat{\mathcal{P}}_{S,t,21} + \hat{\mathcal{P}}_{S,t,22} + \hat{\mathcal{P}}_{S,t,23}. \end{aligned} \quad (100)$$

Then, invoking, once more, uniform consistency of  $(\tilde{\mathbf{S}}_1, \dots, \tilde{\mathbf{S}}_T)$ , part (b) of Lemma 8 and A7, it readily follows that  $\|\hat{\mathcal{P}}_{S,t,21}\| + \|\hat{\mathcal{P}}_{S,t,22}\| + \|\hat{\mathcal{P}}_{S,t,23}\| \leq K \hat{\eta}_{t,N,n} f_d(\|\mathbf{S}_t\|)$ .

Combining the bounds above, we have

$$\left\| \hat{\mathcal{I}}_{S,t}(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) + \hat{\mathcal{P}}_{S,t}(\tilde{\mathbf{S}}_t) - \mathcal{L}_{S,t} \right\| \leq K \hat{\eta}_{t,N,n} f_d(\|\mathbf{S}_t\|). \quad (101)$$

Hence, taking into account the lower bound restriction on the eigenvalues of  $\mathcal{L}_{S,t}$  in A8, we get, for some sufficiently big integer  $q > 0$ ,

$$\left\| \mathcal{L}_{S,t}^{-q} \left( \hat{\mathcal{I}}_{S,t}(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) + \hat{\mathcal{P}}_{S,t}(\tilde{\mathbf{S}}_t) \right) - \mathcal{L}_{S,t}^{-q+1} \right\| \leq \hat{\eta}_{t,N,n}. \quad (102)$$

From here, upon using the triangle inequality, we have

$$\left\| \mathcal{L}_{S,t}^{-q+1} \right\| - \hat{\eta}_{t,N,n} \leq \left\| \mathcal{L}_{S,t}^{-q} \left( \hat{\mathcal{I}}_{S,t}(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) + \hat{\mathcal{P}}_{S,t}(\tilde{\mathbf{S}}_t) \right) \right\| \leq \left\| \mathcal{L}_{S,t}^{-q+1} \right\| + \hat{\eta}_{t,N,n}. \quad (103)$$

This implies that the minimum eigenvalue of  $\mathcal{L}_{S,t}^{-q}(\hat{\mathcal{I}}_{S,t}(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) + \hat{\mathcal{P}}_{S,t}(\tilde{\mathbf{S}}_t))$  is bounded from below by  $K f_d(\|\mathbf{S}_t\|) - \hat{\eta}_{t,N,n}$ . Hence, on a set with probability approaching one, we have

$$\left\| \mathcal{L}_{S,t}^q \left( \hat{\mathcal{I}}_{S,t}(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) + \hat{\mathcal{P}}_{S,t}(\tilde{\mathbf{S}}_t) \right)^{-1} \right\| \leq K f_d(\|\mathbf{S}_t\|).$$

To proceed, we make use of the following identity  $\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{A}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{B}^{-1}$  for two square invertible matrices  $\mathbf{A}$  and  $\mathbf{B}$ . By applying this with the inequality above, it follows that

$$\left\| \mathcal{L}_{S,t}^q \left( \hat{\mathcal{I}}_{S,t}(\tilde{\mathbf{S}}_t, \tilde{\boldsymbol{\theta}}) + \hat{\mathcal{P}}_{S,t}(\tilde{\mathbf{S}}_t) \right)^{-1} - \mathcal{L}_{S,t}^{-1+q} \right\| \leq K \hat{\eta}_{t,N,n} f_d(\|\mathbf{S}_t\|). \quad (104)$$

Finally, by using  $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \times \|\mathbf{B}\|$ , we get the desired result.

**Part (b).** From the bounds in the proof of part (a), we have

$$\|\bar{\mathcal{L}}_{S,t} - \mathcal{L}_{S,t}\| \leq K \left( \frac{\hat{\eta}_{t,N,n}}{\sqrt{N}} + \left| \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{G}_{j,t,0}^{(S)} \right| + \lambda \frac{k_n}{N_t} |\hat{V}_t^n - V_t| \right) f_d(\|\mathbf{S}_t\|). \quad (105)$$

Hence, with  $\bar{\mathbf{A}}_{t,N,n} = \mathcal{L}_{S,t}^{-1} \bar{\mathcal{L}}_{S,t} - \mathbf{I}_p$ , we get, using A7,

$$\|\bar{\mathbf{A}}_{t,N,n}\| \leq K \left( \frac{\hat{\eta}_{t,N,n}}{\sqrt{N}} + \left| \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{G}_{j,t,0}^{(S)} \right| + \lambda \frac{k_n}{N_t} |\hat{V}_t^n - V_t| \right) f_d(\|\mathbf{S}_t\|). \quad (106)$$

Now, by denoting the eigenvalues of  $\mathbf{I}_p + \bar{\mathbf{A}}_{t,N,n}$  with  $\bar{\lambda}_i^n$ , for  $i = 1, \dots, p$ , and applying Gershgorin's circle theorem, we have

$$\max_{i=1, \dots, p} |\bar{\lambda}_i^n - 1| \leq K \left( \frac{\hat{\eta}_{t,N,n}}{\sqrt{N}} + \left| \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{G}_{j,t,0}^{(S)} \right| + \lambda \frac{k_n}{N_t} |\hat{V}_t^n - V_t| \right) f_d(\|\mathbf{S}_t\|),$$

and, using parts (a) and (b) of Lemma 8 for  $\sup_{t=1, \dots, T} \left| \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{G}_{j,t,0}^{(S)} \right|$  and  $\sup_{t=1, \dots, T} |\hat{V}_t^n - V_t|$  as well as the integrability conditions for  $\mathbf{S}_t$  (so that we can make the inversion), the same holds true for eigenvalues of  $|\mathbf{I}_p + \bar{\mathbf{A}}_{t,N,n}|^{-1}$ . Hence, the eigenvalues of  $|\mathbf{I}_p + \bar{\mathbf{A}}_{t,N,n}|^{-1} - \mathbf{I}_p$  are bounded in absolute value by  $K \left( \frac{\hat{\eta}_{t,N,n}}{\sqrt{N}} + \left| \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{G}_{j,t,0}^{(S)} \right| + \lambda \frac{k_n}{N_t} |\hat{V}_t^n - V_t| \right) f_d(\|\mathbf{S}_t\|)$ , and, as a result,

$$\left\| |\mathbf{I}_p + \bar{\mathbf{A}}_{t,N,n}|^{-1} - \mathbf{I}_p \right\| \leq K \left( \frac{\hat{\eta}_{t,N,n}}{\sqrt{N}} + \left| \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{G}_{j,t,0}^{(S)} \right| + \lambda \frac{k_n}{N_t} |\hat{V}_t^n - V_t| \right) f_d(\|\mathbf{S}_t\|). \quad (107)$$

From here, the desired result readily follows.

**Part (c).** First, as for the proof of consistency of the parameter and state vector estimates, we have

$$\sup_{t=1,\dots,T} \|\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}}) - \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)\| = o_p(1). \quad (108)$$

Given the above result, it is no restriction to work on a set (whose probability approaches 1) on which  $\sup_{t=1,\dots,T} \|\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}}) - \mathbf{S}_t\| \leq \epsilon$  for some arbitrarily small  $\epsilon > 0$ . We will assume this to be the case for the remainder of the proof, without further mentioning.

By using the first-order condition,  $\mathbf{M}_t(\tilde{\boldsymbol{\theta}}) = \mathbf{0}$  in (47), then  $\partial \hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})/\partial \boldsymbol{\theta}'$  is recovered by the implicit function theorem, that is, from  $\partial \mathbf{M}_t(\tilde{\boldsymbol{\theta}})/\partial \boldsymbol{\theta}' = \mathbf{0}$ . In particular, the latter provides

$$\left( \hat{\mathcal{I}}_{S,t}(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}}), \tilde{\boldsymbol{\theta}}) + \hat{\mathcal{P}}_{S,t}(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})) \right) \frac{\partial \hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'} = \hat{\mathcal{I}}_{S,\theta,t}(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}}), \tilde{\boldsymbol{\theta}}). \quad (109)$$

From part (a), we have

$$\left\| \left( \hat{\mathcal{I}}_{S,t}(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}}), \tilde{\boldsymbol{\theta}}) + \hat{\mathcal{P}}_{S,t}(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}})) \right)^{-1} - \left( \hat{\mathcal{I}}_{S,t}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0), \boldsymbol{\theta}_0) + \hat{\mathcal{P}}_{S,t}(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0)) \right)^{-1} \right\| \leq K \hat{\eta}_{t,N,n} f_d(\|\mathbf{S}_t\|).$$

Furthermore, using A7 and arguments similar to those for the proof of part (a), we have

$$\left\| \hat{\mathcal{I}}_{S,\theta,t}(\hat{\mathbf{S}}_t(\tilde{\boldsymbol{\theta}}), \tilde{\boldsymbol{\theta}}) - \mathcal{I}_{S,\theta,t} \right\| \leq K f_{2d}(\|\mathbf{S}_t\|) \hat{\eta}_{t,N,n}. \quad (110)$$

By combining the results above with part (a) of the lemma, we easily get the claim.  $\square$

**Lemma 10.** *Assume the conditions of Theorem 2 hold. Denote  $p_1 \times p_2$  matrices  $\mathbf{M}_{j,t}$  and  $\mathbf{M}_t$ , with  $p_1, p_2 \geq 1$ , which are adapted to  $\mathcal{F}_t^{(0)}$ , for  $j = 1, \dots, N_t$  and  $t \in \mathbb{N}$  and for which  $\|\mathbf{M}_{j,t}\| \leq \bar{\mathbf{M}}_t$  with  $\sup_t \mathbb{E}|\bar{\mathbf{M}}_t|^q < \infty$  and  $\sup_t \mathbb{E}(\|\mathbf{M}_t\|^q) < \infty$  for every  $q > 0$ . Then, for some,  $r_1, r_2, r_3 \in \mathbb{N}^3$ , we have*

$$\begin{aligned} \frac{N^{(r_1+r_2+r_3)/2}}{T} \sum_{t=1}^T \left\| \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t \right\|^{r_1} \times \left\| \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{M}_{j,t} \right\|^{r_2} \times \left\| \lambda \frac{k_n}{N_t} \mathbf{M}_t (\hat{V}_t^n - V_t) \right\|^{r_3} \\ = O_p \left( k_n^{1+\frac{r_1+r_3}{2}} \Delta_n \vee 1 \right). \end{aligned} \quad (111)$$

*Proof.* First, using Lemma 9 (a) and  $\sup_{t=1,\dots,T} (k_n/N_t) = O_p(1)$ , we can bound the stated expression by  $(A_1 + A_2) \times O_p(1)$  where

$$\begin{aligned} A_1 &= \frac{1}{T} \sum_{t=1}^T \left\| (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \right\|^{r_1} \times \left\| \sqrt{N} \mathbf{F}_{t,N,n} \right\|^{r_1} \times \left\| \frac{1}{\sqrt{N_t}} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{M}_{j,t} \right\|^{r_2} \times \left\| \sqrt{k_n} (\hat{V}_t^n - V_t) \mathbf{M}_t \right\|^{r_3}, \\ A_2 &= \sup_{t=1,\dots,T} |\hat{\eta}_{t,N,n}|^{r_1} \\ &\quad \times \frac{1}{T} \sum_{t=1}^T \left\| \sqrt{N} \mathbf{F}_{t,N,n} \right\|^{r_1} \times \left\| \frac{1}{\sqrt{N_t}} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{M}_{j,t} \right\|^{r_2} \times \left\| \sqrt{k_n} (\hat{V}_t^n - V_t) \mathbf{M}_t \right\|^{r_3} \times f_d(\|\mathbf{S}_t\|). \end{aligned}$$

Now, since  $\sup_{t=1,\dots,T} |\hat{\eta}_{t,N,n}|^{r_1} = o_p(1)$ , and the analysis of  $A_2$  is done exactly the same way as that of  $A_1$ , we can focus on the latter henceforth. By using the triangle inequality in conjunction with the definition of  $\mathbf{F}_{t,N,n}$ , we may further bound  $A_1 \leq (A_{11} + A_{12}) \times O_p(1)$ , where

$$\begin{aligned} A_{11} &= \frac{1}{T} \sum_{t=1}^T a_t^1 \times \left\| \frac{1}{\sqrt{N_t}} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{G}_{j,t,0} \right\|^{r_1} \times \left\| \frac{1}{\sqrt{N_t}} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{M}_{j,t} \right\|^{r_2} \times \left\| \sqrt{k_n} (\hat{V}_t^n - V_t) \right\|^{r_3}, \\ A_{12} &= \frac{1}{T} \sum_{t=1}^T a_t^2 \times \left\| \frac{1}{\sqrt{N_t}} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{M}_{j,t} \right\|^{r_2} \times \left\| \sqrt{k_n} (\hat{V}_t^n - V_t) \right\|^{r_1+r_3}, \end{aligned}$$

and where we use the shorthand notation

$$a_t^1 = \|(\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1}\|^{r_1} \|\mathbf{M}_t\|^{r_3} \quad \text{and} \quad a_t^2 = \|(\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1}\|^{r_1} \|\mathbf{C}_{t,0}\|^{r_1} \|\mathbf{M}_t\|^{r_3}.$$

Then, by applying the Burkholder-Davis-Gundy inequality, inequality in means as well as the regularity conditions in A5, we have for  $r \geq 2$

$$\begin{aligned} \mathbb{E}_0 \left\| \frac{1}{\sqrt{N_t}} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{M}_{j,t} \right\|^r &\leq K \mathbb{E}_0 \left( \frac{1}{N_t} \sum_{j=1}^{N_t} \epsilon_{j,t}^2 \|\mathbf{M}_{j,t}\|^2 \right)^{r/2} \\ &\leq K \mathbb{E}_0 \left( \frac{1}{N_t} \sum_{j=1}^{N_t} |\epsilon_{j,t}|^r \|\mathbf{M}_{j,t}\|^r \right) = \frac{K}{N_t} \sum_{j=1}^{N_t} \phi_{j,t}^{(r)} \|\mathbf{M}_{j,t}\|^r, \end{aligned} \tag{112}$$

and, for  $r = 1$ , by Jensen's inequality

$$\mathbb{E}_0 \left\| \frac{1}{\sqrt{N_t}} \sum_{j=1}^{N_t} \epsilon_{j,t} \mathbf{M}_{j,t} \right\| \leq K \sqrt{\frac{1}{N_t} \sum_{j=1}^{N_t} \phi_{j,t} \|\mathbf{M}_{j,t}\|^2}. \tag{113}$$

Furthermore, given the integrability assumptions for  $\mathbf{M}_{j,t}$  and the ones for  $\phi_{j,t}^{(k)}$  in A5, we have

$$\mathbb{E} \left( \frac{1}{N_t} \sum_{j=1}^{N_t} \phi_{j,t}^{(r)} \|\mathbf{M}_{j,t}\|^r \right)^m \leq K_{m,r}, \quad \forall m > 0, \quad r \geq 2, \tag{114}$$

where  $K_{m,r} > 0$  denotes some finite constant depending on the value of  $m$  and  $r$ .

Next, using the triangle inequality and the decomposition of  $\hat{V}_t^n - V_t$  from the proof of Lemma 7, we have  $|\hat{V}_t^n - V_t|^q \leq K(|\hat{V}_t^{n,1}|^q + |\hat{V}_t^{n,2}|^q + |\hat{V}_t^{n,3}|^q)$ . Then, we may readily use the bounds from the proof of Lemma 7 to show that for some arbitrary small  $\iota > 0$ , it follows that

$$\mathbb{E}_t |\sqrt{k_n} \hat{V}_t^{n,1}|^q \leq K_{t,n}^{(1)} \begin{cases} \sqrt{k_n} \Delta_n^{\frac{1}{2}-\iota}, & \text{if } q = 1, \\ k_n^{1-\frac{q}{2}} \Delta_n^{1-q+(2q-r)\varpi}, & \text{if } q \geq 2, \end{cases}$$

$$\mathbb{E}_t |\sqrt{k_n} \hat{V}_t^{n,2}|^q \leq K_{t,n}^{(2)} \begin{cases} k_n \sqrt{\Delta_n}, & \text{if } q = 1, \\ k_n^{1+q/2} \Delta_n, & \text{if } q \geq 2, \end{cases}$$

$$\mathbb{E}_t |\sqrt{k_n} \hat{V}_t^{n,3}|^q \leq K_{t,n}^{(3)},$$

where  $K_{t,n}^{(1)}$ ,  $K_{t,n}^{(2)}$  and  $K_{t,n}^{(3)}$  are random variables adapted to  $\mathcal{F}_t^{(0)}$  and having finite moments of all orders. Now, since, from the assumptions of Theorem 2, we have  $\frac{\Delta_n^{(2-\frac{\varpi}{2})\varpi-1}}{k_n} \rightarrow 0$  and  $k_n \sqrt{\Delta_n} \rightarrow 0$ , it follows that

$$\mathbb{E}_t |\sqrt{k_n} (\hat{V}_t^n - V_t)|^q \leq K_{t,n} \begin{cases} 1, & \text{if } q = 1, \\ k_n^{1+q/2} \Delta_n, & \text{if } q \geq 2, \end{cases} \quad (115)$$

where  $K_{t,n}$  is a random variable adapted to  $\mathcal{F}_t^{(0)}$  and having finite moments of all orders. By combining (112)-(114) and (115), using the fact that  $a_t^1$  and  $a_t^2$  have finite moments of all orders, and by applying Hölder's inequality, we get the desired result.  $\square$

**Lemma 11.** *Assume the conditions of Theorem 2 hold. Then, for some,  $r_1, r_2 \in \mathbb{N}^2$ ,*

$$\frac{N^{(r_1+r_2)/2}}{T} \sum_{t=2}^T \left\| \hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t \right\|^{r_1} \times \left\| \hat{\mathbf{S}}_{t-1}(\boldsymbol{\theta}_0) - \mathbf{S}_{t-1} \right\|^{r_2} = O_p \left( k_n^{1+\frac{r_1+r_2}{2}} \Delta_n \vee 1 \right). \quad (116)$$

*Proof.* This follows by arguments almost identical to those for Lemma 10.  $\square$

**Lemma 12.** *Assume the conditions of Theorem 2 hold. Then, we have*

$$\frac{1}{T} \sum_{t=1}^T \left\{ N(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)(\hat{\mathbf{S}}_t(\boldsymbol{\theta}_0) - \mathbf{S}_t)' \right. \\ \left. - (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} (\boldsymbol{\Omega}_{S,t} + \boldsymbol{\Psi}_{S,t}) (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \right\} \mathbf{M}_t = o_p(1), \quad (117)$$

$$\frac{1}{T} \sum_{t=1}^T \left\{ k_n (\hat{V}_t^n - V_t)^2 - 2V_t^2 \right\} \mathbf{M}_t = o_p(1), \quad (118)$$

where  $\mathbf{M}_t$  is  $\mathcal{F}_t^{(0)}$ -adapted matrix, with dimension  $p \times p_1$  for the first result and of dimension  $p_1 \times p_2$  for the second result, where  $p_1, p_2 \geq 1$ , and further  $\sup_t \mathbb{E}(\|\mathbf{M}_t\|^q) < \infty$  for every  $q > 0$ .

*Proof.* We start with (118). First, decompose the term on the left-hand side of (118) into  $\mathbf{C}_1$  and  $\mathbf{C}_2$ , where

$$\mathbf{C}_j = \frac{1}{T} \sum_{t=1}^T c_t^j \mathbf{M}_t, \quad j = 1, 2, \quad (119)$$

with  $c_t^1 = k_n (\hat{V}_t^{n,3})^2 - 2V_t$  and  $c_t^2 = 2k_n \hat{V}_t^{n,3} (\hat{V}_t^{n,1} + \hat{V}_t^{n,2}) + k_n (\hat{V}_t^{n,1} + \hat{V}_t^{n,2})^2$ , using the notation from the proof of Lemma 7. From here, by applying the bound in (78), we have

$$|\mathbb{E}_t(c_t^1)| \leq K_{t,n} k_n \Delta_n \quad \text{and} \quad \mathbb{E}_t(c_t^1)^2 \leq K_{t,n}$$

where  $\mathbb{E}(K_{t,n})^2 \leq K$ . Hence, by splitting  $\mathbf{C}_1$  into  $\frac{1}{T} \sum_{t=1}^T \mathbb{E}_t(c_t^1) \mathbf{M}_t$  and  $\frac{1}{T} \sum_{t=1}^T (c_t^1 - \mathbb{E}_t(c_t^1)) \mathbf{M}_t$ , we get  $\mathbb{E} \|\mathbf{C}_1\|^2 \leq K \left( (k_n \Delta_n)^2 \vee \frac{1}{T} \right)$ , and, as a result,  $\mathbf{C}_1 = o_p(1)$ .

Next, for  $\mathbf{C}_2$ , we invoke the derived bounds for the conditional second moments of  $\hat{V}_t^{n,1}$  and  $\hat{V}_t^{n,2}$  in (76) and (79),

$$\mathbb{E}_t \left( \hat{V}_t^{n,1} + \hat{V}_t^{n,2} \right)^2 \leq K_{t,n} \left( \frac{\Delta_n^{(4-r)\varpi} - 1}{k_n} \vee k_n \Delta_n \right)$$

for some integrable  $K_{t,n}$ . Hence, we have  $\mathbb{E}_t |c_t^2| \leq K_{t,n} (\Delta_n^{\frac{(4-r)\varpi-1}{2}} \vee k_n \sqrt{\Delta_n})$ , where  $\mathbb{E}(K_{t,n})^2 \leq K$ . Therefore, given the restrictions on  $\varpi$  and  $k_n$  in (23), we get  $\mathbf{C}_2 = o_p(1)$ .

We continue with (117). Given part (a) of Lemma 9 and since  $\mathbb{E}_t \|\mathbf{F}_{t,N,n}\|^2 \leq \hat{\eta}_{t,N,n} K_{t,n}/N$  for some  $K_{t,n}$  with finite moments and  $\sup_{t=1,\dots,T} \hat{\eta}_{t,N,n} = O_p(1)$ , we have that the term on the left-hand side of (117) is asymptotically equivalent to

$$\frac{1}{T} \sum_{t=1}^T (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} (N \mathbf{F}_{t,N,n} \mathbf{F}'_{t,N,n} - \Omega_{S,t} - \Psi_{S,t}) (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{M}_t,$$

which, in turn, may be decomposed into  $\mathbf{B}_1$ ,  $\mathbf{B}_2$  and  $\mathbf{B}_3$ , where

$$\begin{aligned} \mathbf{B}_1 &= \frac{1}{T} \sum_{t=1}^T (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \left( N (\mathbf{F}_{t,N,n}^{(1)}) (\mathbf{F}_{t,N,n}^{(1)})' - \Omega_{S,t} \right) (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{M}_t, \\ \mathbf{B}_2 &= \frac{1}{T} \sum_{t=1}^T (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \left( N (\mathbf{F}_{t,N,n}^{(2)}) (\mathbf{F}_{t,N,n}^{(2)})' - \Psi_{S,t} \right) (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{M}_t, \\ \mathbf{B}_3 &= \frac{1}{T} \sum_{t=1}^T (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \left( N (\mathbf{F}_{t,N,n}^{(1)}) (\mathbf{F}_{t,N,n}^{(2)})' + N (\mathbf{F}_{t,N,n}^{(2)}) (\mathbf{F}_{t,N,n}^{(1)})' \right) (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{M}_t. \end{aligned}$$

First, for  $\mathbf{B}_1$ , if we denote  $\hat{\Omega}_{S,t}^n = \frac{1}{N_t} \sum_{j=1}^{N_t} \phi_{j,t} \mathbf{G}_{j,t,0} \mathbf{G}'_{j,t,0}$ , we may decompose  $\mathbf{B}_1$  further as

$$\begin{aligned} \mathbf{B}_{11} &= \frac{1}{T} \sum_{t=1}^T (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \left( N (\mathbf{F}_{t,N,n}^{(1)}) (\mathbf{F}_{t,N,n}^{(1)})' - \hat{\Omega}_{S,t}^n \right) (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{M}_t, \\ \mathbf{B}_{12} &= \frac{1}{T} \sum_{t=1}^T (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \left( \hat{\Omega}_{S,t}^n - \Omega_{S,t} \right) (\mathcal{I}_{S,t} + \mathcal{P}_{S,t})^{-1} \mathbf{M}_t. \end{aligned}$$

When conditioning on  $\mathcal{F}^{(0)}$ , the increments of  $\mathbf{B}_{11}$  are uncorrelated, and by additionally applying A2 for  $N_t/N$ , we have  $\mathbf{B}_{11} = O_p(1/\sqrt{T})$ . Next, using our assumption for the smoothness in  $k$  of  $\mathbf{G}(k, \tau, \mathbf{S}, \boldsymbol{\theta})$ ,  $\mathbf{D}(k, \tau, \mathbf{S}, \boldsymbol{\theta})$  and  $\phi_{t,k,\tau}$  in A5 and A7 as well as the assumption for the mesh of the observation grid in A2, we also have  $\mathbf{B}_{12} = o_p(1)$ . The asymptotic negligibility of  $\mathbf{B}_2$  is proved in exactly the same way as (118), and we are, thus, left with the analysis of  $\mathbf{B}_3$ . For the latter, by successive conditioning, and since  $\mathbb{E}_t \|\mathbf{F}_{t,N,n}^{(1)}\|^2 + \mathbb{E}_t \|\mathbf{F}_{t,N,n}^{(2)}\|^2 \leq \hat{\eta}_{t,N,n} K_{t,n}/N$  for some  $K_{t,n}$  with finite moments as well as  $\sup_{t=1,\dots,T} \hat{\eta}_{t,N,n} = O_p(1)$ , we have  $\mathbf{B}_3 = O_p(1/\sqrt{T})$ . This concludes the proof.  $\square$

**Lemma 13.** *Assume the conditions of Theorem 2 hold. Then, we have the convergence in (25) holding jointly with the one in (24).*

*Proof.* In the proof, without loss of generality, we assume that  $t_1 < \dots < t_h$ . First, given the consistency result of Theorem 1 and Lemma 9 (a), for any  $t$ , we have

$$\hat{\mathcal{I}}_{S,t}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) + \hat{\mathcal{P}}_{S,t}(\hat{\mathbf{S}}_t, \hat{\boldsymbol{\theta}}) \xrightarrow{\mathbb{P}} \mathcal{I}_{S,t} + \mathcal{P}_{S,t}.$$

Furthermore, again for any  $t$ , we have using the bounds in the proof of Lemma 7 that  $\sqrt{N}\mathbf{F}_{t,N,n} = O_p(1)$ . Then, using Lemmas 1-3 in Andersen et al. (2015a), we get the convergence result (25) to hold, with  $\mathbf{E}_{S_{t_1}}, \dots, \mathbf{E}_{S_{t_h}}$  being a sequence of standard Gaussian vectors, defined on an extension of the original probability space, each of them independent of each other, and of  $\mathcal{F}$ .

In addition, by applying Lemma 7, we have the convergence result in (24). What remains to be proved is that the convergence in (24) and (25) holds jointly with  $\mathbf{E}_\theta$  independent from each of the vectors  $\mathbf{E}_{S_{t_1}}, \dots, \mathbf{E}_{S_{t_h}}$ . For this, it is sufficient to show that for an arbitrary bounded random variable  $Y$  adapted to  $\mathcal{F}$  and bounded Lipschitz-continuous functions  $f_1, \dots, f_h$  and  $g$ , we have

$$\begin{aligned} \mathbb{E} \left( Y f_1 \left( \sqrt{N}\mathbf{F}_{t_1,N,n} \right) \cdots f_h \left( \sqrt{N}\mathbf{F}_{t_h,N,n} \right) g \left( \sqrt{NT}(\mathbf{A}_{11} + \mathbf{A}_{31}) \right) \right) \\ \rightarrow \mathbb{E} (Y f_1 (\mathbf{F}_{t_1}) \cdots f_h (\mathbf{F}_{t_h})) \mathbb{E} (g(\mathbf{F})), \end{aligned} \quad (120)$$

where  $\mathbf{F}_{t_1}, \dots, \mathbf{F}_{t_h}$  are  $\mathcal{F}$ -conditionally Gaussian distributed and independent of each other, each of which with  $\mathcal{F}$ -conditional asymptotic variance of  $\boldsymbol{\Omega}_{S,t_1} + \boldsymbol{\Psi}_{S,t_1}, \dots, \boldsymbol{\Omega}_{S,t_h} + \boldsymbol{\Psi}_{S,t_h}$ , respectively, and  $\mathbf{F}$  is Gaussian distributed with variance of  $\boldsymbol{\Omega}_U + \boldsymbol{\Psi}_\theta$ .

Now, since  $\mathcal{F}_{t_h}$  is separable, we also have the  $\mathcal{F}_{t_h}$ -conditional convergence of  $\sqrt{NT}(\mathbf{A}_{11} + \mathbf{A}_{31})$  to its limit by Theorem VIII.5.25 in Jacod and Shiryaev (2003), and, as a result,

$$\mathbb{E}_t \left( Y g \left( \sqrt{NT}(\mathbf{A}_{11} + \mathbf{A}_{31}) \right) \right) \xrightarrow{\mathbb{P}} \mathbb{E}_t(Y) \mathbb{E} (g(\mathbf{F})). \quad (121)$$

Therefore, since  $f_1, \dots, f_h$  are bounded, we have

$$\begin{aligned} \mathbb{E} \left( f_1 \left( \sqrt{N}\mathbf{F}_{t_1,N,n} \right) \cdots f_h \left( \sqrt{N}\mathbf{F}_{t_h,N,n} \right) \right. \\ \left. \times \left( \mathbb{E}_t \left( Y g \left( \sqrt{NT}(\mathbf{A}_{11} + \mathbf{A}_{31}) \right) \right) - \mathbb{E}_t(Y) \mathbb{E} (g(\mathbf{F})) \right) \right) \rightarrow 0. \end{aligned} \quad (122)$$

Furthermore, from the stable convergence result of  $\sqrt{N}\mathbf{F}_{t_1,N,n}, \dots, \sqrt{N}\mathbf{F}_{t_h,N,n}$ , we get

$$\begin{aligned} \mathbb{E} \left( f_1 \left( \sqrt{N}\mathbf{F}_{t_1,N,n} \right) \cdots f_h \left( \sqrt{N}\mathbf{F}_{t_h,N,n} \right) \mathbb{E}_t(Y) \mathbb{E} (g(\mathbf{F})) \right) \\ \rightarrow \mathbb{E} (Y f_1 (\mathbf{F}_{t_1}) \cdots f_h (\mathbf{F}_{t_h})) \mathbb{E} (g(\mathbf{F})). \end{aligned} \quad (123)$$

Combining (122) and (123), we get (120), and this establishes joint convergence in (24) and (25).  $\square$



## B Computational Details

The PLS estimators introduced in Sections 3 and 5, respectively, entail optimization over a large set of parameters as well as state variables. In order to maximize the objective functions in equations (11) and (29), we proceed in two steps: (1) we propose an initial guess,  $\hat{\theta}^1$ , for the unknown parameter vector  $\theta$ ; (2) given  $\hat{\theta}^1$ , on each day in the sample, we optimize equation (12) (or equation (31)) over the factor realizations,  $\mathbf{S}_t$ , thus obtaining an initial estimate  $\hat{\mathbf{S}}_t^1$ . Since the parameter vector is fixed over the entire sample, the optimization of equation (12) can easily be parallelized.

Next, with  $\hat{\theta}^1$  and  $\hat{\mathbf{S}}^1 = (\hat{\mathbf{S}}_1^1, \dots, \hat{\mathbf{S}}_T^1)$ , we compute  $\mathcal{L}_t^1$  and finally  $\mathcal{L}^1$ . The algorithm, then, iterates through step (1), by proposing a new parameter vector  $\hat{\theta}_t^i$ , and step (2), by re-optimizing over the factor values on each day in the sample to obtain a new estimator  $\hat{\mathbf{S}}^i$ , until the minimization criterion in equation (13) (or equation (32)) converges.

From a technical point of view, a local optimizer works reasonably well for the optimization over the latent factors  $\hat{\mathbf{S}}$  in equations (12) and (31). We find that the BOBYQA algorithm contained in the NLOpt (<https://nlopt.readthedocs.io/en/latest/>) library performs well in terms of both accuracy and computational time. As far as the optimization over the parameter vector  $\hat{\theta}$  is concerned, global optimizers such as the Controlled Random Search (CRS) with local mutation or the ISRES (Improved Stochastic Ranking Evolution Strategy) algorithm in the NLOpt library seem to work reasonably well.

## C Additional Numerical Results

In this section, we provide additional evidence regarding the performance of the proposed estimation procedure for different values of  $N$ , the size of the option cross-section on each day. Specifically, we compare the results in Section 6, obtained for  $N = 200$ , to a similar setting with  $N = 100$ . Note that we do not include a penalization term in the objective function in either case.

Figure 7, consistent with Theorem 2 and Lemma 1, shows that the precision with which the factors are recovered scales by a factor  $1/\sqrt{N}$ . Thus, the median standard error approximately increases with a factor  $\sqrt{2}$  when the number of observations per cross-section reduces from 200 to 100.

Next, when considering the correlation between the errors in the factor extraction, Figure 8 shows relatively little sensitivity of this quantity to changes in  $N$ . Recall from our asymptotic results that the correlation between the errors in recovering the factors on different days is a small  $T$  phenomenon, implying that it should remain unchanged when keeping  $T$  fixed.

Finally, Figure 9 shows the impact of different values of  $N$  on the estimates of the standard errors as well as the asymptotic bias of the parameters. Consistent with the asymptotic theory, the standard errors decrease approximately by a factor  $\sqrt{2}$  and the estimates of the bias by a factor 2, when going

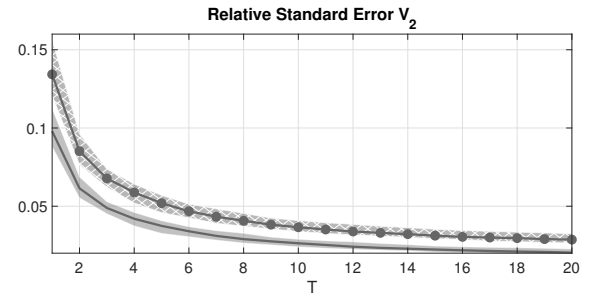
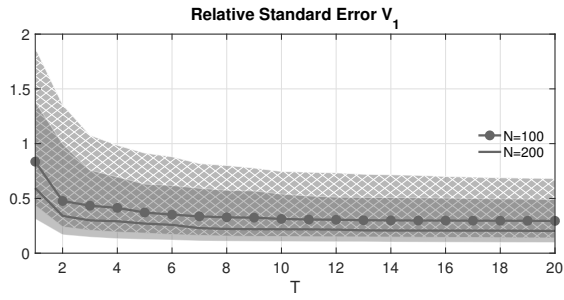


Figure 7: *Relative Standard Errors for Factors: Effect of  $N$* . Standard errors are computed for the factor realization from the last week of the first year in the sample. Standard errors are divided by the absolute value of the true value of the factor realization. The straight line in each of the plots corresponds to the median and the shaded area to the inter-quantile range, both computed from 100 draws of the process  $(V_{1,t}, V_{2,t})$ .

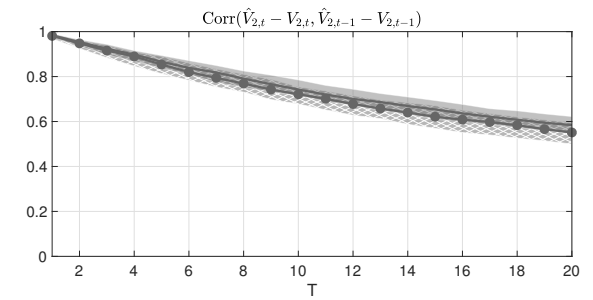
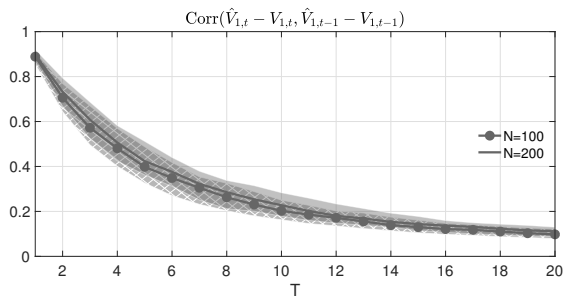


Figure 8: *Persistence of the Error in Factor Extraction: Effect of  $N$* . The plots display correlation between the errors in extracting factors over the last two weeks in the first year of the sample. The straight line in each of the plots corresponds to the median and the shaded area to the inter-quantile range, both computed from 100 draws of the process  $(V_{1,t}, V_{2,t})$ .

from  $N = 100$  to  $N = 200$ , for a given  $T$ .

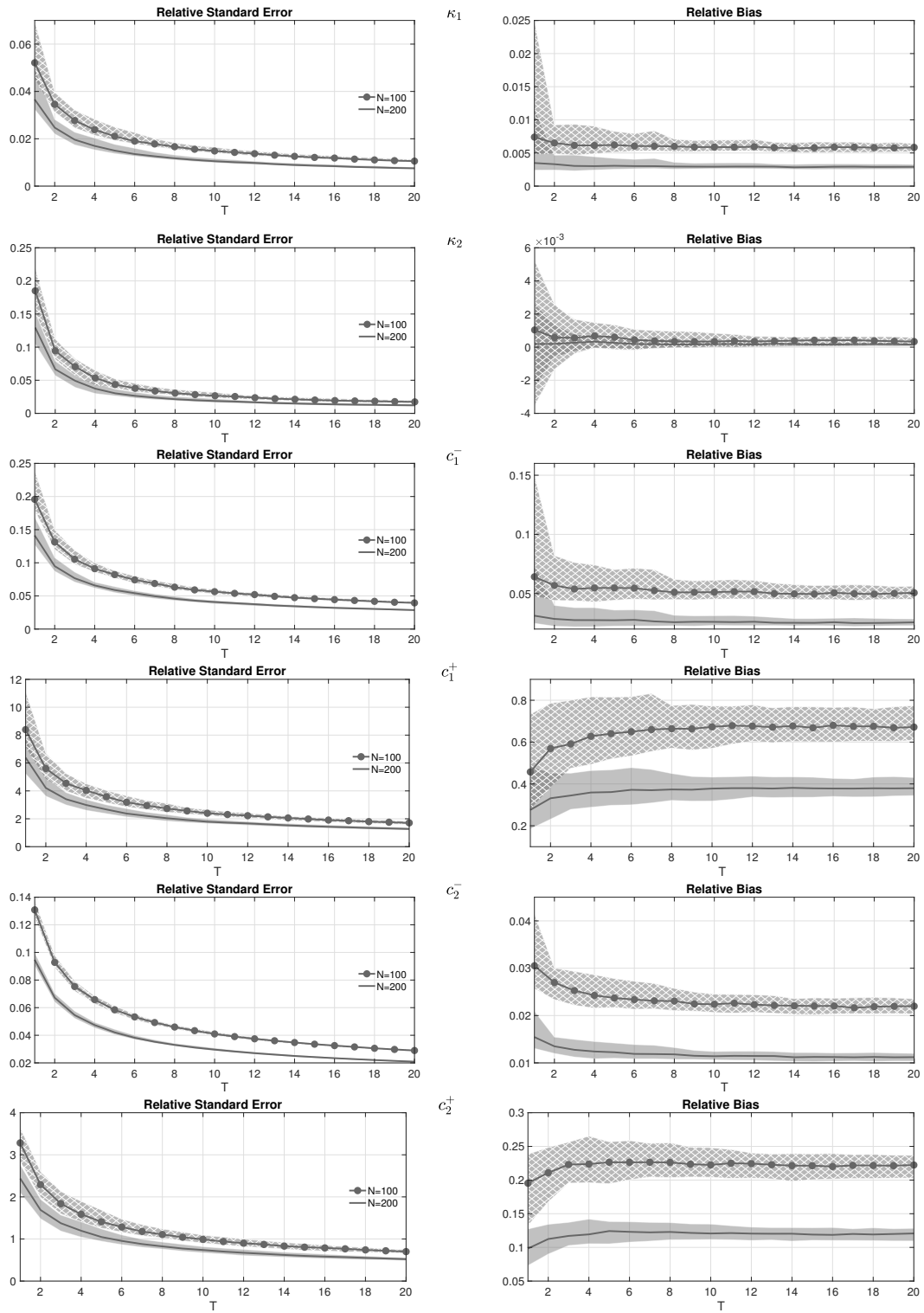


Figure 9: *Relative Standard Error and Bias for Parameters: Effect of  $N$ .* Standard errors and biases are divided by the absolute value of the true value of the parameter reported in Table 1. The straight line in each of the plots corresponds to the median and the shaded area to the inter-quantile range, both computed from 100 draws of the process  $(V_{1,t}, V_{2,t})$ .