Joint ML estimation of all parameters in a discrete time random field HJM type interest rate model

József Gáll*, Gyula Pap** and Martien van Zuijlen***

*Faculty of Economics, University of Debrecen, Hungary,

jozsef.gall@econ.unideb.hu

**Bolyai Institute, University of Szeged, Hungary

papgy@math.u-szeged.hu

***Institute for Mathematics, Astrophysics and Particle Physics,

Radboud University, Nijmegen, The Netherlands

M.vanZuijlen@science.ru.nl

Abstract

We consider discrete time Heath–Jarrow–Morton type interest rate models, where the interest rate curves are driven by a geometric spatial autoregression field. Strong consistency and asymptotic normality of the maximum likelihood estimators of the parameters are proved for stable no-arbitrage models containing a general stochastic discounting factor, where explicit form of the ML estimators is not available given a non-i.i.d. sample. The results form the basis of further statistical problems in such models.

Keywords: Heath–Jarrow–Morton models, interest rate, maximum likelihood estimation, consistency, asymptotic normality, AR random fields.

1 Introduction

Our aim in the present paper is to consider some statistical questions arising in a Heath–Jarrow–Morton (HJM) type interest rate model proposed by Gáll, Pap and Zuijlen [5]. Such models are useful not only for describing the structure of interest rates but also for describing bond price structures in the market. We focus on asymptotic properties of the joint maximum likelihood estimators (MLE) of the parameters of the model, where the non-i.i.d. sample and the lack of an explicit form of the estimators make the derivation of the results difficult. These results give the basis of further statistical problems, such as hypothesis tests, interval estimations or model selection tools.

In the following we specify the model. For $\mathbb{Z}_+$ being the sets of non-negative integers, let $f_{k,\ell}$ denote the forward interest rate at time $k \in \mathbb{Z}_+$ with time to maturity date $\ell \in \mathbb{Z}_+$. Hence it is the interest rate for the future time period $[k+\ell, k+\ell+1)$.

The forward rate dynamics is supposed to be given by the (stochastic) difference equation

$$f_{k+1,\ell} = f_{k,\ell} + \alpha_{k,\ell} + \beta \Delta_1 S_{k,\ell}, \qquad k, \ell \in \mathbb{Z}_+,$$

where the initial values $(f_{0,\ell})_{\ell \in \mathbb{Z}_+}$ are given real numbers, $\beta \in \mathbb{R}$ denotes the volatility and $\Delta_1 S_{k,\ell} := S_{k+1,\ell} - S_{k,\ell}$, where $(S_{k,\ell})_{k,\ell \in \mathbb{Z}_+}$ is a doubly geometric spatial autoregressive process given by

$$\begin{cases} S_{k,\ell} = S_{k-1,\ell} + \varrho S_{k,\ell-1} - \varrho S_{k-1,\ell-1} + \eta_{k,\ell}, \\ S_{k,-1} = S_{0,\ell} = S_{0,-1} := 0, \end{cases} \qquad k \in \mathbb{N}, \quad \ell \in \mathbb{Z}_+,$$

with autoregression parameter $\varrho \in \mathbb{R}$, where $(\eta_{k,\ell})_{k \in \mathbb{N}, \ell \in \mathbb{Z}_+}$ is a set of independent standard normal random variables on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$, and $\mathbb{N}$ denotes the set of positive integers. The drift $\alpha_{k,\ell}$ is supposed to be an $\mathcal{F}_k$-measurable random variable, where the filtration $(\mathcal{F}_k)_{k \in \mathbb{Z}_+}$ is given by the trivial σ-algebra $\mathcal{F}_0 := \{\emptyset, \Omega\}$ and

$$\mathcal{F}_k := \sigma(\eta_{i,j} : 1 \leqslant i \leqslant k \ \text{ and } \ j \geqslant 0), \qquad k \in \mathbb{N}.$$

Let $P_{k,\ell}$ denote the price of a zero coupon bond at time $k \in \mathbb{Z}_+$ with maturity $\ell \in \mathbb{Z}_+$ with $\ell \geqslant k$. Assume that the relationship between the forward interest rates and the prices of a zero coupon bond is given by $P_{k,k} = 1$, $k \in \mathbb{N}$, and

$$P_{k,\ell+1} = \exp\left\{ -\sum_{j=0}^{\ell-k} f_{k,j} \right\}, \qquad k, \ell \in \mathbb{Z}_+ \text{ with } k \leqslant \ell,$$

so that $P_{k,\ell+1} = \mathrm{e}^{-f_{k,\,\ell-k}} P_{k,\ell}$. Next we consider for given positive integer J a stochastic discount factor process $(M_k)_{k \in \mathbb{Z}_+}$ given by $M_0 := 1$ and

$$M_{k+1} := M_k \exp\{-r_k\} \frac{\exp\left\{ \sum_{j=0}^{J} b_j \Delta_1 S_{k,j} \right\}}{\mathsf{E}\left(\exp\left\{ \sum_{j=0}^{J} b_j \Delta_1 S_{k,j} \right\} \big| \mathcal{F}_k \right)}, \qquad k \in \mathbb{Z}_+,$$

where $r_k := f_{k,0}$ are the spot interest rate (corresponding to time k) and $\mathbf{b} = (b_0, b_1, \ldots, b_J) \in \mathbb{R}^{J+1}$ is the vector of the market price of risk parameters. They play an important role in the market when determining the market prices of assets. This role is discussed in detail in [5], where also the reasoning for the choice of the special form of the stochastic discount factors has been given. Note that the collection of unknown parameters we have to deal with are these risk parameters, the volatility β and the autoregression parameter ρ.

We are interested only in models where arbitrage opportunities are excluded in the market. No-arbitrage property follows from a martingale condition, which is satisfied if the M_k-discounted bond price processes $(M_k P_{k,\ell})_{0 \leqslant k \leqslant \ell}$ form martingales for all $\ell \in \mathbb{N}$. Using the equations resulting from the martingale condition, the drifts $\alpha_{k,\ell}$ disappear and we obtain

$$\begin{cases} f_{k,\ell} - f_{k-1,\ell+1} - \varrho(f_{k,\ell-1} - f_{k-1,\ell}) = \beta\eta_{k,\ell} + \frac{\beta^2}{2}\sum\limits_{j=0}^{2\ell}\varrho^j - \beta\sum\limits_{j=\ell}^{J} b_j\varrho^{j-\ell}, \\ f_{k,0} - f_{k-1,1} = \beta\eta_{k,0} + \frac{\beta^2}{2} - \beta\sum\limits_{j=0}^{J} b_j\varrho^j, \end{cases} \tag{1}$$

for $k, \ell \in \mathbb{N}$. The details of the derivation of these no-arbitrage equations together with the role of the market discount factors can be found in [5].

The main goal of this paper is to prove strong consistency and asymptotic normality of the joint MLE of the parameters $(\beta, \varrho, b_0, \dots, b_J)$ based on samples $(f_{k,\ell})_{1\leqslant k\leqslant K_n,\, 0\leqslant\ell\leqslant L_n}$, where $K_n = Kn + o(n)$ and $L_n = Ln + o(n)$ as $n \to \infty$ with some $K > 0$ and $L > 0$. Of course, the main difficulty is that the samples consist of non-independent, non-identically distributed random variables and moreover, no explicit formula is available for the MLE of $(\beta, \varrho, b_0, \dots, b_J)$.

It will turn out that compared to the other two parameters β, and ϱ, the market price of risk parameters have a different asymptotic behaviour.

When dealing with certain problems and in particular with pricing derivatives, for the sake of convenience, many authors started modelling interest rate and bond markets under an equivalent martingale measure. However, statistical properties of the parameter estimations usually cannot be discussed in that way, so that we had to work under the real (objective) measure of the market. We would like to mention that in our opinion statistical tools have to be applied in finance for instance for pricing derivatives, since in many situations the market will not be complete, so that one cannot work under an equivalent martingale measure and one has to fit real date to the model. Unfortunately, in the above sense relatively few papers are written in finance with a real statistical orientation.

Concerning the present literature we mention the following related results. In the type of interest rate framework we have investigated, there are some results already available for the MLE of a single parameter assuming that the true values of the other parameters are known. In [4] the MLE of the volatility β has been investigated, and asymptotic normality has been obtained both in stable and in nearly unstable cases. (A model is called stable, unstable, or explosive, if $|\varrho| < 1$, $|\varrho| = 1$, or $|\varrho| > 1$, respectively. In the nearly unstable case given a sequence of models with corresponding autoregression parameter ϱ_n we have $\lim_{n\to\infty}\varrho_n = 1$.) Volatility estimation has also been studied by Peeters [11] in case of a more complicated volatility structure. Fülöp and Pap [1] tested the autoregression parameter ϱ both in stable and unstable cases, and they succeeded in proving local asymptotic normality of the sequence of the related statistical experiments in the sense of [10]. In a further work, in Fülöp and Pap [2], they also gave results on strong consistency of the MLE estimator of the autoregressive parameter.

The paper is organized as follows. In Section 2 we will formulate our results on consistency (Theorem 2.1) and on the asymptotic normality of the joint ML parameter estimators (Theorem 2.2). In Section 3 we discuss our results together with their consequences, as well as some related problems and future work. In Appendix A we give first the derivation of the likelihood function which is followed by the rigorous mathematical proofs of our main results. In Appendix B we collected some useful general (not model specific) lemmas we apply in the proofs of the main theorems.

2 MLE and results

In this section we present the main results on the joint maximum likelihood estimators of the parameters $(\beta, \varrho, b_0, \ldots, b_J)$ of the model.

Consider a sample $(f_{k,\ell})_{1\leqslant k\leqslant K, 0\leqslant \ell\leqslant L}$ taken from a model (1). One needs first to obtain the log-likelihood function which can be derived based on the no-arbitrage conditions given in [5]. It has the form

$$\begin{aligned} \mathcal{L}_{K,L}(x_{k,\ell} : 1 \leqslant k \leqslant K,\, 0 \leqslant \ell \leqslant L;\, \beta, \varrho, \mathbf{b}) &= -\frac{K(L+1)}{2}\log(2\pi\beta^2) \\ &\quad - \frac{1}{2}\log(K!) - \frac{1}{2\beta^2}\sum_{k=1}^{K}\sum_{\ell=0}^{L-1}\left(y_{k,\ell}(\varrho) - \frac{\beta^2}{2}\sum_{i=0}^{2\ell}\varrho^i + \beta\sum_{j=\ell}^{J} b_j \varrho^{j-\ell}\right)^2 \\ &\quad - \frac{1}{2\beta^2}\sum_{k=1}^{K}\frac{1}{k}\left(\widetilde{y}_{k,L}(\varrho) - \frac{\beta^2}{2}\sum_{j=1}^{k}\sum_{i=0}^{2(k+L-j)}\varrho^i + \beta\sum_{j=0}^{J} b_j q_{j,k,\ell}\right)^2, \end{aligned} \tag{2}$$

where

$$\begin{aligned} y_{k,\ell}(\varrho) &:= \begin{cases} x_{k,\ell} - x_{k-1,\ell+1} - \varrho(x_{k,\ell-1} - x_{k-1,\ell}) & \text{for } \ 1 \leqslant \ell \leqslant L-1, \\ x_{k,0} - x_{k-1,1} & \text{for } \ \ell = 0, \end{cases} \\ \widetilde{y}_{k,L}(\varrho) &:= x_{k,L} - x_{0,k+L} - \varrho(x_{k,L-1} - x_{0,k+L-1}) \end{aligned} \tag{3}$$

for all $k, L \geqslant 1$, and $x_{0,\ell} := f_{0,\ell}$ for $\ell \geqslant 1$. The derivation of the log-likelihood function is given in the Appendix A in Remark A.1.

Unfortunately this log-likelihood function has a complicated form. Hence one cannot hope to get an explicit solutions for the estimators of all the parameters. We mention here that knowing the true values of some parameters, it is possible to give an explicit formula for the estimator(s) of the remaining parameter(s). Such a case is considered in [4], where the volatility estimator is studied in a similar model. In general one has to use numerical procedures to maximise (2) in order to obtain the ML estimators. Although we do not have explicit form for the estimators, the following theorems assure us that they have good statistical properties (like in classical cases): the first theorem is on the consistency, the second is on the asymptotic normality of the joint estimators.

Theorem 2.1 *Let $H \subset \mathbb{R}^{J+3}$ be a compact set such that for all $(\beta, \varrho, \mathbf{b}) \in H$ we have $\beta \neq 0$ and $\varrho \in (-1, 1)$. Let $(\beta_0,\, \varrho_0,\, \mathbf{b}_0) \in H$ denote the true parameters, where we write $\mathbf{b}_0 = (b_{0,0},\, b_{0,1}, \ldots,$*

$b_{0,J})$. *Let* K_n, L_n, $n \in \mathbb{N}$, *be positive integers such that* $K_n = nK + o(n)$ *and* $L_n = nL + o(n)$ *as* $n \to \infty$ *with some* $K > 0$ *and* $L > 0$. *For each* $n \in \mathbb{N}$ *let* $(\widehat{\beta}_n, \widehat{\varrho}_n, \widehat{\mathbf{b}}_n)$ *denote a maximum likelihood estimator of* $(\beta_0, \varrho_0, \mathbf{b}_0)$ *maximising the (log-)likelihood function over* H.

Then the sequence $(\widehat{\beta}_n, \widehat{\varrho}_n, \widehat{\mathbf{b}}_n)_{n\in\mathbb{N}}$ *is a strongly consistent estimator of* $(\beta_0, \varrho_0, \mathbf{b}_0)$, *i.e.*,

$$(\widehat{\beta}_n, \widehat{\varrho}_n, \widehat{\mathbf{b}}_n) \to (\beta_0, \varrho_0, \mathbf{b}_0) \qquad \text{a.s. as } n \to \infty. \tag{4}$$

Theorem 2.2 *Under the assumptions of Theorem 2.1 we have*

$$\begin{bmatrix} n(\widehat{\beta}_n - \beta_0) \\ n(\widehat{\varrho}_n - \varrho_0) \\ \sqrt{n}(\widehat{\mathbf{b}}_n - \mathbf{b}_0) \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Lambda), \qquad \text{as } n \to \infty, \tag{5}$$

such that Λ *is of the form*

$$\Lambda := \begin{bmatrix} \Lambda_1 & \mathbf{0} \\ \mathbf{0} & \Lambda_2 \end{bmatrix},$$

where

$$\Lambda_1 := \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{bmatrix}^{-1} = \left(\sigma_{1,1}\sigma_{2,2} - \sigma_{1,2}^2\right)^{-1} \begin{bmatrix} \sigma_{2,2} & -\sigma_{1,2} \\ -\sigma_{1,2} & \sigma_{1,1} \end{bmatrix}$$

with

$$\sigma_{1,1} := \frac{2KL}{\beta_0^2} + \frac{K(K+2L)}{2\,(1-\varrho_0)^2}, \qquad \sigma_{2,2} := \frac{KL}{1-\varrho_0^2} + \frac{K(K+2L)\beta_0^2}{2\,(1-\varrho_0)^4}, \tag{6}$$

$$\sigma_{1,2} = \sigma_{2,1} := \frac{K(K+2L)\beta_0}{2\,(1-\varrho_0)^3}, \tag{7}$$

furthermore, Λ_2 *of size* $(J+1) \times (J+1)$ *has the form*

$$\Lambda_2 := \frac{1}{K} \begin{bmatrix} 1+\varrho_0^2 & -\varrho_0 & 0 & 0 & 0 & \dots & 0 \\ -\varrho_0 & 1+\varrho_0^2 & -\varrho_0 & 0 & 0 & \dots & 0 \\ 0 & -\varrho_0 & 1+\varrho_0^2 & -\varrho_0 & 0 & \dots & 0 \\ \vdots & \vdots & & & & & \vdots \\ 0 & 0 & \dots & 0 & -\varrho_0 & 1+\varrho_0^2 & -\varrho_0 \\ 0 & 0 & \dots & \dots & 0 & -\varrho_0 & 1 \end{bmatrix}.$$

3 Discussion of the results

In this paper we presented statistical results for discrete time HJM type forward rate models which are driven by autoregressive (AR) random fields. We considered some natural questions that arise when fitting such a model. Our aim was to examine the joint behaviour of the maximum likelihood estimator of all parameters of our model. That is, we considered the joint estimation of the AR field parameter ϱ, the volatility parameter β and the market price of risk parameters $b_0, b_1, \ldots, b_J$.

The challenge we faced was to derive good properties of the estimators (consistency, asymptotic normality) in a model where the observations are neither independent nor identically distributed. Furthermore, as a consequence of the complexity of the likelihood function there is no hope for deriving explicit solutions of the maximum likelihood estimators, which complicated the task.

Therefore, given a real market data set of forward rates, in order to fit the model one needs first to use numerical procedures to reach the maxima of the likelihood function (2). Note that due to Theorem 2.1 one can reach the maxima by the use of first order conditions. On the other hand, due to the same theorem we are assured that the estimators are consistent. We also showed that joint asymptotic normality of the estimator holds like in the well-known cases of MLE for i.i.d. samples (under certain conditions). We emphasise here that the estimators had different normalising factors in Theorem 2.2, which might be interesting for the reader. Namely, in the normalizing factor, the market price of risk parameters differ from the 'classical' ('square-root') factor (of the well-known i.i.d. cases) as the sample size goes to infinity. In that sense it is not classical because it is not proportional to the reciprocal value of the square root of the sample size. (For this notice that the sample size we took in our theorem was of order n^2). Another interesting property of these risk parameters is that their estimators are asymptotically uncorrelated from the estimators of β and ϱ. To see this we refer to the structure of the sample's Fisher information Σ in Theorem A.1.

Gáll, Pap and Peeters [6] discussed more on the numerical problems and gave numerical results of the estimations at issue. It was shown by the tests that even in case of small sample sizes the behaviour of the estimators was still very good, the estimators converged fairly fast. Due to this one can have good hope to fit the model well to real data.

As we mentioned before, Fülöp and Pap [1] considered the separate estimation of the autoregression parameter ϱ both in stable and unstable cases. In the stable case the scaling factor is n^{-1}, like in our case, of course. However, in the unstable cases the scaling factors are different, namely, n^{-2} and n^{-3}. These scaling factors are in accordance with the Fisher information quantity contained in the sample. Based on Example 9.12 of [12], we expect in the explosive case the sequence of the related statistical experiments to be locally asymptotic mixed normal. Finally, we note that Fülöp [3] gave some early numerical results on the estimation of ϱ as well in the above mentioned cases.

Related models and future work

In Gáll, Pap and Zuijlen [5] a general setup has been proposed for discrete time forward rate curves driven by random fields. In this paper we studied an important special case. However, we mention that this is certainly not the only interesting specification of the model one can study. More complicated volatility structures, other forms of market price of risk as well as different random fields can also be the subject of further research. We believe that in order to derive similar statistical results for several modifications of the recent model, the methods we used for the proofs will also work. For this Appendix B contains some useful tools. We appreciate very much the works [13] and [14] of Ying. Though we did not apply directly any of Ying's specific results, his methods and ideas were especially fruitful during the development of the proofs of our main theorems. We note that also other methods might have been also applied in order to derive the asymptotic results. Here we mention among others the excellent papers of Heijmans and Magnus [7] and [8]. However, the approach we have we chosen (motivated by Ying's approach) has turned out to be fairly appropriate and effective for our purposes.

The asymptotic results we have found can form the basis of hypothesis tests that we intend to develop in our forthcoming studies. In this way one can hope to be able to test the goodness of fit of the model and possibly to compare the fits of different models. For model selection, information criteria might also be used. In our present research we are focusing on such problems. In that sense this paper is just the first, however the fundamental step for our purposes. We find these problems important since, unlike in many fields of econometrics, the goodnesses of fit of recently applied financial market models are often not justified by empirical means at all (tests, information criteria). They are often 'just parametrised to be rich enough' so that the model produces (derivative) asset prices being 'close enough' to market data. However, overparametrised models or misspecified models may occur in this way.

Forward rate models are, of course, not only used for pricing interest rate derivatives. We hope that by finding the appropriate models and testing tools, risk management of firms entering to markets of bonds and interest rate related assets can also be better supported. (Here we also refer to the fact that for derivative pricing one needs not necessarily take our way of parameter estimation —under the objective measure—, but one can alternatively use well-known calibration techniques to fit the models.) However, for many problems (e.g. risk management, goodness of fit) we suggest to take our approach to fit and test the model.

Acknowledgements

The research of G. Pap was realized in the frames of TÁMOP 4.2.4. A/2-11-1-2012-0001 ,,National Excellence Program – Elaborating and operating an inland student and researcher personal support system". The project was subsidized by the European Union and co-financed by the European Social Fund.

We are thankful to Ronald Kortram for the fruitful discussions on uniformity issues.

References

[1] Fülöp, E. and Pap, G. (2007), *Asymptotically optimal tests for a discrete time random*

field HJM type interest rate model, Acta Scientiarum Mathematicarum, **73(3-4)**, 637–661.

[2] Fülöp, E. and Pap, G. (2009), *Strong consistency of maximum likelihood estimators for a discrete-time random field HJM-type interest rate model*, Lithuanian Math. J., **49(1)**, 5-25.

[3] Fülöp, E. (2009), *Simulations of a discrete time HJM type forward interest rate model*, unpublished manuscript.

[4] Gáll, J., Pap, G. and Zuijlen, M. v. (2004), *Maximum likelihood estimator of the volatility of forward rates driven by geometric spatial AR sheet*, Journal of Applied Mathematics **2004(4)**, 293–309.

[5] Gáll, J., Pap, G. and Zuijlen, M. v. (2006), *Forward interest rate curves in discrete time settings driven by random fields*, Computers & Mathematics with Applications, **51(3-4)**, 387–396.

[6] Gáll, J., Pap, G. and Peeters, W. (2007), *Random field forward interest rate models, market price of risk and their statistics*, Analli dell'Universita di Ferrara Sez. VII Sci. Mat., **53**, 233–242.

[7] Heijmans, R. D. H. and Magnus, J. R. (1986), *Consistent maximum-likelihood estimation with dependent observations, The general (non-normal) case and the normal case*, Journal of Econometrics, **32**, 253–285.

[8] Heijmans, R. D. H. and Magnus, J. R. (1986), *Asymptotic normality of maximum likelihood estimators obtained from normally distributed but dependent observations*, Econometric Theory, **2**, 374–412.

[9] Jacod, J., Shiryayev, A. N. (1987), *Limit Theorems for Stochastic Processes*, Springer-Verlag, Berlin.

[10] Le Cam, L. (1986), *Asymptotic Methods in Statistical Decision Theory*, New York: Springer-Verlag.

[11] Peeters, W. (2008), *Volatility estimation for different structures of random field interest rate models in discrete time*, Publicationes Mathematicae Debrecen, **72(3-4)**, 317–334.

[12] Vaart, A. W. van der (1998), *Asymptotic Statistics*, Cambridge University Press.

[13] Ying, Z. (1993), *Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process*, Journal of Multivariate Analysis, **36**, 280–296.

[14] Ying, Z. (1993), *Maximum likelihood of parameters under spatial sampling scheme*, The Annals of Statistics, **21**, 1567–1590.

Appendix A: Proofs of the main results

Remark A.1 (Derivation of the likelihood function) It can be seen from the main results of Gáll, Pap and Zuijlen [5] that under the assumption that the common distribution of the $\eta_{i,j}$'s, for $i, j \in \mathbb{Z}_+$, is standard normal, the no–arbitrage criterion is equivalent with

$$f_{k,\ell} - f_{k-1,\ell+1} - \varrho(f_{k,\ell-1} - f_{k-1,\ell}) = \beta\eta_{k,\ell} + \frac{\beta^2}{2}\sum_{i=0}^{2\ell}\varrho^i - \beta\sum_{j=\ell}^{J} b_j\varrho^{j-\ell}, \tag{A.1}$$

and hence

$$f_{k,\ell-1} - f_{k-1,\ell} = \beta\sum_{i=0}^{\ell-1}\varrho^{\ell-i-1}\eta_{k,i} + \frac{\beta^2}{2}\left(\sum_{i=0}^{\ell-1}\varrho^i\right)^2 - \beta\sum_{j=0}^{J} b_j \sum_{i=0}^{j\wedge(\ell-1)}\varrho^{\ell+j-1-2i} \tag{A.2}$$

for $k \geqslant 1$, $\ell \geqslant 1$. Furthermore, we have

$$\begin{aligned} f_{k,\ell} - f_{0,k+\ell} = \sum_{n=0}^{k}\Bigg[&\frac{\beta^2}{2}\left(\sum_{i=0}^{k+\ell-n}\varrho^i\right)^2 + \beta\sum_{i=0}^{k+\ell-n}\varrho^{k+\ell-n-i}\eta_{n,i} \\ &- \beta\sum_{j=0}^{J} b_j\sum_{i=0}^{j\wedge(k+\ell-n)}\varrho^{k+\ell-n+j-2i}\Bigg] \end{aligned} \tag{A.3}$$

and

$$\begin{aligned} &f_{k,\ell} - f_{0,k+\ell} - \varrho(f_{k,\ell-1} - f_{0,k+\ell-1}) \\ &\qquad = \beta\sum_{j=1}^{k}\eta_{j,k+\ell-j} + \frac{\beta^2}{2}\sum_{j=1}^{k}\sum_{i=0}^{2(k+\ell-j)}\varrho^i - \beta\sum_{j=0}^{J} b_j q_{j,k,\ell} \end{aligned} \tag{A.4}$$

for $k \geqslant 1$, $\ell \geqslant 1$, where

$$q_{j,k,\ell} := \begin{cases} \sum_{n=0\vee(j-k-\ell+1)}^{j-\ell}\varrho^n & \text{for } j \geqslant \ell \\ 0 & \text{otherwise.} \end{cases} \tag{A.5}$$

Consider now a sample $(f_{k,\ell})_{1\leqslant k\leqslant K,\,0\leqslant\ell\leqslant L}$ By the help of equations (A.1) and (A.4) one can obtain the joint density function of $(f_{k,\ell})_{1\leqslant k\leqslant K,\,0\leqslant\ell\leqslant L}$ and hence the likelihood function takes the form

$$\begin{aligned} \mathbf{L}_{K,L}(x_{k,\ell} : 1 \leqslant k \leqslant K,\, 0 \leqslant \ell \leqslant L;\, \beta, \varrho, \mathbf{b}) &= \frac{1}{(2\pi\beta^2)^{(L+1)K/2}(K!)^{1/2}} \\ \times \exp\Bigg\{ &- \frac{1}{2\beta^2}\sum_{k=1}^{K}\sum_{\ell=0}^{L-1}\left(y_{k,\ell}(\varrho) - \frac{\beta^2}{2}\sum_{i=0}^{2\ell}\varrho^i + \beta\sum_{j=\ell}^{J} b_j\varrho^{j-\ell}\right)^2 \\ &- \frac{1}{2\beta^2}\sum_{k=1}^{K}\frac{1}{k}\left(\widetilde{y}_{k,L}(\varrho) - \frac{\beta^2}{2}\sum_{j=1}^{k}\sum_{i=0}^{2(k+L-j)}\varrho^i + \beta\sum_{j=0}^{J} b_j q_{j,k,\ell}\right)^2\Bigg\}, \end{aligned}$$

where $y_{k,\ell}(\varrho)$ and $\widetilde{y}_{k,L}(\varrho)$ are given in (3).

Thus the log-likelihood function has, indeed, the form given in (2). Note that $q_{i,k,\ell}$ (see (A.5)) is bounded over a compact subset of $(-1,1)$ (since clearly $|q_{i,k,\ell}| \leqslant 1/(1-|\varrho|)$). Moreover recall that $q_{i,k,\ell}$ vanishes for large ℓ. These facts will simplify many problems in the proofs of the results on the asymptotics of the likelihood estimators.

Notation. For simplicity, in what follows we will write

$$\mathcal{L}_n(\beta,\varrho,\mathbf{b}) = \mathcal{L}_{K_n,L_n}(f_{k,\ell} : 1 \leqslant k \leqslant K_n, 0 \leqslant \ell \leqslant L_n; \beta, \varrho, \mathbf{b})$$

and

$$\partial_1^{i_1}\partial_2^{i_2}\partial_{j_1}^{i_3}\partial_{j_2}^{i_4}\mathcal{L}_n(\beta,\varrho,\mathbf{b}) = \frac{\partial^{i_1}\partial^{i_2}\partial^{i_3}\partial^{i_4}\mathcal{L}_{K_n,L_n}(x_{k,\ell} : 1 \leqslant k \leqslant K_n, 0 \leqslant \ell \leqslant L_n; \beta, \varrho, \mathbf{b})}{\partial\beta^{i_1}\partial\varrho^{i_2}\partial b_{j_1-3}^{i_3}\partial b_{j_2-3}^{i_4}}\bigg|_{x_{k,\ell}=f_{k,\ell}},$$

where $i_1, i_2, i_3, i_4, j_1, j_2$ are non-negative integers and $3 \leqslant j_i \leqslant J+3$ for $i = 1, 2$. Furthermore,

$$\boldsymbol{\partial_3}\mathcal{L}_n(\beta,\varrho,\mathbf{b}) = \begin{bmatrix} \partial_3\mathcal{L}_n(\beta,\varrho,\mathbf{b}) \\ \partial_4\mathcal{L}_n(\beta,\varrho,\mathbf{b}) \\ \vdots \\ \partial_{J+3}\mathcal{L}_n(\beta,\varrho,\mathbf{b}) \end{bmatrix}.$$

Proof of Theorem 2.1. First we show strong consistency of $\widehat{\beta}_n$, and $\widehat{\varrho}_n$. For this, the aim of the following discussion is to derive an asymptotic expansion for the sequence of random variables

$$\mathcal{L}_n(\beta,\varrho,\mathbf{b}) = \mathcal{L}_{K_n,L_n}(f_{k,\ell} : 1 \leqslant k \leqslant K_n,\ 0 \leqslant \ell \leqslant L_n;\ \beta, \varrho, \mathbf{b}), \qquad n \geqslant 1.$$

We have

$$\begin{aligned}\mathcal{L}_n(\beta,\varrho,\mathbf{b}) = &-\frac{K_n(L_n+1)}{2}\log(2\pi\beta^2) - \frac{1}{2}\log(K_n!) \\ &-\frac{1}{2\beta^2}\sum_{k=1}^{K_n}\sum_{\ell=0}^{L_n-1}\xi_{k,\ell}^2(\beta,\varrho,\mathbf{b}) \\ &-\frac{1}{2\beta^2}\sum_{k=1}^{K_n}k^{-1}\left(\sum_{j=1}^{k}\xi_{j,k+L_n-j}(\beta,\varrho,\mathbf{b})\right)^2,\end{aligned} \tag{A.6}$$

where

$$\xi_{k,\ell}(\beta,\varrho,\mathbf{b}) := g_{k,\ell}(\varrho) - \frac{\beta^2}{2}\sum_{i=0}^{2\ell}\varrho^i + \beta\sum_{j=\ell}^{J}b_j\varrho^{j-\ell} \tag{A.7}$$

with

$$g_{k,\ell}(\varrho) := \begin{cases} f_{k,\ell} - f_{k-1,\ell+1} - \varrho(f_{k,\ell-1} - f_{k-1,\ell}) & \text{for } \ \ell \geqslant 1, \\ f_{k,0} - f_{k-1,1} & \text{for } \ \ell = 0, \end{cases}$$

for all $k \geqslant 1$. Since we have $g_{k,\ell}(\varrho) = g_{k,\ell}(\varrho_0) + (\varrho_0 - \varrho)(f_{k,\ell-1} - f_{k-1,\ell})$, $\ell \geqslant 1$, by applying formula (A.2) we obtain

$$\begin{aligned} \xi_{k,\ell}(\beta, \varrho, \mathbf{b}) = \beta_0 \eta_{k,\ell} + \frac{\beta_0^2}{2} \sum_{i=0}^{2\ell} \varrho_0^i - \frac{\beta^2}{2} \sum_{i=0}^{2\ell} \varrho^i - \beta_0 \sum_{j=\ell}^{J} b_{0,j} \varrho_0^{j-\ell} + \beta \sum_{j=\ell}^{J} b_j \varrho^{j-\ell} \\ + (\varrho_0 - \varrho) \Bigg[\beta_0 \sum_{i=0}^{\ell-1} \varrho_0^{\ell-i-1} \eta_{k,i} + \frac{\beta_0^2}{2} \left(\sum_{i=0}^{\ell-1} \varrho_0^i \right)^2 \\ - \beta_0 \sum_{j=0}^{J} b_{0,j} \sum_{i=0}^{j \wedge \ell} \varrho_0^{\ell+j-2i} \Bigg]. \end{aligned} \tag{A.8}$$

We have

$$\begin{aligned} \mathsf{E}\, \xi_{k,\ell}(\beta, \varrho, \mathbf{b}) = \frac{\beta_0^2}{2} \sum_{i=0}^{2\ell} \varrho_0^i - \frac{\beta^2}{2} \sum_{i=0}^{2\ell} \varrho^i - \beta_0 \sum_{j=\ell}^{J} b_{0,j} \varrho_0^{j-\ell} + \beta \sum_{j=\ell}^{J} b_j \varrho^{j-\ell} \\ + (\varrho_0 - \varrho) \left[\frac{\beta_0^2}{2} \left(\sum_{i=0}^{\ell-1} \varrho_0^i \right)^2 - \beta_0 \sum_{j=0}^{J} b_{0,j} \sum_{i=0}^{j \wedge \ell} \varrho_0^{\ell+j-2i} \right] \to m(\beta, \varrho) \end{aligned}$$

as $\ell \to \infty$, where

$$m(\beta, \varrho) := \frac{\beta_0^2}{2(1 - \varrho_0)} - \frac{\beta^2}{2(1 - \varrho)} + \frac{(\varrho_0 - \varrho)\beta_0^2}{2(1 - \varrho_0)^2}. \tag{A.9}$$

Hence $\sup_{k,\ell} |\, \mathsf{E}\, \xi_{k,l}(\beta, \varrho, \mathbf{b})| < \infty$. Moreover,

$$\operatorname{Var} \xi_{k,\ell}(\beta, \varrho, \mathbf{b}) = \beta_0^2 \left(1 + (\varrho_0 - \varrho)^2 \sum_{i=0}^{\ell-1} \varrho_0^{2(\ell-i-1)} \right) \to \sigma^2(\varrho)$$

as $\ell \to \infty$, where

$$\sigma^2(\varrho) := \beta_0^2 \left(1 + \frac{(\varrho_0 - \varrho)^2}{1 - \varrho_0^2} \right). \tag{A.10}$$

Hence $\sup_{k,\ell} |\operatorname{Var} \xi_{k,l}(\beta, \varrho, \mathbf{b})| < \infty$. Since $\xi_{k,\ell}(\beta, \varrho, \mathbf{b})$ has a normal distribution for all $k \geqslant 1$, $\ell \geqslant 0$, we conclude $\sup_{k,\ell} \mathsf{E}\, \xi_{k,l}^8(\beta, \varrho, \mathbf{b}) < \infty$. Furthermore,

$$n^{-2} \sum_{k=1}^{K_n} \sum_{\ell=0}^{L_n-1} \mathsf{E}\, \xi_{k,\ell}^2(\beta, \varrho, \mathbf{b}) \to KL(\sigma^2(\varrho) + m^2(\beta, \varrho))$$

as $n \to \infty$. Obviously the sets $\{\xi_{k,\ell}(\beta, \varrho, \mathbf{b}) : \ell \in \mathbb{N}\}$, $k \in \mathbb{N}$, are independent, hence by Lemma B.1 we obtain

$$n^{-2} \sum_{k=1}^{K_n} \sum_{\ell=0}^{L_n-1} \xi_{k,\ell}^2(\beta, \varrho, \mathbf{b}) \to KL(\sigma^2(\varrho) + m^2(\beta, \varrho)) \qquad \text{a.s. as } \ n \to \infty. \tag{A.11}$$

Clearly $\{\xi_{j,k+L_n-j}(\beta,\varrho,\mathbf{b}) : 1 \leqslant j \leqslant k\}$ are independent for all $k, n \in \mathbb{N}$, hence

$$\mathsf{E}\left(\sum_{j=1}^{k}\xi_{j,k+L_n-j}(\beta,\varrho,\mathbf{b})\right)^2 = \sum_{j=1}^{k}\operatorname{Var}\xi_{j,k+L_n-j}(\beta,\varrho,\mathbf{b}) + \left(\sum_{j=1}^{k}\mathsf{E}\,\xi_{j,k+L_n-j}(\beta,\varrho,\mathbf{b})\right)^2.$$

Applying the above formulas for $\mathsf{E}\,\xi_{k,\ell}(\beta,\varrho,\mathbf{b})$ and $\operatorname{Var}\xi_{k,\ell}(\beta,\varrho,\mathbf{b})$ it follows that

$$n^{-2}\sum_{k=1}^{K_n}k^{-1}\,\mathsf{E}\left(\sum_{j=1}^{k}\xi_{j,k+L_n-j}(\beta,\varrho,\mathbf{b})\right)^2 \to \frac{K^2m^2(\beta,\varrho)}{2}$$

as $n \to \infty$, hence by Lemma B.3 we obtain

$$n^{-2}\sum_{k=1}^{K_n}k^{-1}\left(\sum_{j=1}^{k}\xi_{j,k+L_n-j}(\beta,\varrho,\mathbf{b})\right)^2 \to \frac{K^2m^2(\beta,\varrho)}{2} \qquad \text{a.s. as } n\to\infty. \tag{A.12}$$

Now, equations (A.11) and (A.12) lead us to

$$\begin{aligned}
&\mathcal{L}_n(\beta_0,\varrho_0,\mathbf{b}_0) - \mathcal{L}_n(\beta,\varrho,\mathbf{b})\\
&= \frac{KLn^2}{2}\left(\frac{\beta_0^2}{\beta^2} - 1 - \log\frac{\beta_0^2}{\beta^2}\right) + \frac{KL\,(\varrho_0-\varrho)^2\,\beta_0^2n^2}{2\beta^2\,(1-\varrho_0^2)} \qquad\qquad (\text{A.13})\\
&\quad + \frac{K(K+2L)n^2}{16\beta^2}\left(\frac{\beta_0^2\,(\varrho_0-\varrho)}{(1-\varrho_0)^2} + \frac{\beta_0^2}{1-\varrho_0} - \frac{\beta^2}{1-\varrho}\right)^2 + o(n^2) \qquad \text{a.s. as } n\to\infty.
\end{aligned}$$

Furthermore, notice that (A.13) holds uniformly in $(\beta,\varrho,\mathbf{b})$ over H, due to the special from of the likelihood function. We show the details of the proof of uniformity in Remark A.2.

For a fixed n, one can now consider a maximum likelihood estimator of $(\beta_0,\varrho_0,\mathbf{b}_0)$, say $(\widehat{\beta}_n,\widehat{\varrho}_n,\widehat{\mathbf{b}}_n)$, which is the maximiser of $\mathcal{L}_n(\beta,\varrho,\mathbf{b})$ over H. Hence, after replacing $(\beta,\varrho,\mathbf{b})$ by $(\widehat{\beta}_n,\widehat{\varrho}_n,\widehat{\mathbf{b}}_n)$ in (A.13) one can easily see that the left hand side is non-positive with probability one, that is a.s. $\mathcal{L}_n(\beta_0,\varrho_0,\mathbf{b}_0) - \mathcal{L}_n(\widehat{\beta}_n,\widehat{\varrho}_n,\widehat{\mathbf{b}}_n) \leqslant 0$. On the other hand the leading terms of the right hand side of (A.13) are non-negative and at least one of them is positive if $(\widehat{\beta}_n,\widehat{\varrho}_n) \neq (\beta_0,\varrho_0)$. Therefore, as $n\to\infty$, equation (A.13) can be kept only if $(\widehat{\beta}_n,\widehat{\varrho}_n) \to (\beta_0,\varrho_0)$ a.s., since the right hand side of (A.13) would not tend to 0 as $n\to\infty$ for $\omega\in\Omega$ if we had $\big(\widehat{\beta}_n(\omega),\widehat{\varrho}_n(\omega)\big) \not\to (\beta_0,\varrho_0)$. That is, the strong consistency of the maximum likelihood estimators of (β,ϱ) holds.

Now we turn to showing strong consistency of $\widehat{\mathbf{b}}_n$. Consider the system of equations determined by the first order conditions $\partial_{j+3}\mathcal{L}_n(\widehat{\beta}_n,\widehat{\varrho}_n,\widehat{\mathbf{b}}_n) = 0$ for $j = 0,1,\dots,J$. For large n (e.g. for $L_n > J$, for this recall that due the remark made on the vanishing $q_{j,k,\ell}$'s at the end of Remark A.1 only the second line of the right hand side of (2) will contain the market price of risk parameters) we can rewrite this system of equations in the simple form (see (A.6) and (A.7))

$$\sum_{k=1}^{K_n}\sum_{i=0}^{j}\xi_{k,i}(\widehat{\beta}_n,\widehat{\varrho}_n,\widehat{\mathbf{b}}_n)\widehat{\varrho}^{\,j-i} = 0 \quad \text{for } j = 0,1,\dots,J,$$

which can be reduced to

$$\sum_{k=1}^{K_n} \xi_{k,j}(\widehat{\beta}_n, \widehat{\varrho}_n, \widehat{\mathbf{b}}_n) = 0 \quad \text{for } j = 0, 1, \ldots, J. \tag{A.14}$$

Now, taking (A.14) for $j = J$ we obtain

$$\sum_{k=1}^{K_n} \left[\beta_0 \eta_{k,J} + \frac{\beta_0^2}{2} \sum_{i=0}^{2J} \varrho_0^i - \frac{\widehat{\beta}_n^2}{2} \sum_{i=0}^{2J} \widehat{\varrho}_n^{\,i} - \beta_0 b_{0,J} + \widehat{\beta}_n \, \widehat{b}_{n,J} + (\varrho_0 - \widehat{\varrho}_n) \, c_{k,J}^{(2)} \right] = 0, \tag{A.15}$$

where for $k, \ell \in \mathbb{Z}_+$, $k > 0$ we write

$$c_{k,\ell}^{(2)} := \beta_0 \sum_{i=0}^{\ell-1} \varrho_0^{\ell-i-1} \eta_{k,i} + \frac{\beta_0^2}{2} \left(\sum_{i=0}^{\ell-1} \varrho_0^i \right)^2 - \beta_0 \sum_{j=0}^{J} b_{0,j} \sum_{i=0}^{j \wedge \ell} \varrho_0^{\ell+j-2i}. \tag{A.16}$$

Notice that the random variable $c_{k,\ell}^{(2)}$ does not depend on $(\widehat{\beta}_n, \widehat{\varrho}_n, \widehat{\mathbf{b}}_n)$ and $\{c_{k,\ell}^{(2)}\}_{k \geqslant 0}$ are i.i.d. for a fixed $\ell \in \mathbb{Z}_+$. Reordering (A.15) we obtain

$$\widehat{b}_{n,J} - b_{0,J} = \frac{\beta_0 - \widehat{\beta}_n}{\widehat{\beta}_n} b_{0,J} - \frac{\beta_0^2}{2\widehat{\beta}_n} \sum_{i=0}^{2J} \varrho_0^i + \frac{\widehat{\beta}_n}{2} \sum_{i=0}^{2J} \widehat{\varrho}_n^{\,i} - \frac{1}{K_n \widehat{\beta}_n} \sum_{k=1}^{K_n} \left[\beta_0 \eta_{k,J} + (\varrho_0 - \widehat{\varrho}_n) \, c_{k,J}^{(2)} \right].$$

Hence, by the SLLN and the consistency of $(\widehat{\beta}_n, \widehat{\varrho}_n)$ we obtain that $\widehat{b}_{n,J} \to b_{0,J}$ a.s. as $n \to \infty$, i.e. $\widehat{b}_{n,J}$ is strongly consistent. In a similar way, recursively we can obtain the consistency of $\widehat{b}_{n,J-1}, \widehat{b}_{n,J-2}, \ldots, \widehat{b}_{n,1}$. Indeed, given the consistency of $\widehat{b}_{n,J-1}, \widehat{b}_{n,J-2}, \ldots, \widehat{b}_{n,j+1}$, consider again (A.14) from which we can obtain

$$\begin{aligned} \widehat{b}_{n,j} - b_{0,j} =& \frac{\beta_0 - \widehat{\beta}_n}{\widehat{\beta}_n} b_{0,j} + \sum_{i=j+1}^{J} \left(\frac{\beta_0}{\widehat{\beta}_n} \, b_{0,i} \varrho_0^{i-j} - \widehat{b}_{n,i} \, \widehat{\varrho}_n^{\,i-j} \right) \\ & - \frac{\beta_0^2}{2\widehat{\beta}_n} \sum_{i=0}^{2j} \varrho_0^i + \frac{\widehat{\beta}_n}{2} \sum_{i=0}^{2j} \widehat{\varrho}_n^{\,i} - \frac{1}{K_n \widehat{\beta}_n} \sum_{k=1}^{K_n} \left[\beta_0 \eta_{k,j} + (\varrho_0 - \widehat{\varrho}_n) \, c_{k,j}^{(2)} \right], \end{aligned}$$

from which the consistency of $\widehat{b}_{n,j}$ follows and thus the proof of Theorem 2.1 is complete. □

Remark A.2 (Uniformity in (A.13)) In the derivation of (A.13) in fact we have shown that for any fixed point $(\beta, \varrho, \boldsymbol{b}) \in H$ we have

$$n^{-2} \left(\mathcal{L}_n(\beta, \varrho, \boldsymbol{b}) + \frac{1}{2} \log(K_n!) \right) \to A(\beta, \varrho), \qquad a.s.,$$

where, recalling notations (A.9) and (A.10), the (deterministic) function A is defined as

$$A(\beta, \varrho) = -\frac{KL}{2} \log\left(2\pi\beta^2\right) - \frac{KL\left(\sigma^2(\varrho) + m^2(\beta, \varrho)\right)}{2\beta^2} - \frac{K^2 m^2(\beta, \varrho)}{4\beta^2},$$

for $(\beta,\varrho)\in\mathbb{R}^2$, $\beta\neq 0$.

Now, introducing the notations $c_\ell^{(1)}(\beta,\varrho,\boldsymbol{b}) := \frac{\beta^2\varrho^{2\ell+1}}{2(1-\varrho)}+\beta\sum_{j=\ell}^J b_j\varrho^{j-\ell}$ and $c_\ell^{(3)} := \frac{\beta_0^2}{2}\sum_{i=0}^{2\ell}\varrho_0^i - \beta_0\sum_{j=\ell}^J b_{0,j}\varrho^{j-\ell}$ we can rewrite (A.8) as

$$\xi_{k,\ell}(\beta,\varrho,\mathbf{b}) = \beta_0\eta_{k,\ell} - \frac{\beta^2}{2\,(1-\varrho)} + c_\ell^{(1)}(\beta,\varrho,\boldsymbol{b}) + (\varrho_0-\varrho)c_{k,\ell}^{(2)} + c_\ell^{(3)}, \tag{A.17}$$

where $c_{k,\ell}^{(2)}$ is given in (A.16). Notice that $c_{k,\ell}^{(2)}$ is a random variable, $c_\ell^{(3)}$ is a constant and these latter two terms depend only on $(\beta_0,\varrho_0,\boldsymbol{b}_0)$ but not on $(\beta,\varrho,\boldsymbol{b})$. In this way of writing $\xi_{k,\ell}$ we have displayed only the parts which depend on the parameters $(\beta,\varrho,\boldsymbol{b})$. We can see that this dependence is relatively simple and, say, fairly separated from the random parts.

Now take the square of $\xi_{k,\ell}(\beta,\varrho,\mathbf{b})$ based on (A.17) and substitute it in (A.6). In the followings we will consider the terms we obtain in the square of $\xi_{k,\ell}(\beta,\varrho,\mathbf{b})$. We mention that in (A.17) we displayed the term $-\frac{\beta^2}{2(1-\varrho)}$ rather than embedding it in $c_\ell^{(1)}(\beta,\varrho,\boldsymbol{b})$. The reason for that was that the terms (of the log-likelihood function) which contain $c_\ell^{(1)}(\beta,\varrho,\boldsymbol{b})$ will be shown to vanish uniformly as $n\to\infty$ unlike the terms containing $-\frac{\beta^2}{2(1-\varrho)}$.

By the application of Lemmas B.1, B.2, B.3 and their corollaries (see Appendix B) we can easily see that for $m=1,2$ the terms

$$n^{-2}\sum_{k=1}^{K_n}\sum_{\ell=0}^{L_n-1}(\eta_{k,\ell})^m, \qquad n^{-2}\sum_{k=1}^{K_n}k^{-1}\left(\sum_{j=1}^{k}\eta_{j,k+L_n-j}\right)^m,$$

$$n^{-2}\sum_{k=1}^{K_n}\sum_{\ell=0}^{L_n-1}\left(c_{k,\ell}^{(2)}\right)^m, \qquad n^{-2}\sum_{k=1}^{K_n}k^{-1}\left(\sum_{j=1}^{k}c_{j,k+L_n-j}^{(2)}\right)^m,$$

$$n^{-2}\sum_{k=1}^{K_n}\sum_{\ell=0}^{L_n-1}\eta_{k,\ell}c_{k,\ell}^{(2)}, \qquad n^{-2}\sum_{k=1}^{K_n}k^{-1}\eta_{j,k+L_n-j}c_{j,k+L_n-j}^{(2)}$$

all have an almost sure limit. Therefore, let $\Gamma_{\beta_0,\varrho_0,\boldsymbol{b}_0}\subset\Omega$ denote the set over which these terms all converge to the their limits (given by Lemmas B.1, B.2, B.3 and their corollaries). Thus, $\mathsf{P}(\Gamma_{\beta_0,\varrho_0,\boldsymbol{b}_0})=1$. We will show that the uniformity of the almost sure convergence at issue is fulfilled over this set.

Next consider the terms we obtain in (A.6) after taking the square of $\xi_{k,\ell}(\beta,\varrho,\mathbf{b})$ based on (A.17) which contain $c_\ell^{(1)}$. According to our assumptions $\sup_{(\beta,\varrho,\boldsymbol{b})\in H}|\varrho|<1$. Hence observe that $\left|\sum_{\ell=0}^{L_n-1}c_\ell^{(1)}\right|$ and $\left|k^{-1}\sum_{j=1}^{k}c_{k+L_n-j}^{(1)}\right|$ are both bounded above for k,n. Therefore for $m=1,2$ the terms

$$n^{-2}\sum_{k=1}^{K_n}\sum_{\ell=0}^{L_n-1}\left(c_\ell^{(1)}\right)^m \qquad \textit{and} \qquad n^{-2}\sum_{k=1}^{K_n}k^{-1}\left(\sum_{j=1}^{k}c_{k+L_n-j}^{(1)}\right)^m$$

vanish uniformly over H.

Furthermore, the cross product terms containing $c_\ell^{(1)}$ all vanish uniformly in H. One can see this by applying the Cauchy-Schwartz inequality. For instance,

$$\left| n^{-2} \sum_{k=1}^{K_n} \sum_{\ell=0}^{L_n-1} c_\ell^{(1)} c_{k,\ell}^{(2)} \right| \leqslant \left[n^{-2} \sum_{k=1}^{K_n} \sum_{\ell=0}^{L_n-1} \left(c_\ell^{(1)} \right)^2 \right]^{1/2} \left[n^{-2} \sum_{k=1}^{K_n} \sum_{\ell=0}^{L_n-1} \left(c_{k,\ell}^{(2)} \right)^2 \right]^{1/2} \to 0$$

as $n \to \infty$ uniformly in H.

It is easy to check that the remaining terms we obtained in (A.6) also converge almost surely and uniformly over H.

Summarising the above results we obtain that on the one hand

$$n^{-2} \left(\mathcal{L}_n(\beta, \varrho, \boldsymbol{b}) + \frac{1}{2} \log(K_n!) \right) \to A(\beta, \varrho), \quad \forall \omega \in \Gamma_{\beta_0, \varrho_0, \boldsymbol{b}_0}, \ \ \forall (\beta, \varrho, \boldsymbol{b}) \in H, \tag{A.18}$$

and on the other hand the uniformity of the convergences detailed in the last paragraphs imply that (A.18) holds uniformly over H, which means that the almost sure expansion (A.13) holds uniformly in $(\beta, \varrho, \boldsymbol{b}) \in H$, indeed.

Proof of Theorem 2.2. We apply again Taylor's expansion for the gradient vector of $\mathcal{L}_n(\beta, \varrho, b)$ up to order 2. Write

$$\begin{bmatrix} n^{-1}\partial_1 \mathcal{L}_n(\widehat{\beta}_n, \widehat{\varrho}_n, \widehat{\mathbf{b}}_n) \\ n^{-1}\partial_2 \mathcal{L}_n(\widehat{\beta}_n, \widehat{\varrho}_n, \widehat{\mathbf{b}}_n) \\ n^{-1/2}\boldsymbol{\partial_3} \mathcal{L}_n(\widehat{\beta}_n, \widehat{\varrho}_n, \widehat{\mathbf{b}}_n) \end{bmatrix} - \begin{bmatrix} n^{-1}\partial_1 \mathcal{L}_n(\beta_0, \varrho_0, \mathbf{b}_0) \\ n^{-1}\partial_2 \mathcal{L}_n(\beta_0, \varrho_0, \mathbf{b}_0) \\ n^{-1/2}\boldsymbol{\partial_3} \mathcal{L}_n(\beta_0, \varrho_0, \mathbf{b}_0) \end{bmatrix} = (\mathbf{A}_n + \mathbf{B}_n) \begin{bmatrix} n(\widehat{\beta} - \beta_0) \\ n(\widehat{\varrho} - \varrho_0) \\ \sqrt{n}(\widehat{\mathbf{b}} - \mathbf{b}_0) \end{bmatrix} \tag{A.19}$$

where $\mathbf{A}_n$ and $\mathbf{B}_n$ are $(J+3) \times (J+3)$ matrices defined as follows. Write $\mathbf{A}_n := (a_{i,j}^n)_{i,j=1,\ldots,J+3}$ and define

$$a_{i,j}^n := n^{l_{i,j}} \partial_i \partial_j \mathcal{L}_n(\beta_0, \varrho_0, \mathbf{b}_0),$$

where

$$l_{i,j} := \begin{cases} -2 & \text{if } i \vee j \leqslant 2 \\ -1 & \text{if } i \wedge j \geqslant 3 \\ -3/2 & \text{otherwise.} \end{cases}$$

Denoting the ith row of $\mathbf{B}_n$ by B_n^i we will write it in the form $B_n^i = D_n^\top R_n^i$, where the superscript $^\top$ denotes the transposed,

$$D_n^\top := (\widehat{\beta}_n - \beta_0, \ \widehat{\varrho}_n - \varrho_0, \ \widehat{\mathbf{b}}_n - \mathbf{b}_0),$$

$$R_n^i := \left(r_{j_1,j_2}^{n,i} \right)_{j_1,j_2=1,2,\ldots,J+3}$$

and

$$r_{j_1,j_2}^{n,i} := \frac{1}{2} n^{l_{i,j_2}} \partial_i \partial_{j_1} \partial_{j_2} \mathcal{L}_n(\tilde{\beta}, \tilde{\varrho}, \tilde{\mathbf{b}})$$

with appropriate $(\tilde{\beta}, \tilde{\varrho}, \tilde{\mathbf{b}})$ taking values —coordinate-wise— between $(\beta_0, \varrho_0, \mathbf{b}_0)$ and $(\widehat{\beta}_n, \widehat{\varrho}_n, \widehat{\mathbf{b}}_n)$.

Under the assumptions of Theorem 2.1 we will need the following lemmas on the terms introduced in (A.19). The proofs of these lemmas follow this proof.

Lemma A.1 *Under the assumptions of Theorem 2.1 we have*

$$\begin{bmatrix} n^{-1}\partial_1\mathcal{L}_n(\beta_0, \varrho_0, \mathbf{b}_0) \\ n^{-1}\partial_2\mathcal{L}_n(\beta_0, \varrho_0, \mathbf{b}_0) \\ n^{-1/2}\boldsymbol{\partial_3}\mathcal{L}_n(\beta_0, \varrho_0, \mathbf{b}_0) \end{bmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma), \tag{A.20}$$

where $\Sigma = (\sigma_{i,j})_{i,j=1,\dots,J+3}$ *with* $\sigma_{i,j}$ *for* $i \vee j \leqslant 2$ *defined in* (6) *and* (7) *in Theorem 2.2,*

$$\sigma_{i+3,j+3} = K \sum_{k=0}^{i\wedge j} \varrho^{i+j-2k} \quad \text{for } i, j = 0, 1, \dots, J,$$

and the remaining entries of Σ *are zero.*

Lemma A.2

$$\mathbf{A}_n \to -\Sigma \qquad \text{a.s. as } n \to \infty.$$

Lemma A.3 *For* $i = 1, 2, \dots, J+3$ *we have*

$$B_n^i \xrightarrow{\mathsf{P}} 0.$$

Clearly, the first term on the left hand side of (A.19) tends to zero almost surely, in fact it takes value 0 a.s. for large n due to Theorem 2.1. Hence, by Slutsky's Lemma the limit distribution of the left hand side of (A.19) is $\mathcal{N}(0, \Sigma)$, which is given in Lemma A.1. Lemma A.2 and Lemma A.3 together with Slutsky's Lemma give

$$A_n + B_n \xrightarrow{\mathsf{P}} -\Sigma.$$

Having these asymptotic results and recalling (A.19) we can apply Lemma B.5 to obtain (5). For this note that $\Lambda = \Sigma^{-1}$ where Σ is given in Lemma A.1. Thus, for the proof of Theorem 2.2 there remains to prove Lemma A.1, Lemma A.2 and Lemma A.3. □

Proof of Lemma A.1. First we will show that

$$\mathbf{Z}_n := \begin{bmatrix} n^{-1}\partial_1\mathcal{L}_n(\beta_0, \varrho_0, \boldsymbol{b}_0) \\ n^{-1}\partial_2\mathcal{L}_n(\beta_0, \varrho_0, \boldsymbol{b}_0) \\ n^{-1/2}\boldsymbol{\partial}_3\mathcal{L}_n(\beta_0, \varrho_0, \boldsymbol{b}_0) \end{bmatrix}$$

can be considered as a martingale with respect to an appropriate filtration. Namely, rewriting the terms of $\mathbf{Z}_n$ in an appropriate order we will obtain the form

$$\mathbf{Z}_n = \sum_{m=1}^{K_n(L_n+1)} M_m^{(n)},$$

where $M_m^{(n)}$ is defined below. The idea of reordering the terms is simple: starting with and fixing $k = 0$ we increase ℓ step by step (as m increases) from 0 to L_n. When $\ell = L_n$ is reached after $L_n + 1$ steps than we consider the next value of k $(= 1)$ and we take again the possible values of ℓ from 0 to L_n. We continue this as long as $k = K_n$ is reached. Thus the number of summands is $K_n(L_n + 1)$. This means that in each step a new $\eta_{k,\ell}$ will occur in the martingale sum (which is independent of the previous terms). The case $\ell = L_n$ is a little bit special, since it involves a number of new variables, namely, η_{k,L_n}, η_{k-1,L_n+1}, ... η_{1,L_n+k-1} (which are also independent of the previous terms).

Let us turn now to the rigorous definition of the martingale difference $M_m^{(n)}$, and the corresponding filtration. Fix $n \in \mathbb{N}$ and notice that for any positive integer m there exist uniquely determined integers k_m, ℓ_m such that $m = (k_m - 1)(L_n + 1) + \ell_m + 1$ with $0 < k_m$ and $0 \leqslant \ell_m \leqslant L_n$. We remark that k_m and ℓ_m depend on n as well, however, for simplicity we omit to denote their dependence on n. Now, define $\mathcal{G}_0^{(n)} := \{\emptyset, \Omega\}$ and

$$\mathcal{G}_m^{(n)} := \begin{cases} \sigma\left\{\mathcal{G}_{m-1}^{(n)} \bigcup \sigma\left\{\eta_{k_m,\ell_m}\right\}\right\} & \text{if } 0 \leqslant \ell_m < L_n, \\ \sigma\left\{\mathcal{G}_{m-1}^{(n)} \bigcup \sigma\left\{\eta_{k_m-i,L_n+i} \mid 0 \leqslant i < k_m\right\}\right\} & \text{if } \ell_m = L_n. \end{cases}$$

Furthermore, write $M_m^{(n)} := (M_m^{1,(n)}, M_m^{2,(n)}, \ldots, M_m^{J+3,(n)})^\top$ and

$$\bar{\eta}_{k,L_n} := \sum_{j=0}^{k-1} \eta_{j+1,k+L_n-j-1}, \qquad \tilde{\eta}_{k,L_n} := \sum_{j=0}^{k-1} \sum_{i=0}^{k+L_n-j-1} \varrho_0^{k+L_n-j-i-1} \eta_{j+1,i}. \tag{A.21}$$

For the sake of convenience and better readability, in the following definition of $M_m^{i,(n)}$ we will simply write ℓ instead of ℓ_m and k instead of k_m. Define

$$M_m^{1,(n)} := \begin{cases} \frac{1}{n\beta_0}\left[\eta_{k,\ell}^2 - 1 + \eta_{k,\ell}\left(\beta_0 \sum_{j=0}^{2\ell} \varrho_0^j - \sum_{j=\ell}^J b_{0,j}\varrho_0^{j-\ell}\right)\right] & \text{if } 0 \leqslant \ell < L_n, \\ \frac{1}{n\beta_0 k}\left[\bar{\eta}_{k,L_n}^2 - k + \bar{\eta}_{k,L_n}\left(\beta_0 \sum_{j=0}^{k-1} \sum_{i=0}^{2(k+L_n-j-1)} \varrho_0^i\right)\right] & \text{if } \ell = L_n, \end{cases}$$

$$M_m^{2,(n)} := \begin{cases} \frac{1}{n}\eta_{k,\ell}\Bigg[\sum_{j=0}^{\ell-1} \varrho_0^{\ell-j-1}\eta_{k,j} + \frac{\beta_0}{2}\left(\sum_{j=0}^{\ell-1}\varrho_0^j\right)^2 + \frac{\beta_0}{2}\sum_{j=1}^{2\ell} j\varrho_0^{j-1} \\ \qquad -\beta_0 \sum_{j=0}^J b_{0,j} \sum_{i=0}^{j\wedge(\ell-1)} \varrho_0^{\ell+j-1-2i} - \beta_0 \sum_{j=\ell+1}^J b_{0,j}(j-\ell)\varrho_0^{j-\ell-1}\Bigg] \\ \qquad\qquad \text{if } 0 \leqslant \ell < L_n, \\ \frac{1}{nk}\bar{\eta}_{k,L_n}\Bigg[\tilde{\eta}_{k,L_n} + \frac{\beta_0}{2}\sum_{j=0}^{k-1}\left(\sum_{i=0}^{k+L_n-j-2}\varrho_0^i\right)^2 + \frac{\beta_0}{2}\sum_{j=0}^{k-1}\sum_{i=1}^{2(k+L_n-j-1)} i\varrho_0^{i-1} \\ \qquad - \sum_{i=0}^{k-1}\sum_{j=0}^J b_{0,j}\sum_{n=0}^j \varrho_0^{j+k+L_n-i-1-2n}\Bigg] \\ \qquad\qquad \text{if } \ell = L_n, \end{cases}$$

and for $j = 0, \ldots, J$

$$M_m^{j+3,(n)} := \begin{cases} -\mathbf{1}_{\{j \geqslant \ell\}} n^{-1/2} \varrho_0^{j-\ell}\eta_{k,\ell} & \text{if } 0 \leqslant \ell < L_n, \\ 0 & \text{if } \ell = L_n. \end{cases}$$

Now, by the independence of the $\eta_{k,l}$'s, it is easy to see that

$$\mathsf{E}\left(M_m^{(n)} | \mathcal{G}_{m-1}^{(n)}\right) = 0$$

for $1 \leqslant m \leqslant K_n(L_n+1)$. Hence, we can see that $\left(M_m^{(n)}\right)_{1\leqslant m\leqslant K_n(L_n+1)}$ are martingale differences with respect to the filtration $\left(\mathcal{G}_m^{(n)}\right)_{0\leqslant m\leqslant K_n(L_n+1)}$. Furthermore, recalling (A.1), (A.2), (A.3) and (A.4) we can see that for sufficiently large n (s.t. $L_n > J$) we clearly have

$$\mathbf{Z}_n = \sum_{m=1}^{K_n(L_n+1)} M_m^{(n)}.$$

Next observe that the sequence consisting of the conditional covariances of $M_m^{(n)}$ tends to Σ in probability, i.e.

$$\gamma_{i,j}^{(n)} := \sum_{m=1}^{K_n(L_n+1)} \mathsf{E}\left(M_m^{i,(n)} M_m^{j,(n)} \mid \mathcal{G}_{m-1}^{(n)}\right) \xrightarrow{\mathsf{P}} \sigma_{i,j} \qquad \text{for } i,j = 1,2,\ldots,J+3 \tag{A.22}$$

as well as the conditional Liapounov condition holds, i.e.

$$\sum_{m=1}^{K_n(L_n+1)} \mathsf{E}\left(\left\|M_m^{(n)}\right\|^4 \mid \mathcal{G}_{m-1}^{(n)}\right) \xrightarrow{\mathsf{P}} 0. \tag{A.23}$$

In fact, we will show more: the convergence results in (A.22) and (A.23) are valid even in almost sure sense. For what follows (for the martingale limit theorem that we shall apply), however, the convergence in probability is sufficient.

To show how to check (A.22) we only demonstrate two cases. Firstly, for $i = 1$ and $j = 2$ write

$$\delta_{k,\ell} := \sum_{i=0}^{\ell-1} \varrho_0^{\ell-i-1}\eta_{k,i} + \frac{\beta_0}{2}\left(\sum_{i=1}^{\ell-1}\varrho_0^i\right)^2 + \frac{\beta_0}{2}\sum_{i=1}^{2\ell} i\varrho_0^{i-1}, \tag{A.24}$$

$$\tilde{\delta}_{k,L_n} := \tilde{\eta}_{k,L_n} + \frac{\beta_0}{2}\sum_{j=0}^{k-1}\left(\sum_{i=0}^{k+L_n-j-2}\varrho_0^i\right)^2 + \frac{\beta_0}{2}\sum_{j=0}^{k-1}\sum_{i=1}^{2(k+L_n-j-1)} i\varrho_0^{i-1}. \tag{A.25}$$

Thus we obtain

$$\begin{aligned}
\gamma_{1,2}^{(n)} &= \frac{1}{\beta_0 n^2}\sum_{k=1}^{K_n}\sum_{\ell=0}^{L_n-1}\delta_{k,\ell}\,\mathsf{E}\,\eta_{k,\ell}\left(\eta_{k,\ell}^2 - 1 + \beta_0\eta_{k,\ell}\sum_{i=0}^{2\ell}\varrho_0^i\right) \\
&\quad + \frac{1}{\beta_0 n^2}\sum_{k=1}^{K_n}\frac{1}{k^2}\tilde{\delta}_{k,L_n}\,\mathsf{E}\,\bar{\eta}_{k,L_n}\left(\bar{\eta}_{k,L_n}^2 - k + \beta_0\bar{\eta}_{k,L_n}\sum_{j=0}^{k-1}\sum_{i=0}^{2(k+L_n-j-1)}\varrho_0^i\right) + o(1) \\
&= \frac{KL}{\beta_0}\left[\frac{\beta_0}{2(1-\varrho_0)^2} + \frac{\beta_0}{2(1-\varrho_0)^2}\right]\frac{\beta_0}{1-\varrho_0} \\
&\quad + \frac{1}{\beta_0 n^2}\sum_{k=1}^{K_n}\frac{1}{k^2}\left[\frac{\beta_0 k}{2(1-\varrho_0)^2} + \frac{\beta_0 k}{2(1-\varrho_0)^2}\right]\frac{\beta_0 k^2}{1-\varrho_0} + o(1) \to \sigma_{1,2}
\end{aligned}$$

a.s. as $n \to \infty$. Note that all the terms containing the market price of risk parameters vanish, i.e. their order is $o(n^2)$, hence we omit to display these terms. Secondly, take $i, j \in \{0, 1, \dots, J\}$ and consider $\gamma^{(n)}_{i+3,\, j+3}$. Now we obtain

$$\gamma^{(n)}_{i+3,\, j+3} = n^{-1} K_n \sum_{\ell=0}^{i \wedge j} \varrho_0^{i+j-2\ell} \, \mathsf{E}\, \eta^2_{k,\ell} \to \sigma_{i+3,j+3} \quad \text{a.s. as } n \to \infty.$$

The remaining cases can be derived in a similar way.

To show (A.23) notice that even

$$n^z \sum_{m=1}^{K_n(L_n+1)} \mathsf{E}\left(\left(M_m^{i,(n)} \right)^4 \mid \mathcal{G}^{(n)}_{m-1} \right)$$

has an almost sure limit, where $z = 2$ for $i = 1, 2$, and $z = 1$ for $3 \leqslant i \leqslant J + 3$. This can be shown easily by the application of Lemmas B.1, B.2, B.3 and their corollaries. From this (A.23) is immediate.

Finally, it is known that according to the martingale limit theorem (see Theorem VIII.3.33. in [9]) that (A.22) and (A.23) are together sufficient to imply (A.20). □

Proof of Lemma A.2. First consider the case $i \vee j \leqslant 2$. Then one can easily show that

$$\frac{1}{n^2} \mathsf{E}\, \partial_i \partial_j \mathcal{L}_n(\beta_0, \varrho_0, \mathbf{b}_0) \to -\sigma_{i,j} \quad \text{as } n \to \infty.$$

Hence, by Lemma B.1, Lemma B.3 (with $\kappa = 2$) and/or by the corollaries following from these lemmas one can easily see that

$$\frac{1}{n^2} \partial_i \partial_j \mathcal{L}_n(\beta_0, \varrho_0, \mathbf{b}_0) \to -\sigma_{i,j} \quad \text{a.s. as } n \to \infty.$$

To demonstrate the method, consider the most complicated case, where $i = 1$, $j = 2$. Recalling notations (A.21), (A.24) and (A.25) we have a.s.

$$\begin{aligned}
\partial_1 \partial_2 \mathcal{L}_n(\beta_0, \varrho_0, \mathbf{b}_0) = &- \sum_{k=1}^{K_n} \sum_{\ell=0}^{L_n-1} \eta_{k,\ell} \left[\frac{2}{\beta_0} \delta_{k,\ell} - \sum_{i=1}^{2\ell} i \varrho_0^{i-1} \right] + \sum_{k=1}^{K_n} \sum_{\ell=0}^{L_n-1} \delta_{k,\ell} \sum_{i=1}^{2\ell} \varrho_0^i \\
&- \sum_{k=1}^{K_n} \frac{1}{k} \overline{\eta}_{k,L_n} \left[\frac{2}{\beta_0} \tilde{\delta}_{k,L_n} - \sum_{j=0}^{k-1} \sum_{i=1}^{2(k+L_n-j-1)} i \varrho_0^{i-1} \right] \\
&+ \sum_{k=1}^{K_n} \frac{1}{k} \tilde{\delta}_{k,L_n} \sum_{j=0}^{k-1} \sum_{i=0}^{2(k+L_n-j-1)} \varrho_0^i + o(n^2).
\end{aligned}$$

Note that all the terms containing the market price of risk parameters vanish, i.e. their order is $o(n^2)$, hence we omit to display these terms. For the expected values we have

$$\frac{1}{n^2}\,\mathsf{E}\,\partial_1\partial_2\mathcal{L}_n(\beta_0,\varrho_0,\mathbf{b}_0) = -\frac{\beta_0}{n^2}\sum_{k=1}^{K_n}\sum_{\ell=0}^{L_n-1}\left[\sum_{i=0}^{2\ell}\varrho_0^i\right]\left[\frac{1}{2}\left(\sum_{i=0}^{\ell-1}\varrho_0^i\right)^2+\frac{1}{2}\sum_{i=0}^{2\ell}i\varrho_0^{i-1}\right]$$

$$-\frac{\beta_0}{n^2}\sum_{k=1}^{K_n}\frac{1}{k}\left[\sum_{j=0}^{k-1}\sum_{i=1}^{2(k+L_n-j-1)}\varrho_0^i\right]$$

$$\times\left[\frac{1}{2}\sum_{j=0}^{k-1}\left(\sum_{i=0}^{k+L_n-j-2}\varrho_0^i\right)^2+\frac{1}{2}\sum_{j=0}^{k-1}\sum_{i=1}^{2(k+L_n-j-1)}i\varrho_0^{i-1}\right]+o(1)$$

$$=-\frac{\beta_0 KL}{1-\varrho_0}\left[\frac{1}{2}\frac{1}{(1-\varrho_0)^2}+\frac{1}{2}\frac{1}{(1-\varrho_0)^2}\right]$$

$$-\frac{\beta_0 K^2}{2\,(1-\varrho_0)}\left[\frac{1}{2}\frac{1}{(1-\varrho_0)^2}+\frac{1}{2}\frac{1}{(1-\varrho_0)^2}\right]+o(1)\to-\sigma_{1,2}$$

as $n\to\infty$.

Now consider the case $i\vee j>2$. Then in a similar way one can easily show that

$$\frac{1}{n}\,\mathsf{E}\,\partial_i\partial_j\mathcal{L}_n(\beta_0,\varrho_0,\mathbf{b}_0)\to\lambda\quad\text{as } n\to\infty,$$

where $\lambda\in\mathbb{R}$ and $\lambda=-\sigma_{i,j}$ for $i\wedge j>2$. Hence, by the application of Lemma B.2, Lemma B.3 (with $\kappa=1$) one can easily see that

$$\frac{1}{n}\partial_i\partial_j\mathcal{L}_n(\beta_0,\varrho_0,\mathbf{b}_0)\to-\sigma_{i,j}\quad\text{a.s. as } n\to\infty$$

for $i\wedge j>2$, and

$$\frac{1}{n^{3/2}}\partial_i\partial_j\mathcal{L}_n(\beta_0,\varrho_0,\mathbf{b}_0)\to 0=-\sigma_{i,j}\quad\text{a.s. as } n\to\infty$$

for $i\wedge j\leqslant 2$. For instance, taking $i,j\in\{0,1,\dots,J\}$ we have

$$\partial_{i+3}\partial_{j+3}\mathcal{L}_n(\beta_0,\varrho_0,\mathbf{b}_0)=-K_n\sum_{\ell=0}^{i\wedge j}\varrho_0^{i+j-2\ell},$$

from which the statement is immediate. The remaining cases can easily be calculated in a similar way. □

Proof of Lemma A.3. Based on Lemma B.1, Lemma B.2, Lemma B.3 one can show that

$$\frac{1}{2}n^{l_{i,j_2}}\partial_i\partial_{j_1}\partial_{j_2}\mathcal{L}_n(\beta,\varrho,\mathbf{b})$$

has an almost sure limit uniformly in $(\beta,\varrho,\mathbf{b})\in H$. This can be shown similarly to the uniform convergence in (A.13). (Recall also Remark A.2 for this. Notice that in fact the higher order derivatives of the likelihood function will have at most the same speed of convergence as the first order ones in their asymptotic expansion due to its relatively easy dependence on the parameters.) Thus, by Lemma B.4 we obtain that $r^{n,i}_{j_1,j_2}$ is stochastically bounded for all $1 \leqslant j_1, j_2 \leqslant J+3$. On the other hand, the estimators $\widehat{\beta}_n, \widehat{\varrho}_n, \widehat{\mathbf{b}}_n$ are proved to be strongly consistent and thus $D_n^\top \to 0$ a.s. as $n \to \infty$. Hence we obtain that $B_n^i = D_n^\top R_n^i$ converges to zero in probability. □

Appendix B

In what follows we summarise some simple but useful lemmas that are often used in the proofs of the main results. They give some general statements which are not model specific (which was the reason for presenting them in a separate appendix).

Lemma B.1 *Let $\xi_{k,\ell,n}$, $k, \ell, n \in \mathbb{N}$, be random variables such that for each $n \in \mathbb{N}$ the sets $\{\xi_{k,\ell,n} : \ell \in \mathbb{N}\}$, $k \in \mathbb{N}$, are independent (i.e., the σ-algebras $\sigma(\xi_{k,\ell,n} : \ell \in \mathbb{N})$, $k \in \mathbb{N}$, are independent), and $\sup_{k,\ell,n\in\mathbb{N}} \mathsf{E}\,\xi^8_{k,\ell,n} < \infty$. Let K_n, L_n, $n \in \mathbb{N}$, be positive integers such that $K_n = nK + o(n)$ and $L_n = nL + o(n)$ as $n \to \infty$ with some $K > 0$ and $L > 0$. Then*

$$n^{-2} \sum_{k=1}^{K_n} \sum_{\ell=1}^{L_n} \left(\xi^2_{k,\ell,n} - \mathsf{E}\,\xi^2_{k,\ell,n}\right) \to 0 \qquad a.s. \text{ as } n \to \infty.$$

Proof. It suffices to show that for all $\varepsilon > 0$ we have

$$\sum_{n=1}^{\infty} \mathsf{P}(|\zeta_n| > \varepsilon n^2) < \infty,$$

where

$$\zeta_n := \sum_{k=1}^{K_n} \sum_{\ell=1}^{L_n} \left(\xi^2_{k,\ell,n} - \mathsf{E}\,\xi^2_{k,\ell,n}\right).$$

By Markov inequality we obtain $\mathsf{P}(|\zeta_n| > \varepsilon n^2) \leqslant \varepsilon^{-4} n^{-8}\,\mathsf{E}\,\zeta_n^4$, hence it is enough to show that $\mathsf{E}\,\zeta_n^4 = O(n^{7-\delta})$ as $n \to \infty$ with some $\delta > 0$. We have

$$\mathsf{E}\,\zeta_n^4 = \sum_{k_1,k_2,k_3,k_4=1}^{K_n} \sum_{\ell_1,\ell_2,\ell_3,\ell_4=1}^{L_n} \mathsf{E}\,\zeta_{k_1,\ell_1,n}\zeta_{k_2,\ell_2,n}\zeta_{k_3,\ell_3,n}\zeta_{k_4,\ell_4,n},$$

where $\zeta_{k,\ell,n} := \xi^2_{k,\ell,n} - \mathsf{E}\,\xi^2_{k,\ell,n}$. By the Cauchy–Schwartz inequality

$$|\,\mathsf{E}\,\zeta_{k_1,\ell_1,n}\zeta_{k_2,\ell_2,n}\zeta_{k_3,\ell_3,n}\zeta_{k_4,\ell_4,n}| \leqslant \left(\mathsf{E}\,\zeta^4_{k_1,\ell_1,n}\,\mathsf{E}\,\zeta^4_{k_2,\ell_2,n}\,\mathsf{E}\,\zeta^4_{k_3,\ell_3,n}\,\mathsf{E}\,\zeta^4_{k_4,\ell_4,n}\right)^{1/4}.$$

Moreover

$$\mathsf{E}\,\zeta^4_{k,\ell,n} = \mathsf{E}(\xi^2_{k,\ell,n} - \mathsf{E}\,\xi^2_{k,\ell,n})^4 \leqslant 2^3\left(\mathsf{E}\,\xi^8_{k,\ell,n} + (\mathsf{E}\,\xi^2_{k,\ell,n})^4\right) \leqslant 16\,\mathsf{E}\,\xi^8_{k,\ell,n} \leqslant 16 M_8,$$

where $M_8 := \sup_{k,\ell,n\in\mathbb{N}} \mathsf{E}\,\xi_{k,\ell,n}^8 < \infty$ by the assumptions. Hence we conclude

$$|\,\mathsf{E}\,\zeta_{k_1,\ell_1,n}\zeta_{k_2,\ell_2,n}\zeta_{k_3,\ell_3,n}\zeta_{k_4,\ell_4,n}| \leqslant 16M_8.$$

By the assumptions the sets $\{\zeta_{k,\ell,n} : \ell \in \mathbb{N}\}$, $k \in \mathbb{N}$, are independent for each $n \in \mathbb{N}$, and $\mathsf{E}\,\zeta_{k,\ell,n} = 0$ for all $k, \ell, n \in \mathbb{N}$, hence

$$\begin{aligned}\mathsf{E}\,\zeta_n^4 = \sum_{\ell_1,\,\ell_2,\,\ell_3,\,\ell_4\,=\,1}^{L_n} \Bigg(&\sum_{k=1}^{K_n} \mathsf{E}\,\zeta_{k,\ell_1,n}\zeta_{k,\ell_2,n}\zeta_{k,\ell_3,n}\zeta_{k,\ell_4,n} \\ &+ 6 \sum_{1\leqslant k_1<k_2\leqslant K_n} \mathsf{E}\,\zeta_{k_1,\ell_1,n}\zeta_{k_1,\ell_2,n}\zeta_{k_2,\ell_3,n}\zeta_{k_2,\ell_4,n}\Bigg).\end{aligned} \tag{B.1}$$

Consequently we obtain $\mathsf{E}\,\zeta_n^4 = O(n^6)$ as $n \to \infty$. □

Lemma B.2 *Let $\xi_{k,n}$, $k, n \in \mathbb{N}$, be random variables such that for each $n \in \mathbb{N}$ the random variables $\xi_{k,n}$, $k \in \mathbb{N}$, are independent and $\sup_{k,n\in\mathbb{N}} \mathsf{E}\,\xi_{k,n}^4 < \infty$. Let K_n, $n \in \mathbb{N}$, be positive integers such that $K_n = nK + o(n)$ as $n \to \infty$ with some $K > 0$. Then*

$$n^{-1} \sum_{k=1}^{K_n} (\xi_{k,n} - \mathsf{E}\,\xi_{k,n}) \to 0 \qquad \text{a.s. as } n \to \infty.$$

Proof. This statement can be proved almost readily in the same way as Lemma B.1. □

Corollary B.1 *Let $\xi_{k,\ell,n}$ and $\zeta_{k,\ell,n}$, $k, \ell, n \in \mathbb{N}$, be random variables such that for each $n \in \mathbb{N}$ the sets $\{\xi_{k,\ell,n}, \zeta_{k,\ell,n} : \ell \in \mathbb{N}\}$, $k \in \mathbb{N}$, are independent (i.e., the σ-algebras $\sigma(\xi_{k,\ell,n}, \zeta_{k,\ell,n} : \ell \in \mathbb{N})$, $k \in \mathbb{N}$, are independent), and $\sup_{k,\ell,n\in\mathbb{N}} \mathsf{E}(\xi_{k,\ell,n}^8 + \zeta_{k,\ell,n}^8) < \infty$. Let K_n, L_n, $n \in \mathbb{N}$, be positive integers such that $K_n = nK + o(n)$ and $L_n = nL + o(n)$ as $n \to \infty$ with some $K > 0$ and $L > 0$. Then*

$$n^{-2} \sum_{k=1}^{K_n} \sum_{\ell=1}^{L_n} (\xi_{k,\ell,n}\zeta_{k,\ell,n} - \mathsf{E}\,\xi_{k,\ell,n}\zeta_{k,\ell,n}) \to 0 \qquad \text{a.s. as } n \to \infty.$$

Proof. Clearly

$$\begin{aligned}&\xi_{k,\ell,n}\zeta_{k,\ell,n} - \mathsf{E}\,\xi_{k,\ell,n}\zeta_{k,\ell,n} \\ &\quad = \frac{1}{4}\Big[\Big\{(\xi_{k,\ell,n} + \zeta_{k,\ell,n})^2 - \mathsf{E}(\xi_{k,\ell,n} + \zeta_{k,\ell,n})^2\Big\} - \Big\{(\xi_{k,\ell,n} - \zeta_{k,\ell,n})^2 - \mathsf{E}(\xi_{k,\ell,n} - \zeta_{k,\ell,n})^2\Big\}\Big],\end{aligned}$$

and we can apply Lemma B.1 for $\{\xi_{k,\ell,n} + \zeta_{k,\ell,n} : k, \ell, n \in \mathbb{N}\}$ and $\{\xi_{k,\ell,n} - \zeta_{k,\ell,n} : k, \ell, n \in \mathbb{N}\}$.
□

Corollary B.2 *Let $\xi_{k,\ell,n}$, $k, \ell, n \in \mathbb{N}$, be random variables such that for each $n \in \mathbb{N}$ the sets $\{\xi_{k,\ell,n} : \ell \in \mathbb{N}\}$, $k \in \mathbb{N}$, are independent (i.e., the σ-algebras $\sigma(\xi_{k,\ell,n} : \ell \in \mathbb{N})$, $k \in \mathbb{N}$, are independent), and $\sup_{k,\ell,n\in\mathbb{N}} \mathsf{E}\, \xi_{k,\ell,n}^8 < \infty$. Let K_n, L_n, $n \in \mathbb{N}$, be positive integers such that $K_n = nK + o(n)$ and $L_n = nL + o(n)$ as $n \to \infty$ with some $K > 0$ and $L > 0$. Then*

$$n^{-2} \sum_{k=1}^{K_n} \sum_{\ell=1}^{L_n} (\xi_{k,\ell,n} - \mathsf{E}\, \xi_{k,\ell,n}) \to 0 \qquad \text{a.s. as} \quad n \to \infty.$$

Proof. Corollary B.1 applies with $\zeta_{k,\ell,n} = 1$, $k, \ell, n \in \mathbb{N}$. □

Lemma B.3 *Let $\kappa \in \{1, 2\}$. Let $\xi_{k,j,n}$, $k, j, n \in \mathbb{N}$, be random variables such that for each $n \in \mathbb{N}$ the sets $\{\xi_{k,j,n} : k \in \mathbb{N}\}$, $j \in \mathbb{N}$, are independent (i.e., the σ-algebras $\sigma(\xi_{k,j,n} : k \in \mathbb{N})$, $j \in \mathbb{N}$, are independent), and $\sup_{k,j,n\in\mathbb{N}} \mathsf{E}\, \xi_{k,j,n}^{4\kappa} < \infty$. Let K_n, $n \in \mathbb{N}$, be positive integers such that $K_n = nK + o(n)$ as $n \to \infty$ with some $K > 0$. Then*

$$n^{-\kappa} \sum_{k=1}^{K_n} k^{-1} \left[\left(\sum_{j=1}^{k} \xi_{k,j,n} \right)^{\kappa} - \mathsf{E} \left(\sum_{j=1}^{k} \xi_{k,j,n} \right)^{\kappa} \right] \to 0 \qquad \text{a.s. as } n \to \infty.$$

Proof. Consider the case $\kappa = 2$. Clearly

$$\left(\sum_{j=1}^{k} \xi_{k,j,n} \right)^2 - \mathsf{E} \left(\sum_{j=1}^{k} \xi_{k,j,n} \right)^2 = \sum_{j_1=1}^{k} \sum_{j_2=1}^{k} (\xi_{k,j_1,n}\xi_{k,j_2,n} - \mathsf{E}\, \xi_{k,j_1,n}\xi_{k,j_2,n}).$$

As in the proof of Lemma B.1 it suffices to show that $\mathsf{E}\, \zeta_n^4 = O(n^{7-\delta})$ as $n \to \infty$ with some $\delta > 0$, where

$$\zeta_n := \sum_{k=1}^{K_n} k^{-1} \sum_{j_1=1}^{k} \sum_{j_2=1}^{k} \zeta_{k,j_1,j_2,n}$$

with

$$\zeta_{k,j_1,j_2,n} := \xi_{k,j_1,n}\xi_{k,j_2,n} - \mathsf{E}\, \xi_{k,j_1,n}\xi_{k,j_2,n}.$$

We have

$$\begin{aligned}\zeta_n^4 = \sum_{k_1,\, k_2,\, k_3,\, k_4 \,=\, 1}^{K_n} \Bigg[&(k_1k_2k_3k_4)^{-1} \\ &\times \sum_{j_1,\, j_2 \,=\, 1}^{k_1} \sum_{j_3,\, j_4 \,=\, 1}^{k_2} \sum_{j_5,\, j_6 \,=\, 1}^{k_3} \sum_{j_7,\, j_8 \,=\, 1}^{k_4} \mathsf{E}\, \zeta_{k_1,j_1,j_2,n}\zeta_{k_2,j_3,j_4,n}\zeta_{k_3,j_5,j_6,n}\zeta_{k_4,j_7,j_8,n} \Bigg].\end{aligned} \tag{B.2}$$

As in the proof of Lemma B.1 we obtain

$$|\, \mathsf{E}\, \zeta_{k_1,j_1,j_2,n}\zeta_{k_2,j_3,j_4,n}\zeta_{k_3,j_5,j_6,n}\zeta_{k_4,j_7,j_8,n}| \leqslant 16M_8,$$

where $M_8 := \sup_{k,j,n\in\mathbb{N}} \mathsf{E}\,\xi_{k,j,n}^8 < \infty$ by the assumptions. By the independence of the sets $\{\xi_{k,j,n} : k \in \mathbb{N}\}$, $j \in \mathbb{N}$, we obtain that $\mathsf{E}\,\zeta_{k_1,j_1,j_2,n}\zeta_{k_2,j_3,j_4,n}\zeta_{k_3,j_5,j_6,n}\zeta_{k_4,j_7,j_8,n} = 0$ if one of the sets $\{j_1, j_2\}$, $\{j_3, j_4\}$, $\{j_5, j_6\}$, $\{j_7, j_8\}$ is disjoint from the other three sets. Consequently

$$\sum_{j_1,\, j_2\,=\,1}^{k_1} \sum_{j_3,\, j_4\,=\,1}^{k_2} \sum_{j_5,\, j_6\,=\,1}^{k_3} \sum_{j_7,\, j_8\,=\,1}^{k_4} \mathsf{E}\,\zeta_{k_1,j_1,j_2,n}\zeta_{k_2,j_3,j_4,n}\zeta_{k_3,j_5,j_6,n}\zeta_{k_4,j_7,j_8,n} = O(n^6)$$

as $n \to \infty$. Using $\sum_{k=1}^n k^{-1} = O(\log n)$ we conclude $\mathsf{E}\,\zeta_n^4 = O\big(n^6(\log n)^4\big)$ as $n \to \infty$.

The case $\kappa = 1$ can be proved almost readily in the same way. $\square$

Corollary B.3 *Let $\xi_{k,j,n}$, $\zeta_{k,j,n}$, $k, j, n \in \mathbb{N}$, be random variables such that for each $n \in \mathbb{N}$ the sets $\{\xi_{k,j,n} : k \in \mathbb{N}\}$, $j \in \mathbb{N}$, are independent (i.e., the σ-algebras $\sigma(\xi_{k,j,n}, \zeta_{k,j,n} : k \in \mathbb{N})$, $j \in \mathbb{N}$, are independent), and $\sup_{k,\ell,n\in\mathbb{N}} \mathsf{E}(\xi_{k,\ell,n}^8 + \zeta_{k,\ell,n}^8) < \infty$. Let K_n, $n \in \mathbb{N}$, be positive integers such that $K_n = nK + o(n)$ as $n \to \infty$ with some $K > 0$. Then*

$$n^{-2} \sum_{k=1}^{K_n} k^{-1} \sum_{j=1}^{k} (\xi_{k,j,n}\zeta_{k,j,n} - \mathsf{E}\,\xi_{k,j,n}\zeta_{k,j,n}) \to 0 \qquad \text{a.s. as } \; n \to \infty.$$

Proof. Similar to the proof of Corollary B.1. $\square$

Lemma B.4 *Let $H_n : \mathbb{R}^{n+1} \to \mathbb{R}$, $n \in \mathbb{N}$, be measurable functions and $\{\xi_n\}_{n\in\mathbb{N}}$ be a sequence of random variables. Suppose that $H : \mathbb{R} \to \mathbb{R}$ is continuous and C is a compact subset of $\mathbb{R}$ such that*

$$\sup_{\alpha\in C} |H_n(\xi_1, \ldots, \xi_n, \alpha) - H(\alpha)| \to 0 \qquad \mathsf{P}\text{-a.s.}$$

Then, given random variables α_n, $n \in \mathbb{N}$, with $\mathsf{P}\,(\alpha_n \in C) = 1$, the sequence

$$\{H_n(\xi_1, \ldots, \xi_n, \alpha_n)\}_{n\in\mathbb{N}}$$

is stochastically bounded in the following sense:

$$\lim_{R\to\infty} \limsup_{n\to\infty} \mathsf{P}\,(|H_n(\xi_1, \ldots, \xi_n, \alpha_n)| > R) = 0. \tag{B.3}$$

Proof. Let $R > 0$. We have

$$\begin{aligned}
&\mathsf{P}(|H_n(\xi_1, \ldots, \xi_n, \alpha_n)| > R) \\
&\qquad \leqslant \mathsf{P}\,(|H_n(\xi_1, \ldots, \xi_n, \alpha_n) - H(\alpha_n)| + |H(\alpha_n)| > R) \\
&\qquad \leqslant \mathsf{P}\left(|H_n(\xi_1, \ldots, \xi_n, \alpha_n) - H(\alpha_n)| > \frac{R}{2}\right) + \mathsf{P}\left(|H(\alpha_n)| > \frac{R}{2}\right) \qquad \text{(B.4)} \\
&\qquad \leqslant \mathsf{P}\left(\sup_{\alpha\in C} |H_n(\xi_1, \ldots, \xi_n, \alpha) - H(\alpha)| > \frac{R}{2}\right) + \mathsf{P}\left(\sup_{\alpha\in C} |H(\alpha)| > \frac{R}{2}\right),
\end{aligned}$$

hence, taking in both sides of (B.4) first the '$\limsup$' as $n \to \infty$ and then the limit as $R \to \infty$, one gets the desired statement. $\square$

Let us remark that stochastic boundedness is not necessarily defined as in (B.3) in the literature. However, it suffices for our purpose. For this, we note that given a sequence of random variables, say $\{X_n\}_{n\in\mathbb{N}}$, with limit 0 in the sense of convergence in probability and given a stochastically bounded (in the sense of (B.3)) sequence of r.v.'s, say $\{Y_n\}_{n\in\mathbb{N}}$, one easily gets that $X_n \cdot Y_n$ converges to 0 in probability, as well.

Lemma B.5 *Let $\{X_n\}_{n\in\mathbb{Z}_+}$, $\{Y_n\}_{n\in\mathbb{Z}_+}$ and $\{Z_n\}_{n\in\mathbb{Z}_+}$ be sequences of random matrices of type $m \times 1$, $m \times m$ and $m \times 1$, respectively, such that $X_n = Y_n Z_n$, where $m \in \mathbb{N}$. Suppose that $X_n \xrightarrow{\mathcal{D}} X$, $Y_n \xrightarrow{\mathsf{P}} A$, where A is a non-degenerate matrix. Then $Z_n \xrightarrow{\mathcal{D}} A^{-1}X$.*

Proof. Define

$$Y_n^{\ominus} := \begin{cases} Y_n^{-1} & \text{if } Y_n \text{ invertible,} \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, we have $\mathsf{P}(Y_n^{\ominus} = Y_n^{-1}) \to 1$, as $n \to \infty$. Hence,

$$\mathsf{P}(||Y_n^{\ominus} - A^{-1}|| \geqslant \varepsilon) \leqslant \mathsf{P}(||Y_n^{-1} - A^{-1}|| \geqslant \varepsilon \text{ and } Y_n^{\ominus} = Y_n^{-1}) + \mathsf{P}(Y_n^{\ominus} \neq Y_n^{-1}) \to 0$$

as $n \to \infty$, that is $Y_n^{\ominus} \xrightarrow{\mathsf{P}} A^{-1}$. By Slutsky's Lemma, $Y_n^{\ominus} X_n \xrightarrow{\mathsf{P}} A^{-1}X$. Further, $\mathsf{P}(Z_n = Y_n^{\ominus} X_n) \to 1$, and hence $Z_n \xrightarrow{\mathsf{P}} A^{-1}X$. Thus, the proof of Lemma B.5 is completed, and so is the proof of Theorem 2.1. $\square$