Web Appendix to:
"Spillover Dynamics for Systemic Risk Measurement Using Spatial Financial Time Series Models"

Francisco Blasques[(a)], Siem Jan Koopman[(a,b)], Andre Lucas[(a)], Julia Schaumburg[(a)]

[(a)] VU University Amsterdam and Tinbergen Institute, The Netherlands

[(b)] CREATES, Aarhus University, Denmark

Appendix B Proofs

In this web appendix we establish the existence, strong consistency and asymptotic normality of the MLE of the static parameters $\boldsymbol{\theta}$ that define the stochastic properties of the spatial score model from Section 2. We first present the results in a more general setting than the spatial score model, thus extending the results in Blasques et al. (2014) to allow for the presence of exogenous regressors.

B.1 Stochastic properties of the filtered spatial dependence parameter

To establish the consistency and asymptotic normality of the MLE, we first study the stochastic properties of the filtered parameter f_t defined through equations (8), (11), and (12). The filtered f_ts directly determine the time-varying spatial parameter $\rho_t = h(f_t)$. Understanding the properties of the filtered parameters is key to understanding the stochastic properties of the likelihood function over the parameter space $\boldsymbol{\Theta}$.

We first introduce some additional notation. Let the T-period sequences $\{y_t(\omega)\}_{t=1}^T$ and $\{X_t(\omega)\}_{t=1}^T$ be subsets of the realized path of n and k-variate stochastic sequences $\boldsymbol{y}(\omega) := \{y_t(\omega)\}_{t\in\mathbb{Z}}$ and $\boldsymbol{X}(\omega) := \{X_t(\omega)\}_{t\in\mathbb{Z}}$, for some ω in the event space Ω. In particular,[1] we let

[1]The random sequences $\boldsymbol{y}$ and $\boldsymbol{X}$ are thus $\mathfrak{F}/\mathfrak{B}(\mathcal{Y}_\infty)$ and $\mathfrak{F}/\mathfrak{B}(\mathcal{X}_\infty)$-measurable mappings $\boldsymbol{y}:\Omega\rightarrow\mathcal{Y}_\infty\subseteq\mathbb{R}^n_\infty$ and $\boldsymbol{X}:\Omega\rightarrow\mathcal{X}_\infty\subseteq\mathbb{R}^k_\infty$ where $\mathbb{R}^n_\infty := \times_{t=-\infty}^{t=\infty}\mathbb{R}^n$ and $\mathbb{R}^k_\infty := \times_{t=-\infty}^{t=\infty}\mathbb{R}^k$ denote Cartesian products of infinite copies of $\mathbb{R}^n$ and $\mathbb{R}^k$ respectively, and $\mathcal{Y}_\infty = \times_{t=-\infty}^{t=\infty}\mathcal{Y}$ and $\mathcal{X}_\infty = \times_{t=-\infty}^{t=\infty}\mathcal{X}$ with $\mathfrak{B}(\mathcal{Y}_\infty)\equiv\mathfrak{B}(\mathbb{R}^n_\infty)\cap\mathcal{Y}_\infty$ and $\mathfrak{B}(\mathcal{X}_\infty)\equiv\mathfrak{B}(\mathbb{R}^k_\infty)\cap\mathcal{X}_\infty$; see (Billingsley, 1995, p.159). Here, $\mathfrak{B}(\mathbb{R}^n_\infty)$ and $\mathfrak{B}(\mathbb{R}^k_\infty)$ denote the Borel σ-algebras generated by the finite dimensional product cylinders of $\mathbb{R}^n_\infty$ and $\mathbb{R}^k_\infty$ respectively, $\mathfrak{F}$ denotes a σ-field defined on the event space Ω, and together with the probability measure P_0 on $\mathfrak{F}$, the triplet $(\Omega,\mathfrak{F},\mathrm{P}_0)$ denotes the common underlying complete probability space of interest.

$y_t(\omega) \in \mathcal{Y} \subseteq \mathbb{R}^n$and $X_t(\omega) \in \mathcal{X} \subseteq \mathbb{R}^k$ for all $(\omega, t) \in \Omega \times \mathbb{Z}$. For every $\omega \in \Omega$, the stochastic sequences $\boldsymbol{y}(\omega)$ and $\boldsymbol{X}(\omega)$ thus live on the spaces $(\mathcal{Y}_\infty, \mathfrak{B}(\mathcal{Y}_\infty), \mathrm{P}_0^y)$ and $(\mathcal{X}_\infty, \mathfrak{B}(\mathcal{X}_\infty), \mathrm{P}_0^X)$ where the probability measures P_0^y are P_0^X are defined over the elements of the Borel σ-algebras $\mathfrak{B}(\mathcal{Y}_\infty)$ and $\mathfrak{B}(\mathcal{X}_\infty)$. We write the filtered time-varying parameter as $\tilde{f}_t$ to distinguish it from the true time-varying parameter f_t. More precisely, we write the filtered time-varying parameter as $\{\tilde{f}_t(y^{1:t-1}, X^{1:t-1}; \boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$, which depends naturally on the initialization $\bar{f}_1 \in \mathcal{F} \subseteq \mathbb{R}$, the past data $y^{1:t-1} = \{y_s\}_{s=1}^{t-1}$ and $X^{1:t-1} = \{X_s\}_{s=1}^{t-1}$, and the parameter vector $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. For notational simplicity we often omit the dependence on the data and write $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ instead.

We can now rewrite the score update in (8) as

$$\tilde{f}_{t+1}(\boldsymbol{\theta}, \bar{f}_1) = \omega + A\, s\big(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1), y_t, X_t; \beta, \lambda\big) + B\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1) \quad \forall\, t \in \mathbb{N},$$

where $s(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1), y_t, X_t; \beta, \lambda)$ denotes the unit scaled score function. To shorten the notation, we define the random function

$$\begin{aligned} \phi_t\big(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1); \boldsymbol{\theta}\big) \;&:=\; \phi\big(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1), y_t, X_t; \boldsymbol{\theta}\big) \\ &:=\; \omega + A\, s(f_t(\boldsymbol{\theta}, \bar{f}_1), y_t, X_t; \beta, \lambda) + B f_t(\boldsymbol{\theta}, \bar{f}_1), \end{aligned}$$

as well as the supremum of its derivative,

$$\bar{\phi}_t'(\boldsymbol{\theta}) := \sup_{f\in\mathcal{F}} \Big| A\, \frac{\partial s(f, y_t, X_t; \beta, \lambda)}{\partial f} + B \Big|. \tag{B.1}$$

Note that $\bar{\phi}_t(\boldsymbol{\theta})$ is also a random variable due to its dependence on (y_t, X_t).

The following theorem states sufficient conditions for the stochastic sequence $\{\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\}_{t\in\mathbb{N}}$ initialized at $\bar{f}_1 \in \mathcal{F}$ to converge almost surely, uniformly in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, and exponentially fast to a limit stationary and ergodic (SE) sequence $\{\tilde{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ that has N_f bounded moments. We repeatedly make use of this notion of uniform exponentially fast almost sure convergence (e.a.s.), which means that there exists a $\gamma > 1$ such that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \gamma^t \Big| \tilde{f}_t\big(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}, \bar{f}_1\big) - \tilde{f}_t\big(y^{t-1}, X^{t-1}, \boldsymbol{\theta}\big) \Big| \overset{a.s.}{\to} 0 \quad \text{as} \quad t \to \infty;$$

see Straumann and Mikosch (2006). Note that the limit sequence starts in the infinite past and hence depends on the infinite past data $y^{t-1} := \{y_s\}_{s=-\infty}^{t-1}$ and $X^{t-1} := \{X_s\}_{s=-\infty}^{t-1}$, i.e., $\{\tilde{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}} \equiv \{\tilde{f}_t(y^{t-1}, X^{t-1}; \boldsymbol{\theta})\}_{t\in\mathbb{Z}}$. We thus establish the convergence of the sequence of random functions $\{\tilde{f}_t(\cdot,\bar{f}_1)\}_{t\in\mathbb{N}}$ defined on $\mathbf{\Theta}$ with random elements taking values in the Banach space $(\mathbb{C}(\mathbf{\Theta},\mathcal{F}),\|\cdot\|_{\mathbf{\Theta}})$ for every $t\in\mathbb{N}$, to an SE limit $\{\tilde{f}_t(\cdot)\}_{t\in\mathbb{Z}}$ with elements taking values in $(\mathbb{C}(\mathbf{\Theta}),\|\cdot\|_{\mathbf{\Theta}})$, where $\|\cdot\|_{\mathbf{\Theta}}$ denotes the supremum norm on $\mathbf{\Theta}$. We have the following result.

Theorem B.1. *Let $\mathcal{F}$ be convex, $\mathbf{\Theta}$ be compact, $\{y_t\}_{t\in\mathbb{Z}}$ and $\{X_t\}_{t\in\mathbb{Z}}$ be SE, $s\in\mathbb{C}(\mathcal{F}\times\mathcal{Y}\times\mathcal{X}\times\mathcal{B}\times\Lambda)$ and assume there exists a non-random $\bar{f}_1\in\mathcal{F}$ such that*

(i) $\mathbb{E}\log^+\sup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda}|s(\bar{f}_1,y_t,X_t;\beta,\lambda)|<\infty$*;*

(ii) $\mathbb{E}\log\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}}\bar{\phi}_1'(\boldsymbol{\theta})<0$.

Then $\{\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1)\}_{t\in\mathbb{N}}$ converges e.a.s. to the unique limit SE process $\{\tilde{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$.

If furthermore $\exists\ N_f\geq 1$ such that

(iii) $\mathbb{E}\sup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda}|s(\bar{f}_1,y_t,X_t;\beta,\lambda)|^{N_f}<\infty$*;*

and either

(iv) $\sup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda}|s(f,y,X;\beta,\lambda)-s(f',X,f;\beta,\lambda)|<|f-f'|\ \forall\,(f,f',y,X)\in\mathcal{F}\times\mathcal{F}\times\mathcal{Y}\times\mathcal{X}$*;*

or

(iv′) $\mathbb{E}\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}}\bar{\phi}_1'(\boldsymbol{\theta})^{N_f}<1$ *and* $\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1)\perp\bar{\phi}_t'(\boldsymbol{\theta})\ \forall\,(t,\bar{f}_1)\in\mathbb{N}\times\mathcal{F}$*, where $\perp$ denotes independence;*

then both $\{\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1)\}_{t\in\mathbb{N}}$ and the limit SE process $\{\tilde{f}_t(\boldsymbol{\theta})\}_{t\in\mathbb{Z}}$ have N_f bounded moments, i.e., $\sup_t\mathbb{E}\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}}|\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1)|^{N_f}<\infty$ and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}}|\tilde{f}_t(\boldsymbol{\theta})|^{N_f}<\infty$.

The first claim of Theorem 1 makes use of the conditions in Bougerol (1993a). Condition *(i)* requires the existence of an arbitrarily small moment for the score, and condition *(ii)* requires the spatial score update to be contracting on average. The uniqueness of the SE limit follows from Straumann and Mikosch (2006). The second claim of Theorem B.1 uses stricter moment conditions and contraction conditions to obtain bounded moments of higher order for the filtered sequence. This constitutes an extension of Proposition 1 in Blasques et al. (2014) to the spatial score setting with exogenous random variables X_t as well as vector and matrix arguments. Remark B.2 below highlights that in the special case where the score is uniformly bounded, then the filter has infinitely many bounded moments under simpler conditions.

REMARK B.2. *Let $|B|<1$. If $\bar{s}:=\sup_{(\beta,\lambda,f,y,X)\in\mathcal{B}\times\mathcal{F}\times\mathcal{Y}\times\mathcal{X}}|s(f,y,X;\beta,\lambda)|<\infty$, then $\sup_t\mathbb{E}\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}}|\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1)|^{N_f}<\infty$ and $\mathbb{E}\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}}|\tilde{f}_t(\boldsymbol{\theta})|^{N_f}<\infty$ hold for every $N_f\geq 1$.*

The proof of this statement follows immediately by noting that $|\tilde{f}_{t+1}|\leq\sum_{j=0}^{t-2}|B|^j(|\omega|+|A|\,\bar{s})+|B^{t-1}\bar{f}_1|<\infty$.

B.2 Asymptotic properties of the maximum likelihood estimator

The observation-driven structure of the time-varying spatial lag model allows us to perform maximum likelihood (ML) estimation in a straightforward way. Following equation (10), we define the ML estimator (MLE) of the static parameter vector $\boldsymbol{\theta}$ as an element of the $\arg\max$ set of the sample log likelihood function $\mathcal{L}_T(\boldsymbol{\theta}, \bar{f}_1)$,

$$\hat{\boldsymbol{\theta}}_T(\bar{f}_1) \in \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathcal{L}_T(\boldsymbol{\theta}, \bar{f}_1), \tag{B.2}$$

where

$$\begin{aligned}\mathcal{L}_T(\boldsymbol{\theta}, \bar{f}_1) &= \frac{1}{T}\sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}, \bar{f}_1) \\ &= \frac{1}{T}\sum_{t=1}^{T} \log p_e\Big(y_t - h\big((\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\big)Wy_t - X_t\beta\ ;\ \lambda\Big) - \log|Z\big(\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)\big)|.\end{aligned}$$

with $Z(f_t)$ defined below (11).

We can now use the stationarity, ergodicity, and moment results from Theorem B.1 to establish existence, consistency and asymptotic normality of the MLE. For existence, we make the following assumption.

ASSUMPTION B.3. $(\boldsymbol{\Theta}, \mathfrak{B}(\boldsymbol{\Theta}))$ *is a measurable space and* $\boldsymbol{\Theta}$ *is a compact set. Furthermore,* $h : \mathcal{F} \to \mathcal{F} \subseteq \mathbb{R}$ *and* $p_e : \mathbb{R}^n \times \Lambda \to \mathbb{R}$ *are continuously differentiable in their arguments.*

In Section 2, we have opted for the unit scaling of the score in our model. We can easily generalize all results below to the case of a non-constant scaling function S as long as we assume $S : \mathcal{F} \to \mathbb{R}$ is sufficiently smooth. Theorem B.4 below establishes the existence and measurability of the MLE.

Theorem B.4. (Existence) *Let Assumption B.3 hold. Then there exists a.s. an* $\mathfrak{F}/\mathfrak{B}(\boldsymbol{\Theta})$*-measurable map* $\hat{\boldsymbol{\theta}}_T(\bar{f}_1) : \Omega \to \boldsymbol{\Theta}$ *satisfying (B.2) for all* $T \in \mathbb{N}$ *and every initialization* $\bar{f}_1 \in \mathcal{F}$.

To obtain consistency of the MLE, we impose conditions that ensure that the likelihood function satisfies a uniform law of large numbers for SE processes. We first ensures that the filter $\tilde{f}(\boldsymbol{\theta}, \bar{f}_1)$ is SE and has N_f bounded moments by application of Theorem B.1.

ASSUMPTION B.5. $\exists\,(N_f, f) \in [1, \infty) \times \mathcal{F}$ *and a* $\boldsymbol{\Theta} \subset \mathbb{R}^{3+d_\lambda}$ *such that*

(i) $\sup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda} \mathbb{E}|s(f, y_t, X_t; \beta, \lambda)|^{N_f} < \infty$,

and either

(ii) $\sup_{(f,y,X,\beta,\lambda)\in\mathbb{R}\times\mathcal{Y}\times\mathcal{X}\times\mathcal{B}\times\Lambda} |B + A\,\partial s(f,y,X;\beta,\lambda)/\partial f| < 1$,

or

(ii′) $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \bar{\phi}'_{t,N_f}(\boldsymbol{\theta}) = \mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sup_f |B + A\,\partial s(f,y_t,X_t;\beta,\lambda)/\partial f| < 1$
and $\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta}, \bar{f}_1) \perp \bar{\phi}'_{t+1,N_f}(\boldsymbol{\theta})\ \forall\,(t,\bar{f}_1)\in\mathbb{N}\times\mathcal{F}$.

Next, we ensure a bounded expectation for the likelihood function. To do this, we use the notion of 'moment preserving map'; see Blasques et al. (2014)for a detailed description of the moment preserving properties of a wide catalogue of functions. This allows us to derive the appropriate number of bounded moments of the likelihood function from the moments of its arguments

DEFINITION B.6. (Moment Preserving Maps) *A function* $H : \mathbb{R}^{k_1}\times\boldsymbol{\Theta}\to\mathbb{R}^{k_2}$ *is said to be* n/m*-moment preserving, denoted as* $H\in\mathbb{M}_{\boldsymbol{\Theta}}(n,m)$*, if and only if* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathbf{x}_t(\boldsymbol{\theta})|^n < \infty$ *implies* $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |H(\mathbf{x}_t(\boldsymbol{\theta});\boldsymbol{\theta})|^m < \infty$.[2]

ASSUMPTION 1. $N_\ell = \min\{N_{\log p_e}, N_{\log|Z|}\} \geq 1$*, where* $\log|Z| \in \mathbb{M}_{\boldsymbol{\Theta}}(N_f, N_{\log|Z|})$ *and* $\log p_e \in \mathbb{M}_{\boldsymbol{\Theta}}\big(N, N_{\log p_e}\big)$*, with* $N = \min\big\{N_y, N_x\big\}$*, where* N_y *and* N_x *denote the moments of* y_t *and* X_t*, respectively.*

The moment N_ℓ in Assumption 1 corresponds to the number of moments of the likelihood function. Rather than assuming $N_\ell \geq 1$ as a high-level assumption, we follow Blasques et al. (2014) and define N_ℓ as a function of the score model constituents directly, thus obtaining a set of low-level conditions for strong consistency. The requirements imposed in Assumption 1 follow easily by application of a generalized Holder inequality to the likelihood expression below (B.2). Note that $N = \min\big\{N_y, N_x\big\}$ follows directly by the fact that the argument $(y_t - h(\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1)Wy_t - X_t\beta)$ of p_e is linear in both y_t and X_t, and $\sup_{f\in\mathcal{F}} |h(f)| \leq 1$. The current conditions extend those of Blasques et al. (2014) by accounting for the presence of exogenous variables X_t in the model.

Theorem B.7 now establishes the strong consistency of the MLE for the parameters of our time-varying spatial score model if the data are SE.

Theorem B.7. (Consistency) *Let* $\{y_t\}_{t\in\mathbb{Z}}$ *and* $\{X_t\}_{t\in\mathbb{Z}}$ *be SE sequences satisfying* $\mathbb{E}|y_t|^{N_y} < \infty$ *and* $\mathbb{E}|X_t|^{N_x} < \infty$ *for some* $N_y > 0$ *and* $N_x > 0$ *and let Assumptions B.3, B.5, and 1 hold. Furthermore, let* $\boldsymbol{\theta}_0\in\boldsymbol{\Theta}$ *be the unique maximizer of* $\mathcal{L}_\infty(\boldsymbol{\theta})$ *on the parameter space* $\boldsymbol{\Theta}$*. Then the MLE satisfies* $\hat{\boldsymbol{\theta}}_T(\bar{f}_1) \stackrel{a.s.}{\to} \boldsymbol{\theta}_0$ *as* $T\to\infty$ *for every* $\bar{f}_1\in\mathcal{F}$.

[2] The $(k_1\times 1)$-vector $\mathbf{x}_t$ satisfies $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathbf{x}_t(\boldsymbol{\theta})|^n < \infty$ if its elements $x_{i,t}(\boldsymbol{\theta})$ satisfy $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |x_{i,t}(\boldsymbol{\theta})|^n < \infty$, $i = 1, ..., k_1$. The same element-wise definition applies when $\mathbf{x}_t(\boldsymbol{\theta})$ is a matrix.

Remark B.8 below highlights that if the score s is uniformly bounded, we can change Assumption B.5 in line with Remark B.2.

REMARK B.8. *We can substitute Assumption B.5 in Theorem B.7 by*

(i) $\sup_{(\beta,\lambda,f,y,X)\in\mathcal{B}\times\Lambda\times\mathcal{F}\times\mathcal{Y}\times\mathcal{X}} |s(f,y,X;\beta,\lambda)| < \infty$;

(ii) $\mathbb{E}\log\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \bar{\phi}'_{1,1}(\boldsymbol{\theta}) < 0$ *and* $|B| < 1$.

Finally, we establish the asymptotic normality of the MLE. For this, we require the existence of a sufficient number of bounded moments for the likelihood function and its derivatives. For notational simplicity, we define the function $q_t := q(\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1), y_t, X_t; \beta,\lambda) := \log p_e(y_t - h(\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1)Wy_t - X_t\beta;\lambda)$, as well as the cross-derivatives

$$s^{(K_1,K_2,K_3)}(f,y,X;\beta,\lambda) := \frac{\partial^{K_1+K_2+K_3} s(f,y,X;\beta,\lambda)}{\partial f^{K_1}\partial\beta^{K_2}\partial\lambda^{K_3}}.$$

The (cross)-derivatives $q^{(K_1,K_2,K_3)}$ and $(\log|Z|)^{(K_1)}$ are defined similarly. Assumption B.9 now imposes sufficient moment conditions for the asymptotic normality of the MLE.

ASSUMPTION B.9. *(i)* $s^{(\boldsymbol{K})} \in \mathbb{M}_{\boldsymbol{\Theta}}(\boldsymbol{N}, N_s^{(\boldsymbol{K})})$, $q^{(\boldsymbol{K}')} \in \mathbb{M}_{\boldsymbol{\Theta}}(N, N_q^{(\boldsymbol{K}')})$, $\boldsymbol{N} := (N_f, N_y, N_x)$, *with* N *as defined in Assumption 1;*

(ii) $N_{\ell'} \geq 2$, $N_{\ell''} \geq 1$, $N_f^{(1)} > 0$, *and* $N_f^{(2)} > 0$, *with*

$$\begin{aligned}
N_{\ell'} &= \min\left\{ N_q^{(0,1,0)}\,,\; N_q^{(0,0,1)}\,,\; \frac{N_{\log|Z|}^{(1)} N_f^{(1)}}{N_{\log|Z|}^{(1)} + N_f^{(1)}}\,,\; \frac{N_q^{(1,0,0)} N_f^{(1)}}{N_q^{(1,0,0)} + N_f^{(1)}} \right\},\\
N_{\ell''} &= \min\left\{ N_q^{(0,2,0)}\,,\; N_q^{(0,0,2)}\,,\; N_q^{(0,1,1)}\,,\; \frac{N_q^{(1,1,0)} N_f^{(1)}}{N_q^{(1,1,0)} + N_f^{(1)}}\,,\; \frac{N_q^{(1,0,1)} N_f^{(1)}}{N_q^{(1,0,1)} + N_f^{(1)}}\,,\right.\\
&\qquad \left. \frac{N_q^{(2,0,0)} N_f^{(1)}}{2N_q^{(2,0,0)} + N_f^{(1)}}\,,\; \frac{N_q^{(1,0,0)} N_f^{(2)}}{N_q^{(1,0,0)} + N_f^{(2)}}\,,\; \frac{N_{\log|Z|}^{(1)} N_f^{(2)}}{N_{\log|Z|}^{(1)} + N_f^{(2)}}\,,\; \frac{N_{\log|Z|}^{(2)} N_f^{(1)}}{2N_{\log|Z|}^{(2)} + N_f^{(1)}} \right\},\\
N_f^{(1)} &= \min\left\{ N_f, N_s, N_s^{(0,1,0)}, N_s^{(0,0,1)} \right\},\\
N_f^{(2)} &= \min\left\{ N_f^{(1)}\,,\; N_s^{(0,1,0)}\,,\; N_s^{(0,0,1)}\,,\; N_s^{(0,2,0)}\,,\; N_s^{(0,0,2)}\,,\; N_s^{(0,1,1)}\,,\right.\\
&\qquad \left. \frac{N_s^{(1,0,0)} N_f^{(1)}}{N_s^{(1,0,0)} + N_f^{(1)}}\,,\; \frac{N_s^{(2,0,0)} N_f^{(1)}}{2N_s^{(2,0,0)} + N_f^{(1)}}\,,\; \frac{N_s^{(1,1,0)} N_f^{(1)}}{N_s^{(1,1,0)} + N_f^{(1)}}\,,\; \frac{N_s^{(1,0,1)} N_f^{(1)}}{N_s^{(1,0,1)} + N_f^{(1)}} \right\}.
\end{aligned}$$

Rather than assuming $N_{\ell'} \geq 2$ and $N_{\ell''} \geq 1$ directly as a high-level condition, we define $N_{\ell'}$ and $N_{\ell''}$ explicitly in terms of their lower-level constituents. The moment conditions in Assumption B.9 extend those of Blasques et al. (2014) by allowing for exogenous regressors. The expressions may seem complicated at first, but we show below that their verification is often straightforward; see also Blasques et al. (2014) for the verification of similar moment conditions in a wide range of observation-driven models.

The quantities $N_f^{(1)}$ and $N_f^{(2)}$ in Assumption B.9 correspond to the moments of the first and second derivatives of the filter $\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)$ with respect to the parameter $\boldsymbol{\theta}$. Similarly, $N_{\ell'}$ and $N_{\ell''}$ denote the moments of the first and second derivatives of the likelihood function, respectively.

Theorem B.10 now establishes the asymptotic normality of the MLE. Here, $\operatorname{int}(\boldsymbol{\Theta})$ denotes the interior of $\boldsymbol{\Theta}$.

Theorem B.10. (Asymptotic Normality) *Let $\{y_t\}_{t\in\mathbb{Z}}$ and $\{X_t\}_{t\in\mathbb{Z}}$ be SE sequences that satisfy $\mathbb{E}|y_t|^{N_y} < \infty$ and $\mathbb{E}|X_t|^{N_x} < \infty$ for some $N_y > 0$ and $N_x > 0$ and let Assumptions B.3–B.9 hold. Furthermore, let $\boldsymbol{\theta}_0 \in \operatorname{int}(\boldsymbol{\Theta})$ be the unique maximizer of $\mathcal{L}_\infty(\boldsymbol{\theta})$ on $\boldsymbol{\Theta}$. Then,*

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_T(\bar{f}_1) - \boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}\big(0, \mathcal{I}^{-1}(\boldsymbol{\theta}_0)\mathcal{J}(\boldsymbol{\theta}_0)\mathcal{I}^{-1}(\boldsymbol{\theta}_0)\big) \ \ as \ \ T \to \infty,$$

where $\mathcal{J}(\boldsymbol{\theta}_0) := \mathbb{E}\tilde{\ell}'_t(\boldsymbol{\theta}_0)\tilde{\ell}'_t(\boldsymbol{\theta}_0)^\top$ is the expected outer product of gradients and $\mathcal{I}(\boldsymbol{\theta}_0) := \mathbb{E}\tilde{\ell}''_t(\boldsymbol{\theta}_0)$ is the Fisher information matrix.

Theorem 1 now follows directly from the previous (more general) theorems as a special case.

B.3 Proofs of main theorems

The lines of proof adopted here closely follow the original lines of proof in Blasques et al. (2014), extended to the case of exogenous variables. For ease of reference and to make the paper self-contained, we repeat the arguments in full.

Proof of Theorem B.1: Define the norms $\|\cdot\|^{\boldsymbol{\Theta}} := \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\cdot|$ and $\|\cdot\|^{\boldsymbol{\Theta}}_{N_f} := \mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\cdot\|^{N_f}$. Following Straumann and Mikosch (2006, Proposition 3.12), we have

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}, \bar{f}_1) - \tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta})| \xrightarrow{e.a.s.} 0.$$

This follows directly from Bougerol (1993b, Theorem 3.1) in the context of the random sequence $\{\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \cdot, \bar{f}_1^{\boldsymbol{\Theta}})\}_{t\in\mathbb{N}}$ with elements $\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \cdot, \bar{f}_1^{\boldsymbol{\Theta}})$ taking values in the separable Banach space $\mathcal{F}_{\boldsymbol{\Theta}} \subseteq (\mathbb{C}(\boldsymbol{\Theta}, \mathcal{F}), \|\cdot\|_{\boldsymbol{\Theta}})$, with initialization $\bar{f}_1^{\boldsymbol{\Theta}}$ in $\mathbb{C}(\boldsymbol{\Theta}, \mathcal{F})$, where $\bar{f}_1^{\boldsymbol{\Theta}}(\boldsymbol{\theta}) =$

$\bar{f}_1 \; \forall \, \boldsymbol{\theta} \in \boldsymbol{\Theta}$, and[3]

$$\begin{aligned} \tilde{f}_t(y^{1:t}, X^{1:t}, \cdot, \bar{f}_1^{\boldsymbol{\Theta}}) &= \phi_t\big(\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \cdot, \bar{f}_1^{\boldsymbol{\Theta}})\big), \\ &:= \phi\big(\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \cdot, \bar{f}_1^{\boldsymbol{\Theta}}) \, , \, y_t, X_t; \cdot \big) \; \forall \, t \in \mathbb{N}, \end{aligned}$$

where $\{\phi_t\}_{t\in\mathbb{Z}}$ is a stationary and ergodic (SE) sequence of stochastic recurrence equations $\phi_t : \Xi \times \mathbb{C}(\boldsymbol{\Theta}, \mathcal{F}) \to \mathbb{C}(\boldsymbol{\Theta}, \mathcal{F}) \; \forall \, t \in \Xi$ as in Straumann and Mikosch (2006, Proposition 3.12). Note that with a slight abuse of notation we use ϕ both to denote the functional $\phi : \mathbb{C}(\boldsymbol{\Theta}, \mathcal{F}) \times \mathcal{Y} \times \mathcal{X} \to \mathbb{C}(\boldsymbol{\Theta}, \mathcal{F})$ as well as the function $\phi : \mathcal{F} \times \mathcal{Y} \times \mathcal{X} \times \boldsymbol{\Theta} \to \mathcal{F}$. Continuity of ϕ follows from $s \in \mathbb{C}(\mathcal{F} \times \mathcal{Y} \times \mathcal{X} \times \mathcal{B} \times \Lambda)$, where $\mathcal{B}$ is the domain of the regression parameters β.

The assumption that $\{y_t\}_{t\in\mathbb{Z}}$ and $\{X_t\}_{t\in\mathbb{Z}}$ are SE and the continuity of ϕ together imply that $\{\phi_t\}_{t\in\mathbb{Z}}$ is SE by Krengel (1985, Proposition 4.3). Condition C1 in Bougerol (1993b, Theorem 3.1) follows from $\mathbb{E} \log^+ \|s(f^{\boldsymbol{\Theta}}, y_t, X_t; \cdot, \cdot)\|^{\boldsymbol{\Theta}} < \infty$ since, by norm sub-additivity and positive homogeneity, for any $f^{\boldsymbol{\Theta}} \in \mathbb{C}(\boldsymbol{\Theta}, \mathcal{F})$,

$$\begin{aligned} \mathbb{E} \log^+ \big\|\phi_t(f^{\boldsymbol{\Theta}})\big\|^{\boldsymbol{\Theta}} &= \mathbb{E} \log^+ \|\omega + As(f^{\boldsymbol{\Theta}}, y_t, X_t; \cdot, \cdot) + Bf^{\boldsymbol{\Theta}}\|^{\boldsymbol{\Theta}} \\ &\leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \Big(\log^+ |\omega| + \log^+ |A| + \log^+ |B| \Big) + \mathbb{E} \log^+ \|s(f^{\boldsymbol{\Theta}}, y_t, X_t; \cdot, \cdot)\|^{\boldsymbol{\Theta}} + \log^+ \|f^{\boldsymbol{\Theta}}\|^{\boldsymbol{\Theta}} < \infty, \end{aligned}$$

because $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\omega| < \infty$, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |A| < \infty$, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |B| < \infty$, and $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|f^{\boldsymbol{\Theta}}\|^{\boldsymbol{\Theta}} < \infty$ hold by compactness of $\boldsymbol{\Theta}$ and continuity of $f^{\boldsymbol{\Theta}}$, and $\mathbb{E} \log^+ \|s(f^{\boldsymbol{\Theta}}, y_t, X_t; \cdot, \cdot)\|^{\boldsymbol{\Theta}} < \infty$ holds by assumption. This implies that $f_{\boldsymbol{\Theta}} \in \mathbb{C}(\boldsymbol{\Theta}, \mathcal{F})$ satisfies

$$\begin{aligned} \mathbb{E} \log^+ \|\phi_0(f^{\boldsymbol{\Theta}}) - f^{\boldsymbol{\Theta}}\|_{\boldsymbol{\Theta}} \;\leq\; \mathbb{E}\|\phi_0(f^{\boldsymbol{\Theta}}) - f^{\boldsymbol{\Theta}}\|^{\boldsymbol{\Theta}} \leq\; \mathbb{E}\|\phi(f^{\boldsymbol{\Theta}}, y_t, X_t; \cdot)\|^{\boldsymbol{\Theta}} + \|f^{\boldsymbol{\Theta}}\|^{\boldsymbol{\Theta}} \\ = \mathbb{E} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\phi(f^{\boldsymbol{\Theta}}(\boldsymbol{\theta}), y_t, X_t, \boldsymbol{\theta})| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |f^{\boldsymbol{\Theta}}(\boldsymbol{\theta})| < \infty. \end{aligned}$$

By a similar argument $\mathbb{E} \log^+ \sup_{(\beta,\lambda)\in\mathcal{B}\times\Lambda} |s(\bar{f}_1, y_t, X_t; \beta, \lambda)| < \infty$ implies $\mathbb{E} \log^+ \|\phi_0(f^{\boldsymbol{\Theta}}) - f^{\boldsymbol{\Theta}}\|_{\boldsymbol{\Theta}}^{N_f} < \infty$.

For any pair $(f^{\boldsymbol{\Theta}}, f'^{\boldsymbol{\Theta}}) \in \mathbb{C}(\boldsymbol{\Theta}) \times \mathbb{C}(\boldsymbol{\Theta})$, define

$$\rho_t = \rho(\phi_t) = \sup_{(f^{\boldsymbol{\Theta}}, f'^{\boldsymbol{\Theta}}) \in \mathcal{F}_{\boldsymbol{\Theta}} \times \mathcal{F}_{\boldsymbol{\Theta}}} \frac{\|\phi_t(f^{\boldsymbol{\Theta}}) - \phi_t(f'^{\boldsymbol{\Theta}})\|_{\boldsymbol{\Theta}}}{\|f^{\boldsymbol{\Theta}} - f'^{\boldsymbol{\Theta}}\|_{\boldsymbol{\Theta}}}.$$

Condition C2 in Bougerol (1993b, Theorem 3.1) holds if $\mathbb{E} \log \rho_t < 0$. This is ensured by

[3] That $(\mathbb{C}(\boldsymbol{\Theta}, \mathcal{F}), \|\cdot\|_{\boldsymbol{\Theta}})$ is a separable Banach space under compact $\boldsymbol{\Theta}$ follows from application of the Arzeláscoli theorem to obtain completeness and the Stone-Weierstrass theorem for separability.

$\mathbb{E}\log\|\bar{\phi}_t'\|^{\boldsymbol{\Theta}} < 0$, with $\bar{\phi}_t'(\boldsymbol{\theta})$ as defined in (B.1). It becomes apparent by noting that

$$\begin{aligned}
\mathbb{E}\log\rho(\phi_t) &= \mathbb{E}\log \sup_{\|f^{\boldsymbol{\Theta}}-f'^{\boldsymbol{\Theta}}\|>0} \frac{\|\phi_t(f^{\boldsymbol{\Theta}})-\phi_t(f'^{\boldsymbol{\Theta}})\|^{\boldsymbol{\Theta}}}{\|f^{\boldsymbol{\Theta}}-f'^{\boldsymbol{\Theta}}\|^{\boldsymbol{\Theta}}} \\
&= \mathbb{E}\log \sup_{\|f^{\boldsymbol{\Theta}}-f'^{\boldsymbol{\Theta}}\|>0} \frac{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\phi(f(\boldsymbol{\theta}),y_t,X_t,\boldsymbol{\theta})-\phi(f'(\boldsymbol{\theta}),y_t,X_t,\boldsymbol{\theta})|}{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|f(\boldsymbol{\theta})-f'(\boldsymbol{\theta})|} \\
&\leq \mathbb{E}\log \sup_{\|f^{\boldsymbol{\Theta}}-f'^{\boldsymbol{\Theta}}\|>0} \frac{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\bar{\phi}_t'(\boldsymbol{\theta})\ \ \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|f(\boldsymbol{\theta})-f'(\boldsymbol{\theta})|}{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|f(\boldsymbol{\theta})-f'(\boldsymbol{\theta})|} \\
&= \mathbb{E}\log\|\bar{\phi}_t'\|^{\boldsymbol{\Theta}} < 0.
\end{aligned}$$

Also note that for the t period composition of the stochastic recurrence equation, we have $\mathbb{E}\log\rho(\phi_t\circ\ldots\circ\phi_1) \leq \mathbb{E}\log\prod_{j=1}^t\rho(\phi_j) \leq \sum_{j=1}^t \log\|\bar{\phi}_j'\|^{\boldsymbol{\Theta}} < 0$, where $\circ$ denotes composition. As a result, $\{\tilde{f}_t(\cdot,\bar{f}_1)\}_{t\in\mathbb{N}}$ converges e.a.s. to an SE solution $\{\tilde{f}_t(\cdot)\}_{t\in\mathbb{Z}}$ in $\|\cdot\|^{\boldsymbol{\Theta}}$-norm. Uniqueness and e.a.s. convergence is obtained in Straumann and Mikosch (2006, Theorem 2.8).

Finally, we show that $\sup_t\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\tilde{f}_t(y^{1:t-1},X^{1:t-1},\boldsymbol{\theta},\bar{f}_1)|^{N_f} < \infty$ and also $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\tilde{f}_t(y^{1:t-1},X^{1:t-1},\boldsymbol{\theta})|^{N_f} < \infty$. We have $\sup_t\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\tilde{f}_t(y^{1:t-1},X^{1:t-1},\boldsymbol{\theta},\bar{f}_1)|^{N_f} < \infty$ if and only if $\sup_t(\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\tilde{f}_t(y^{1:t-1},X^{1:t-1},\boldsymbol{\theta},\bar{f}_1)|^{N_f})^{1/N_f} = \sup_t\|\tilde{f}_t(\cdot,\bar{f}_1)\|_{N_f}^{\boldsymbol{\Theta}} < \infty$. Furthermore, for any $f^{\boldsymbol{\Theta}}\in\mathbb{C}(\boldsymbol{\Theta},\mathcal{F})$, having $\|\tilde{f}_t(\cdot,\bar{f}_1)-f^{\boldsymbol{\Theta}}\|_{N_f}^{\boldsymbol{\Theta}} < \infty$ implies $\|\tilde{f}_t(\cdot,\bar{f}_1^{\boldsymbol{\Theta}})\|_{N_f}^{\boldsymbol{\Theta}} < \infty$ since continuity on the compact $\boldsymbol{\Theta}$ implies $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|f(\boldsymbol{\theta})| < \infty$. For $f^{\boldsymbol{\Theta}}\in\mathbb{C}(\boldsymbol{\Theta},\mathcal{F})$, we define $f_*^{\boldsymbol{\Theta}}$, y_*, and X_* such that $f^{\boldsymbol{\Theta}} = \phi(y,X,f_*^{\boldsymbol{\Theta}},\cdot)\in\mathbb{C}(\boldsymbol{\Theta},\mathcal{F})$. Using similar arguments as above, we can show that under the current assumptions $\exists\, f^{\boldsymbol{\Theta}}\in\mathbb{C}(\boldsymbol{\Theta},\mathcal{F})$ satisfying $\|\phi(f^{\boldsymbol{\Theta}},y_t,X_t;\cdot)\|_{N_f}^{\boldsymbol{\Theta}} \leq \bar{\bar{\phi}} < \infty$ and $\|\bar{f}_1^{\boldsymbol{\Theta}}-f^{\boldsymbol{\Theta}}\|_{N_f}^{\boldsymbol{\Theta}} = \|\bar{f}_1^{\boldsymbol{\Theta}}-\phi(f_*^{\boldsymbol{\Theta}},y_*,X_*;\cdot)\|_{N_f}^{\boldsymbol{\Theta}} < \infty$. From this, we obtain

$$\begin{aligned}
\sup_t&\|\tilde{f}_{t+1}(\cdot,\bar{f}_1^{\boldsymbol{\Theta}})-f^{\boldsymbol{\Theta}}\|_{N_f}^{\boldsymbol{\Theta}} = \sup_t\|\phi(\tilde{f}_t(\cdot,\bar{f}_1^{\boldsymbol{\Theta}}),y_t,X_t;\cdot)-\phi(f_*^{\boldsymbol{\Theta}},y_*,X_*;\cdot)\|_{N_f}^{\boldsymbol{\Theta}} \\
&\leq \sup_t\|\phi(\tilde{f}_t(\cdot,\bar{f}_1^{\boldsymbol{\Theta}}),y_t,X_t;\cdot)-\phi(f_*^{\boldsymbol{\Theta}},y_t,X_t;\cdot)\|_{N_f}^{\boldsymbol{\Theta}}+ \\
&\qquad \sup_t\|\phi(f_*^{\boldsymbol{\Theta}},y_t,X_t;\cdot)\|_{N_f}^{\boldsymbol{\Theta}} + \sup_t\|\phi(f_*^{\boldsymbol{\Theta}},y_*,X_*;\cdot)\|_{N_f}^{\boldsymbol{\Theta}} \\
&\leq \sup_t\left(\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1)-f_*^{\boldsymbol{\Theta}}|^{N_f}\times\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\frac{|\phi(\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1^{\boldsymbol{\Theta}}),y_t,X_t;\boldsymbol{\theta})-\phi(f_*^{\boldsymbol{\Theta}}(\boldsymbol{\theta}),y_t,X_t;\boldsymbol{\theta})|^{N_f}}{|\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1^{\boldsymbol{\Theta}})-f_*^{\boldsymbol{\Theta}}(\boldsymbol{\theta})|^{N_f}}\right)^{1/N_f} \\
&\qquad + \sup_t\|\phi(f_*^{\boldsymbol{\Theta}},y_t,X_t;\cdot)\|_{N_f}^{\boldsymbol{\Theta}} + \|f^{\boldsymbol{\Theta}}\|_{N_f}^{\boldsymbol{\Theta}} \\
&\leq \sup_t\left(\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\tilde{f}_t(\boldsymbol{\theta},\bar{f}_1)-f_*^{\boldsymbol{\Theta}}|^{N_f}\times\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\bar{\phi}_t'(\boldsymbol{\theta})^{N_f}\right)^{1/N_f} + \sup_t\|\phi(f_*^{\boldsymbol{\Theta}},y_t,X_t;\cdot)\|_{N_f}^{\boldsymbol{\Theta}} + \|f^{\boldsymbol{\Theta}}\|_{N_f}^{\boldsymbol{\Theta}}.
\end{aligned}$$

Using the orthogonality condition in *(iv′)*, we can write the expectation of the product as the product of the expectations and continue

$$
\begin{aligned}
&\leq \sup_t \|\tilde{f}_t(\,\cdot\,,\bar{f}_1^{\boldsymbol{\Theta}}) - f_*^{\boldsymbol{\Theta}}\|_{N_f}^{\boldsymbol{\Theta}} \cdot \|\bar{\phi}_t'\|_{N_f}^{\boldsymbol{\Theta}} + \sup_t \|\phi(f_*^{\boldsymbol{\Theta}}, y_t, X_t;\cdot)\|_{N_f}^{\boldsymbol{\Theta}} + \|f^{\boldsymbol{\Theta}}\|_{N_f}^{\boldsymbol{\Theta}} \\
&\leq \|\bar{\phi}_t'\|_{N_f}^{\boldsymbol{\Theta}} \times \Big(\sup_t \|\tilde{f}_t(\,\cdot\,,\bar{f}_1^{\boldsymbol{\Theta}}) - f^{\boldsymbol{\Theta}}\|_{N_f}^{\boldsymbol{\Theta}}\Big) + \bar{\bar{\phi}} + \bar{f},
\end{aligned}
$$

with $\bar{c} = \|\bar{\phi}_t'\|_{N_f}^{\boldsymbol{\Theta}} < 1$ by condition *(iv′)*, $\bar{\bar{\phi}} < \infty$, and $\bar{f} = \|f^{\boldsymbol{\Theta}}\| + \bar{c} \cdot \|f^{\boldsymbol{\Theta}} - f_*^{\boldsymbol{\Theta}}\|_{N_f}^{\boldsymbol{\Theta}} < \infty$. As a result we have the recursion $\sup_t \|\tilde{f}_{t+1}(\cdot, \bar{f}_1^{\boldsymbol{\Theta}}) - f^{\boldsymbol{\Theta}}\|_{N_f}^{\boldsymbol{\Theta}} \leq \bar{c} \cdot \sup_t \|\tilde{f}_t(\cdot, \bar{f}_1^{\boldsymbol{\Theta}}) - f^{\boldsymbol{\Theta}}\|_{N_f}^{\boldsymbol{\Theta}} + \bar{\bar{\phi}} + \bar{f}$. Hence,

$$
\sup_t \|\tilde{f}_t(\cdot, \bar{f}_1^{\boldsymbol{\Theta}}) - f^{\boldsymbol{\Theta}}\|_{N_f}^{\boldsymbol{\Theta}} \leq \sum_{j=0}^{t-2} (\bar{c})^j (\bar{f} + \bar{\bar{\phi}}) + \bar{c}^{t-1} \sup_t \|\bar{f}_1^{\boldsymbol{\Theta}} - f^{\boldsymbol{\Theta}}\|_{N_f}^{\boldsymbol{\Theta}} \leq \frac{\bar{f} + \bar{\bar{\phi}}}{1 - \bar{c}} + \|\bar{f}_1^{\boldsymbol{\Theta}} - f^{\boldsymbol{\Theta}}\|_{N_f}^{\boldsymbol{\Theta}} < \infty.
$$

The same result holds using the uniform contraction in *(iv)* by taking a further supremum in y_t and X_t instead of the orthogonality condition. □

Proof of Theorem B.4: Assumption B.3 implies that $\mathcal{L}_T(\boldsymbol{\theta}, \bar{f}_1) = (1/T)\sum_{t=1}^T \ell_t(\boldsymbol{\theta}, \bar{f}_1)$ is a.s. continuous (a.s.c.) in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ through continuity (c.) of each

$$
\begin{aligned}
\ell_t(\boldsymbol{\theta}, \bar{f}_1) &= \ell(y_t, X_t, \tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \bar{f}_1, \boldsymbol{\theta}), \boldsymbol{\theta}) \\
&= \log p_e(Z_t(f_t)^{-1} y_t - X_t\beta; \lambda) - \log |Z_t(f_t)|
\end{aligned}
$$

ensured in turn by the differentiability of S, p_e and h and the implied a.s.c. of

$$
\begin{aligned}
\nabla(\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \bar{f}_1, \boldsymbol{\theta}), y_t, X_t; \beta, \lambda) = &\frac{\partial \log p_e(Z_t(\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \bar{f}_1, \boldsymbol{\theta}))^{-1} y_t - X_t\beta; \lambda)}{\partial f} \\
&- \frac{\partial \log |Z_t(\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \bar{f}_1, \boldsymbol{\theta}))|}{\partial f}
\end{aligned}
$$

in $(\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \bar{f}_1, \boldsymbol{\theta}); \lambda)$ and the resulting c. of $\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \bar{f}_1, \boldsymbol{\theta})$ in $\boldsymbol{\theta}$ as a composition of t c. maps. Together with the compactness of $\boldsymbol{\Theta}$ this implies by Weierstrass' theorem that the $\arg\max$ set is non-empty a.s. and hence that $\hat{\boldsymbol{\theta}}_T$ exists a.s. $\forall T \in \mathbb{N}$. Assumption B.3 implies also by a similar argument that

$$
\mathcal{L}_T(\boldsymbol{\theta}, \bar{f}_1) = \mathcal{L}_T\big(\tilde{f}^{1:T}(y^{1:t-1}, X^{1:t-1}, \bar{f}_1, \boldsymbol{\theta}); y^{1:T}, X^{1:T}, \boldsymbol{\theta}\big)
$$

is continuous in $(y^{1:T}, X^{1:T})$ $\forall\, \boldsymbol{\theta} \in \boldsymbol{\Theta}$ and hence measurable w.r.t. the product Borel σ-algebra

$\mathfrak{B}(\mathcal{Y}) \otimes \mathfrak{B}(\mathcal{X})$ that are, in turn, measurable maps w.r.t. $\mathcal{F}$ by Proposition 4.1.7 in Dudley (2002).[4] The measurability of $\hat{\boldsymbol{\theta}}_T$ follows from Foland (2009, p.24) and White (1994, Theorem 2.11) or Gallant and White (1988, Lemma 2.1, Theorem 2.2).[5] □

Proof of Theorem B.7: We obtain $\hat{\boldsymbol{\theta}}_T(\bar{f}_1) \xrightarrow{a.s.} \boldsymbol{\theta}_0$ from the uniform convergence of the criterion function

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathcal{L}_T(\boldsymbol{\theta}, \bar{f}_1) - \mathcal{L}_\infty(\boldsymbol{\theta})| \xrightarrow{a.s.} 0 \;\forall\, \bar{f}_1 \in \mathcal{F} \;\text{ as }\; T \to \infty, \tag{B.3}$$

and the identifiable uniqueness of the maximizer $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ introduced in White (1994),

$$\sup_{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|>\epsilon} \mathcal{L}_\infty(\boldsymbol{\theta}) < \mathcal{L}_\infty(\boldsymbol{\theta}_0) \;\forall\, \epsilon > 0; \tag{B.4}$$

see for example White (1994, Theorem 3.4) or Theorem 3.3 in Gallant and White (1988) for further details.

The uniform convergence is obtained by norm sub-additivity,[6]

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathcal{L}_T(\boldsymbol{\theta}, \bar{f}_1) - \mathcal{L}_\infty(\boldsymbol{\theta})| \;\le\; \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathcal{L}_T(\boldsymbol{\theta}, \bar{f}_1) - \mathcal{L}_T(\boldsymbol{\theta})| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathcal{L}_T(\boldsymbol{\theta}) - \mathcal{L}_\infty(\boldsymbol{\theta})|,$$

and then showing that the initialization effect vanishes asymptotically,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathcal{L}_T(\boldsymbol{\theta}, \bar{f}_1) - \mathcal{L}_T(\boldsymbol{\theta})| \xrightarrow{a.s.} 0 \;\text{ as }\; T \to \infty, \tag{B.5}$$

and for the second term applying the ergodic theorem for separable Banach spaces in Ranga Rao (1962), as in Straumann and Mikosch (2006, Theorem 2.7), to the sequence $\{\mathcal{L}_T(\cdot)\}$ with elements taking values in $\mathbb{C}(\boldsymbol{\Theta}, \mathbb{R})$ so that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\mathcal{L}_T(\boldsymbol{\theta}) - \mathcal{L}_\infty(\boldsymbol{\theta})| \xrightarrow{a.s.} 0 \quad \text{where} \quad \mathcal{L}_\infty(\boldsymbol{\theta}) = \mathbb{E}\ell_t(\boldsymbol{\theta}) \;\forall\, \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

The criterion $\mathcal{L}_T(\boldsymbol{\theta}, \bar{f}_1)$ satisfies (B.5) if

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} |\ell_t(\boldsymbol{\theta}, \bar{f}_1) - \ell_t(\boldsymbol{\theta})| \xrightarrow{a.s.} 0 \quad \text{as} \quad t \to \infty.$$

[4]Dudley's proposition states that the Borel σ-algebra $\mathfrak{B}(\mathbb{A} \times \mathbb{B})$ generated by the Tychonoff's product topology $\mathcal{T}_{\mathbb{A}\times\mathbb{B}}$ on the space $\mathbb{A} \times \mathbb{B}$ includes the product σ-algebra $\mathfrak{B}(\mathbb{A}) \otimes \mathbb{B}(\mathbb{B})$.

[5]The reference of Foland (2009) is used here to establish that a map into a product space is measurable if and only if its projections are measurable.

[6]$\mathcal{L}_T(\boldsymbol{\theta})$ denotes $\mathcal{L}_T(\boldsymbol{\theta}, \bar{f}_1)$ with $\tilde{f}(\boldsymbol{\theta}, \bar{f}_1)$ replaced by its limit $\tilde{f}(\boldsymbol{\theta})$.

The continuity of p_e ensures that $\ell_t(\cdot, \bar{f}_1) = \ell(\tilde{f}_t(y^t, X^t, \cdot, \bar{f}_1), y_t, X_t, \cdot)$ is continuous in $(\tilde{f}_t(y^t, X^t, \cdot, \bar{f}_1), y_t, X_t)$. Since all the assumptions of Theorem 1 are satisfied we know that there exists a unique SE sequence $\{\tilde{f}_t(y^t, X^t, \cdot)\}_{t\in\mathbb{Z}}$ with elements taking values in $\mathbb{C}(\mathbf{\Theta}, \mathcal{F})$ such that

$$\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} \left|(\tilde{f}_t(y^{t-1}, X^{t-1}, \bar{f}_1, \boldsymbol{\theta}), y_t, X_t) - (\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta}), y_t, X_t)\right| \overset{a.s.}{\to} 0,$$

and

$$\sup_t \mathbb{E} \sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} |\tilde{f}_t(y^{t-1}, X^{t-1}, \bar{f}_1, \boldsymbol{\theta})^{N_f}| < \infty \quad \text{and} \quad \mathbb{E} \sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} |\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta})|^{N_f} < \infty,$$

with $N_f \geq 1$. Hence, (B.5) follows by application of a continuous mapping theorem for $\ell : \mathbb{C}(\mathbf{\Theta}, \mathcal{F}) \to \mathbb{C}(\mathbf{\Theta}, \mathcal{F})$.

The ULLN $\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} |\mathcal{L}_T(\boldsymbol{\theta}) - \mathbb{E}\ell_t(\boldsymbol{\theta})| \overset{a.s.}{\to} 0$ as $T \to \infty$ follows, under a moment bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} |\ell_t(\boldsymbol{\theta})| < \infty$, by the SE nature of $\{\ell_t\}_{t\in\mathbb{Z}}$ which is implied by continuity of ℓ on the SE sequence $\{(y_t, X_t, \tilde{f}_t(y^{t-1}, X^{t-1}, \cdot))\}_{t\in\mathbb{Z}}$ and Proposition 4.3 in Krengel (1985). The moment bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} |\ell_t(\boldsymbol{\theta})| < \infty$ can be established as follows. First note that

$$\begin{aligned}
\mathbb{E} \sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} |\ell_t(\boldsymbol{\theta})| &= \sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} \mathbb{E}| \log p_e(y_t - h(\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta}))Wy_{t-1} - X_t\beta) \\
&\qquad - \log \det Z(\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta}))| \\
&\leq \sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} \mathbb{E}| \log p_e(y_t - h(\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta}))Wy_{t-1} - X_t\beta)| \\
&\qquad - \sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} \mathbb{E}| \log \det Z(\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta}))| < \infty,
\end{aligned}$$

then the bounded first moment for the likelihood is implied by having

$$\mathbb{E}|y_t|^{N_y} < \infty \quad , \quad \mathbb{E}|X_t|^{N_X} < \infty, \quad \text{and} \quad \sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} \mathbb{E}|\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta})|^{N_f} < \infty.$$

since then

$$\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} \mathbb{E}| \log \det Z(\tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta}))| < \infty,$$

$$\sup_{\boldsymbol{\theta}\in\mathbf{\Theta}} \mathbb{E}| \log p_e(y_t - h(\tilde{f}_t(\boldsymbol{\theta}))Wy_{t-1} - X_t\beta)| < \infty,$$

because of the moment preserving properties of $\log |Z|$ and $\log p_e$ with with $N_{\log |Z|} \geq 1$ and $N_{\log p_e} \geq 1$ by assumption.

Finally, the identifiable uniqueness (see e.g. White (1994)) of $\boldsymbol{\theta}_0 \in \mathbf{\Theta}$ in (B.4) follows from the assumed uniqueness, the compactness of $\mathbf{\Theta}$, and the continuity of the limit $\mathbb{E}\ell_t(\boldsymbol{\theta})$ in $\boldsymbol{\theta} \in \mathbf{\Theta}$

which is implied by the continuity of $\mathcal{L}_T$ in $\boldsymbol{\theta} \in \boldsymbol{\Theta}\ \forall\ T \in \mathbb{N}$ and the uniform convergence in (B.3). $\square$

Proof of Theorem B.10: As the likelihood and its derivatives depend on the derivatives of $\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)$ with respect to $\boldsymbol{\theta}$, we introduce the notation $\boldsymbol{f}_t^{(0:m)}$ as the vector containing $\tilde{f}_t(\boldsymbol{\theta}, \bar{f}_1)$ and its derivatives up to order m, with initial condition $\bar{\boldsymbol{f}}^{(0:m)}$. We obtain the desired result from: (i) the strong consistency of $\hat{\boldsymbol{\theta}}_T \stackrel{a.s.}{\to} \boldsymbol{\theta}_0 \in \text{int}(\boldsymbol{\Theta})$; (ii) the a.s. twice continuous differentiability of $\mathcal{L}_T(\boldsymbol{\theta}, \bar{f}_1)$ in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$; (iii) the asymptotic normality of the score

$$\sqrt{T}\mathcal{L}_T'\big(\boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)}\big) \stackrel{d}{\to} \mathrm{N}(0, \mathcal{J}(\boldsymbol{\theta}_0)), \qquad \mathcal{J}(\boldsymbol{\theta}_0) = \mathbb{E}\big(\tilde{\ell}_t'(\boldsymbol{\theta}_0)\tilde{\ell}_t'(\boldsymbol{\theta}_0)^\top\big); \tag{B.6}$$

(iv) the uniform convergence of the likelihood's second derivative,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \big\|\mathcal{L}_T''(\boldsymbol{\theta}, \bar{\boldsymbol{f}}_1^{(0:2)}) - \mathcal{L}_\infty''(\boldsymbol{\theta})\big\| \stackrel{a.s.}{\to} 0; \tag{B.7}$$

and finally, (v) the non-singularity of the limit $\mathcal{L}_\infty''(\boldsymbol{\theta}) = \mathbb{E}\tilde{\ell}_t''(\boldsymbol{\theta}) = \mathcal{I}(\boldsymbol{\theta})$. See e.g. in White (1994, Theorem 6.2) for further details.

The consistency condition $\hat{\boldsymbol{\theta}}_T \stackrel{a.s.}{\to} \boldsymbol{\theta}_0 \in \text{int}(\boldsymbol{\Theta})$ in (i) follows under the maintained assumptions from Theorem B.7 and the additional assumption in Theorem B.10 that $\boldsymbol{\theta}_0 \in \text{int}(\boldsymbol{\Theta})$. The smoothness condition in (ii) follows immediately from Assumption B.5 and the likelihood expressions in Appendix B.4.

The asymptotic normality of the score in (B.9) follows by Theorem 18.10[iv] in van der Vaart (2000) by showing that

$$\|\mathcal{L}_T'\big(\boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)}\big) - \mathcal{L}_T'\big(\boldsymbol{\theta}_0\big)\| \stackrel{e.a.s.}{\to} 0 \ \text{ as } \ T \to \infty, \tag{B.8}$$

plus a CLT result for $\mathcal{L}_T'(\boldsymbol{\theta}_0)$. Note that from (B.8) we obtain that $\sqrt{T}\|\mathcal{L}_T'\big(\boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)}\big) - \mathcal{L}_T'\big(\boldsymbol{\theta}_0\big)\| \stackrel{a.s.}{\to} 0$ as $T \to \infty$. The desired CLT result follows by an application of the CLT for SE martingales in Billingsley (1961),

$$\sqrt{T}\mathcal{L}_T'\big(\boldsymbol{\theta}_0\big) \stackrel{d}{\to} \mathrm{N}\big(0, \mathcal{J}(\boldsymbol{\theta}_0)\big) \ \text{ as } \ T \to \infty, \tag{B.9}$$

where $\mathcal{J}(\boldsymbol{\theta}_0) = \mathbb{E}(\tilde{\ell}_t'\big(\boldsymbol{\theta}_0\big)\tilde{\ell}_t'\big(\boldsymbol{\theta}_0\big)^\top) < \infty$, where finite (co)variances follow from the assumption $N_{\ell'} \geq 2$ in Assumption B.9 and the expressions for the likelihood in Appendix B.4.

To establish the e.a.s. convergence in (B.8), we use the e.a.s. convergence

$$|\tilde{f}_t(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}_0, \bar{f}_1) - \tilde{f}_t(y^{t-1}, X^{t-1}, \boldsymbol{\theta}_0)| \overset{e.a.s.}{\to} 0, \tag{B.10}$$

and

$$\|\tilde{\boldsymbol{f}}_t^{(1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)}) - \tilde{\boldsymbol{f}}_t^{(1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}_0)\| \overset{e.a.s.}{\to} 0. \tag{B.11}$$

The e.a.s. convergence in (B.10) is obtained directly by application of Theorem 1 under the maintained assumptions. The e.a.s. convergence in (B.11) is obtained by the same argument as in the proof of Theorem 1 since: (a) the expressions for the derivative process $\{\tilde{\boldsymbol{f}}_t^{(1)}\}$ in Appendix B.4 show that the contraction condition

$$\mathbb{E} \log \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \bar{\phi}'_{1,1}(\boldsymbol{\theta}) < 0$$

for the recursion of the filter $\{\tilde{f}_t\}$ is the same as the contraction condition for the derivative process $\{\tilde{\boldsymbol{f}}_t^{(1)}\}$; and (b) the expressions in Appendix B.4 also reveal that the counterpart of the moment condition

$$\mathbb{E} \log^+ \sup_{(B,\lambda) \in \mathcal{B} \times \Lambda} |s(\bar{f}_1, y_t, X_t; B, \lambda)| < \infty,$$

used in Theorem 1 for the filtered process $\{\tilde{f}_t\}$, is implied by the condition that

$$\min\{N_f, N_s, N_s^{0,1,0}, N_s^{(0,0,1)}\} > 0,$$

as imposed in Assumption B.9.

From the differentiability of

$$\tilde{\ell}'_t(\boldsymbol{\theta}, \bar{\boldsymbol{f}}_1^{(0:1)}) = \ell'\big(\boldsymbol{\theta}, y^{1:t}, X^{1:t}, \tilde{\boldsymbol{f}}_t^{(0:1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}, \bar{\boldsymbol{f}}_1^{(0:1)})\big)$$

in $\tilde{\boldsymbol{f}}_t^{(0:1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}, \bar{\boldsymbol{f}}_1^{(0:1)})$ and the convexity of $\mathcal{F}$, we use the mean-value theorem to obtain

$$\begin{aligned} \|\mathcal{L}'_T(\boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)}) - \mathcal{L}'_T(\boldsymbol{\theta}_0)\| \leq & \sum_{j=1}^{4+d_\lambda} \Big|\frac{\partial \ell'(y^{1:t}, X^{1:t}, \hat{\boldsymbol{f}}_t^{(0:1)})}{\partial f_j}\Big| \\ & \times \big|\tilde{\boldsymbol{f}}_{j,t}^{(0:1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)}) - \tilde{\boldsymbol{f}}_{j,t}^{(0:1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}_0)\big|, \end{aligned} \tag{B.12}$$

where d_λ denotes the dimension of λ, and $\tilde{\boldsymbol{f}}_{j,t}^{(0:1)}$ denotes the j-th element of $\tilde{\boldsymbol{f}}_t^{(0:1)}$, and $\hat{\boldsymbol{f}}^{(0:1)}$ is on the segment connecting $\tilde{\boldsymbol{f}}_{j,t}^{(0:1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)})$ and $\tilde{\boldsymbol{f}}_{j,t}^{(0:1)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}_0)$. Note that $\tilde{\boldsymbol{f}}_t^{(0:1)} \in \mathbb{R}^{4+d_\lambda}$ because it contains $\tilde{f}_t \in \mathbb{R}$ (the first element) as well as $\tilde{\boldsymbol{f}}_t^{(1)} \in \mathbb{R}^{3+d_\lambda}$ (the derivatives with respect to ω, A, B, and λ). Using the expressions of the likelihood and its derivatives in Appendix B.4, the moment bounds and the moment preserving properties in Assumption B.9, and the expressions in Appendix B.4 shows that

$$\left|\partial \ell'(y^{1:t}, X^{1:t}, \hat{\boldsymbol{f}}_t^{(0:1)})/\partial f_j\right| = O_p(1) \quad \forall j = 1, \ldots, 4 + d_\lambda.$$

The strong convergence in (B.12) is now ensured by

$$\|\mathcal{L}_T'\big(\boldsymbol{\theta}_0, \bar{\boldsymbol{f}}_1^{(0:1)}\big) - \mathcal{L}_T'\big(\boldsymbol{\theta}_0\big)\| = \sum_{i=1}^{4+d_\lambda} O_p(1) o_{e.a.s}(1) = o_{e.a.s.}(1). \tag{B.13}$$

The proof of the uniform convergence in (B.7) is similar to that of Theorem B.4. We note

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathcal{L}_T''(\boldsymbol{\theta}, \bar{f}_1) - \mathcal{L}_\infty''(\boldsymbol{\theta})\| \leq \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathcal{L}_T''(\boldsymbol{\theta}, \bar{f}_1) - \mathcal{L}_T''(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\mathcal{L}_T''(\boldsymbol{\theta}) - \mathcal{L}_\infty''(\boldsymbol{\theta})\|. \tag{B.14}$$

To prove that the first term vanishes a.s., we show that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\tilde{\ell}_t''(\boldsymbol{\theta}, \bar{f}_1) - \tilde{\ell}_t''(\boldsymbol{\theta})\| \overset{a.s.}{\to} 0 \quad \text{as} \quad t \to \infty.$$

The differentiability of $\tilde{g}$, $\tilde{g}'$, $\tilde{p}$, and S from Assumption B.5 ensure that

$$\tilde{\ell}_t''(\cdot, \bar{f}_1) = \ell''(y_t, \tilde{\boldsymbol{f}}_t^{(0:2)}(y^{1:t-1}, X^{1:t-1}, \cdot, \bar{\boldsymbol{f}}_{0:2}), \cdot)$$

is continuous in $(y_t, \tilde{\boldsymbol{f}}_t^{(0:2)}(y^{1:t-1}, X^{1:t-1}, \cdot, \boldsymbol{f}_{0:2}))$. Again, we note that the proof of Theorem B.1 can be easily adapted to show that there exists a unique SE sequence $\{\tilde{\boldsymbol{f}}_t^{(0:2)}(y^{t-1}, X^{t-1}, \cdot)\}_{t\in\mathbb{Z}}$ such that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\|(y_t, \tilde{\boldsymbol{f}}_t^{(0:2)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}, \bar{\boldsymbol{f}}_{0:2})) - (y_t, \tilde{\boldsymbol{f}}_t^{(0:2)}(y^{t-1}, X^{t-1}, \boldsymbol{\theta})\right\| \overset{a.s.}{\to} 0,$$

and satisfying, for for $N_f \geq 1$,

$$\sup_t \mathbb{E} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\tilde{\boldsymbol{f}}_t^{(0:2)}(y^{1:t-1}, X^{1:t-1}, \boldsymbol{\theta}, \bar{\boldsymbol{f}}_{0:2})\|^{N_f} < \infty,$$

and also

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\tilde{\boldsymbol{f}}_t^{(0:2)}(y^{t-1}, X^{t-1}, \boldsymbol{\theta})\|^{N_f} < \infty,$$

because (a) the expressions for the derivative process $\{\tilde{\boldsymbol{f}}_t^{(1)}\}$ in Appendix B.4 show that the contraction condition

$$\mathbb{E}\log\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \bar{\phi}'_{1,1}(\boldsymbol{\theta}) < 0$$

for the recursion of the filter $\{\tilde{f}_t\}$ is the same as the contraction condition for the second derivative process $\{\tilde{\boldsymbol{f}}_t^{(2)}\}$; and (b) the expressions in Appendix B.4 show also that the counterpart of the moment condition

$$\mathbb{E}\log^{+}\sup_{(B,\lambda)\in\mathcal{B}\times\Lambda} |s(\bar{f}_1, y_t, X_t; B, \lambda)| < \infty,$$

used in Theorem 1 for the filtered process $\{\tilde{f}_t\}$, is implied by the condition that

$$\min\Bigg\{ N_f^{(1)}\ ,\ N_s^{(0,1,0)}\ ,\ N_s^{(0,0,1)}\ ,\ N_s^{(0,2,0)}\ ,\ N_s^{(0,0,2)}\ ,\ N_s^{(0,1,1)}\ ,\ \frac{N_s^{(1,0,0)}N_f^{(1)}}{N_s^{(1,0,0)}+N_f^{(1)}}\ ,$$
$$\frac{N_s^{(2,0,0)}N_f^{(1)}}{2N_s^{(2,0,0)}+N_f^{(1)}}\ ,\ \frac{N_s^{(1,1,0)}N_f^{(1)}}{N_s^{(1,1,0)}+N_f^{(1)}}\ ,\ \frac{N_s^{(1,0,1)}N_f^{(1)}}{N_s^{(1,0,1)}+N_f^{(1)}}\Bigg\} > 0\ ,$$

imposed in Assumption B.9. By application of a continuous mapping theorem for $\ell'' : \mathbb{C}(\boldsymbol{\Theta}\times\mathcal{F}^{(0:2)}) \to \mathbb{R}$ we thus conclude that the first term in (B.14) converges to 0 a.s..

The second term in (B.14) converges under a bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\tilde{\ell}''_t(\boldsymbol{\theta})\| < \infty$ by the SE nature of $\{\mathcal{L}''_T\}_{t\in\mathbb{Z}}$. The latter is implied by continuity of ℓ'' on the SE sequence

$$\{(y_t, X_t, \tilde{\boldsymbol{f}}_t^{(0:2)}(y^{1:t-1}, X^{1:t-1}, \cdot))\}_{t\in\mathbb{Z}}.$$

The moment bound $\mathbb{E}\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\tilde{\ell}''_t(\boldsymbol{\theta})\| < \infty$ follows from $N_{\ell''} \geq 1$ in Assumption B.9 and the expressions in Appendix B.4. Finally, the non-singularity of the limit $\mathcal{L}''_\infty(\boldsymbol{\theta}) = \mathbb{E}\tilde{\ell}''_t(\boldsymbol{\theta}) = \mathcal{I}(\boldsymbol{\theta})$ in (v) below equation (B.7) is implied by the uniqueness of $\boldsymbol{\theta}_0$ as a maximum of $\mathcal{L}''_\infty(\boldsymbol{\theta})$ in $\boldsymbol{\Theta}$. □

B.4 Derivatives of the likelihood function

We take first derivatives of the likelihood with respect to all static parameters $\boldsymbol{\theta} = (\omega, A, B, \beta', \sigma^2)'$:

$$\frac{\partial \ell_t}{\partial \boldsymbol{\theta}} = \left(\frac{\partial \ell_t}{\partial \omega}, \frac{\partial \ell_t}{\partial a}, \frac{\partial \ell_t}{\partial b}, \frac{\partial \ell_t}{\partial \beta}, \frac{\partial \ell_t}{\partial \sigma^2}\right)'$$

Let θ_m denote the mth element of $\boldsymbol{\theta}$, $\tilde{p}_t = \log p_e(y_t - h(\tilde{f}_t)Wy_t - X_t\beta)$, and $\tilde{g}'_t = \log|Z(\tilde{f}_t)^{-1}|$.

We can decompose the derivatives of the likelihood with respect to each θ_m into two parts:

$$\begin{aligned}\frac{\partial \ell_t}{\partial \theta_m} &= \frac{\partial(\tilde{p}_t + \log \tilde{g}_t')}{\partial f_t} \cdot \frac{\partial f_t}{\partial \theta_m} + \frac{\partial \tilde{p}_t}{\partial \theta_m} \\ &= \nabla_t \cdot \frac{\partial f_t}{\partial \theta_m} + \frac{\partial \tilde{p}_t}{\partial \theta_m}, \end{aligned} \tag{B.15}$$

because $\tilde{g}_t'$ does not depend on any of the parameters directly, only through f_t. For $\theta_m \in \{\omega, A, B\}$ the second term is zero, because these parameters enter the likelihood only through f_t.

All partial derivatives contain the term $\partial f_t / \partial \theta_m$ given by

$$\frac{\partial f_t}{\partial \theta_m} = \frac{\partial}{\partial \theta_m}\left(\omega + A s_{t-1} + B f_{t-1}\right) \tag{B.16}$$

$$= \frac{\partial \omega}{\partial \theta_m} + \frac{\partial A}{\partial \theta_m} s_{t-1} + A \frac{\partial s_{t-1}}{\partial f_{t-1}} \cdot \frac{\partial f_{t-1}}{\partial \theta_m} + A \frac{\partial s_{t-1}}{\partial \theta_m} + \frac{\partial B}{\partial \theta_m} f_{t-1} + B \frac{\partial f_{t-1}}{\partial \theta_m} \tag{B.17}$$

$$= \frac{\partial \omega}{\partial \theta_m} + \frac{\partial A}{\partial \theta_m} \nabla_{t-1} + A \nabla_{t-1}' \cdot \frac{\partial f_{t-1}}{\partial \theta_m} + A \frac{\partial \nabla_{t-1}}{\partial \theta_m} + \frac{\partial B}{\partial \theta_m} f_{t-1} + B \frac{\partial f_{t-1}}{\partial \theta_m} \tag{B.18}$$

$$= \frac{\partial \omega}{\partial \theta_m} + \frac{\partial A}{\partial \theta_m} \nabla_{t-1} + A \frac{\partial \nabla_{t-1}}{\partial \theta_m} + \frac{\partial B}{\partial \theta_m} f_{t-1} + (A \nabla_{t-1}' + B) \frac{\partial f_{t-1}}{\partial \theta_m} \tag{B.19}$$

We want the matrix of second derivatives of the likelihood function, i.e.

$$\frac{\partial^2 \ell_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}.$$

We take another derivative of (B.15) with respect to θ_o:

$$\frac{\partial^2 \ell_t}{\partial \theta_m \partial \theta_o} = \nabla_t' \cdot \frac{\partial f_t}{\partial \theta_o} \cdot \frac{\partial f_t}{\partial \theta_m} + \frac{\partial \nabla_t}{\partial \theta_o} \cdot \frac{\partial f_t}{\partial \theta_m} + \nabla_t \frac{\partial^2 f_{t-1}^2}{\partial \theta_m \theta_o} + \frac{\partial^2 \tilde{p}_t}{\partial \theta_m \partial \theta_o} \tag{B.20}$$

The second derivative process takes the form

$$
\begin{aligned}
\frac{\partial^2 f_t}{\partial\theta_m\partial\theta_o} =& \frac{\partial A}{\partial\theta_m}\cdot\frac{\partial\nabla_{t-1}}{\partial f_{t-1}}\cdot\frac{\partial f_{t-1}}{\partial\theta_o}+\frac{\partial A}{\partial\theta_m}\frac{\partial\nabla_{t-1}}{\partial\theta_o} \\
&+\frac{\partial A}{\partial\theta_o}\frac{\partial\nabla_{t-1}}{\partial f_{t-1}}\frac{\partial f_{t-1}}{\partial\theta_m}+A\frac{\partial^2\nabla_{t-1}}{\partial f_{t-1}^2}\frac{\partial f_{t-1}}{\partial\theta_o}\frac{\partial f_{t-1}}{\partial\theta_m}+A\frac{\partial^2\nabla_{t-1}}{\partial f_{t-1}\partial\theta_o}\frac{\partial f_{t-1}}{\partial\theta_m} \\
&+A\frac{\partial\nabla_{t-1}}{\partial f_{t-1}}\frac{\partial^2 f_{t-1}}{\partial\theta_m\partial\theta_o}+\frac{\partial A}{\partial\theta_o}\frac{\partial\nabla_{t-1}}{\partial\theta_m}+A\frac{\partial^2\nabla_{t-1}}{\partial\theta_m\partial\theta_o}+A\frac{\partial^2\nabla_{t-1}}{\partial\theta_m\partial f_{t-1}}\frac{\partial f_{t-1}}{\partial\theta_o} \\
&+\frac{\partial B}{\partial\theta_m}\frac{\partial f_{t-1}}{\partial\theta_o}+\frac{\partial B}{\partial\theta_o}\frac{\partial f_{t-1}}{\partial\theta_m}+B\frac{\partial^2 f_{t-1}}{\partial\theta_m\partial\theta_o} \\
=&\frac{\partial A}{\partial\theta_m}\cdot\nabla'_{t-1}\cdot\frac{\partial f_{t-1}}{\partial\theta_o}+\frac{\partial A}{\partial\theta_m}\frac{\partial\nabla_{t-1}}{\partial\theta_o} \\
&+\frac{\partial A}{\partial\theta_o}\cdot\nabla'_{t-1}\cdot\frac{\partial f_{t-1}}{\partial\theta_m}+A\nabla''_{t-1}\cdot\frac{\partial f_{t-1}}{\partial\theta_o}\frac{\partial f_{t-1}}{\partial\theta_m}+A\frac{\partial\nabla'_{t-1}}{\partial\theta_o}\frac{\partial f_{t-1}}{\partial\theta_m} \\
&+A\nabla'_{t-1}\frac{\partial^2 f_{t-1}}{\partial\theta_m\partial\theta_o}+\frac{\partial A}{\partial\theta_o}\frac{\partial\nabla_{t-1}}{\partial\theta_m}+A\frac{\partial^2\nabla_{t-1}}{\partial\theta_m\partial\theta_o}+A\frac{\partial^2\nabla_{t-1}}{\partial\theta_m\partial f_{t-1}}\frac{\partial f_{t-1}}{\partial\theta_o} \\
&+\frac{\partial B}{\partial\theta_m}\frac{\partial f_{t-1}}{\partial\theta_o}+\frac{\partial B}{\partial\theta_o}\frac{\partial f_{t-1}}{\partial\theta_m}+B\frac{\partial^2 f_{t-1}}{\partial\theta_m\partial\theta_o}.
\end{aligned}
\tag{B.21}
$$

Appendix C Additional tables and figures

Table C.1: Key policy events during the Eurozone crisis

Date	Event	Source
Oct. 18, 2009	Greece announces doubling of budget deficit	The Guardian[1]
Mar. 3, 2010	EU offers financial help to Greece	ECB[2]
Dec. 7, 2010	Ireland is bailed out by EU and IMF	ECB[2]
Dec. 22, 2011	ECB launches the first Longer-Term Refinancing Operation (LTRO)	ECB[2]
Mar. 1, 2012	ECB launches the second LTRO	ECB[2]
Jul. 26, 2012	M. Draghi: "[T]he ECB is ready to do whatever it takes to preserve the euro."	ECB[3]
Oct. 8, 2012	European Stability Mechanism (ESM) is inaugurated	ESM[4]
Sep. 12, 2013	European Parliament approves new unified bank supervision system	ECB[2]

[1] http://www.theguardian.com/business/2012/mar/09/greek-debt-crisis-timeline
[2] http://www.ecb.europa.eu/ecb/html/crisis.de.html
[3] http://www.ecb.europa.eu/press/key/date/2012/html/sp120726.en.html
[4] http://www.esm.europa.eu/press/releases/20121008_esm-is-inaugurated.htm
All retrieved on June 19, 2014.

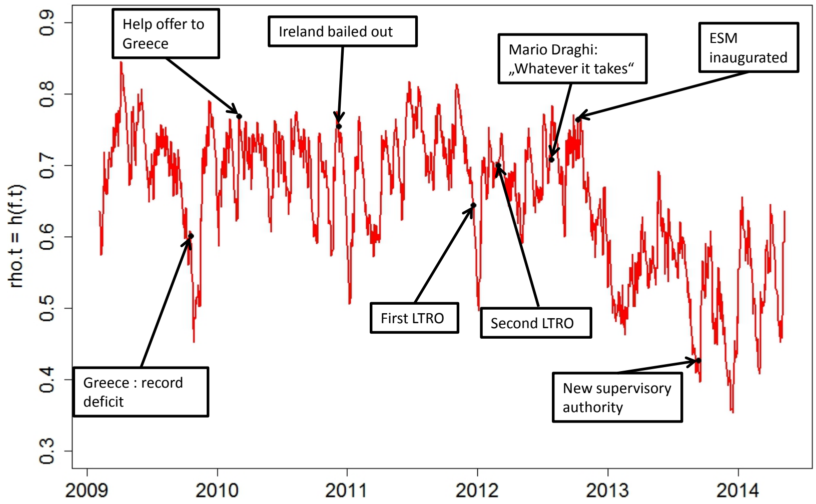

Figure C.1: Filtered spatial dependence parameters obtained from the full model, together with key policy events from Table C.1.

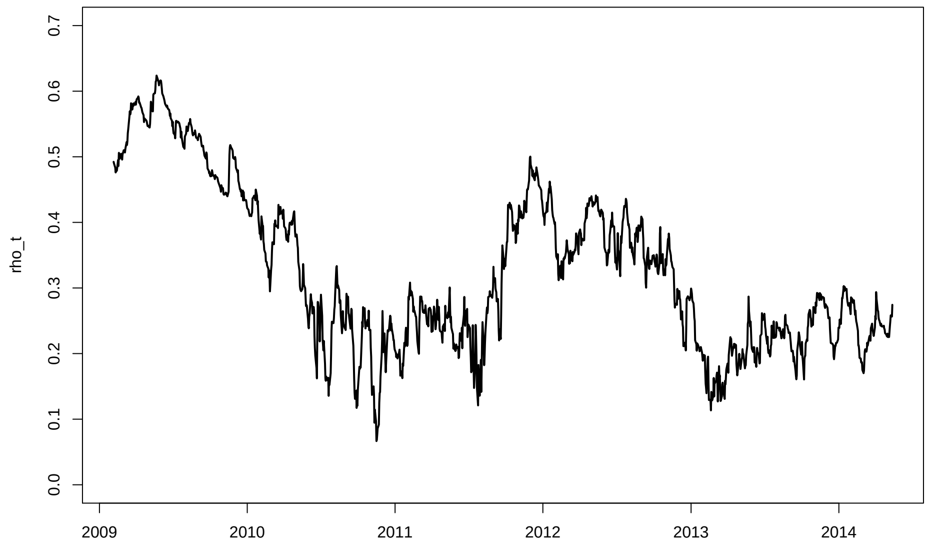

Figure C.2: Filtered spatial parameter obtained from the time-varying spatial score model with time-varying volatilities, using absolute CDS spread changes as dependent variable.

Table C.2: Cross-correlation matrices: raw data and full model residuals

Correlation matrix of raw CDS change data								
	Belgium	France	Germany	Ireland	Italy	Netherlands	Portugal	Spain
Belgium	1.000	0.724	0.697	0.630	0.738	0.737	0.643	0.707
France		1.000	0.724	0.581	0.655	0.678	0.581	0.657
Germany			1.000	0.553	0.609	0.685	0.534	0.577
Ireland				1.000	0.718	0.575	0.724	0.685
Italy					1.000	0.654	0.740	0.847
Netherlands						1.000	0.566	0.620
Portugal							1.000	0.742
Spain								1.000

Correlation matrix of residuals								
	Belgium	France	Germany	Ireland	Italy	Netherlands	Portugal	Spain
Belgium	1.000	0.118	0.146	-0.089	0.204	0.038	0.132	0.199
France		1.000	0.201	-0.136	-0.159	0.032	-0.035	-0.020
Germany			1.000	-0.270	-0.453	-0.002	-0.087	-0.223
Ireland				1.000	0.174	-0.039	0.222	0.038
Italy					1.000	-0.003	0.305	0.577
Netherlands						1.000	-0.034	-0.064
Portugal							1.000	0.041
Spain								1.000

References

Billingsley, P. (1961). The lindeberg-levy theorem for martingales. *Proceedings of the American Mathematical Society*, 12(5):788–792.

Billingsley, P. (1995). *Probability and Measure*. Wiley-Interscience.

Blasques, F., Koopman, S. J., and Lucas, A. (2014). Maximum likelihood estimation for generalized autoregressive score models. *Tinbergen Institute Discussion Papers 14-029/III*.

Bougerol, P. (1993a). *Kalman Filtering with Random Coefficients and Contractions*. Prépublications de l'Institut Elie Cartan. Univ. de Nancy.

Bougerol, P. (1993b). Kalman filtering with random coefficients and contractions. *SIAM J. Control Optim.*, 31(4):942–959.

Dudley, R. M. (2002). *Real Analysis and Probability*. Cambridge Studies in Advanced Mathematics. Cambridge University Press.

Foland, G. B. (2009). *A Guide to Advanced Real Analysis*. Dolciani Mathematical Expositions. Cambridge University Press.

Gallant, R. and White, H. (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Cambridge University Press.

Krengel, U. (1985). *Ergodic theorems*. De Gruyter studies in Mathematics, Berlin.

Ranga Rao, R. (1962). Relations between weak and uniform convergence of measures with applications. *Annals of Mathematical Statistics*, 33:659–680.

Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroeskedastic time series: A stochastic recurrence equations approach. *The Annals of Statistics*, 34(5):2449–2495.

van der Vaart, A. W. (2000). *Asymptotic Statistics (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge Books. Cambridge University Press.