Duisenberg school of finance - Tinbergen Institute Discussion Paper

TI 13-097/IV/59

Stationarity and Ergodicity Regions for Score Driven Dynamic Correlation Models

Francisco Blasques[a,b]
André Lucas[a,b]
Erkki Silde[a,b,c]

[a] *Faculty of Economics and Business Administration, VU University Amsterdam;*
[b] *Tinbergen Institute;*
[c] *Duisenberg school of finance.*

Stationarity and Ergodicity Regions for Score Driven Dynamic Correlation Models*

Francisco Blasques[a,b,c], André Lucas[a,c] and Erkki Silde[a,c,d]

(a) Department of Finance, VU University Amsterdam

(b) Department of Econometrics, VU University Amsterdam

(c) Tinbergen Institute (d) Duisenberg school of finance

July 16, 2013

Abstract

We describe stationarity and ergodicity (SE) regions for a recently proposed class of score driven dynamic correlation models. These models have important applications in empirical work. The regions are derived from sufficiency conditions in Bougerol (1993) and take a non-standard form. We show that the non-standard shape of the sufficiency regions cannot be avoided by reparameterizing the model or by rescaling the score steps in the transition equation for the correlation parameter. This makes the result markedly different from the volatility case. Observationally equivalent decompositions of the stochastic recurrence equation yield regions with different sizes and shapes. We illustrate our results with an analysis of time-varying correlations between UK and Greek equity indices. We find that also in empirical applications different decompositions can give rise to different conclusions regarding the stability of the estimated model.

Keywords: dynamic copulas, generalized autoregressive score (GAS) models, stochastic recurrence equations, observation driven models, contraction properties.

JEL Classifications: C22, C32, C58.

*We thank Peter Boswijk, Andrew Harvey, Pawel Janus and participants of the Amsterdam-Cambridge Conference on Score Driven Models for helpful discussions. All three authors also thank the Dutch Science Foundation (NWO) for financial support. Francisco Blasques: f.blasques@vu.nl, André Lucas: a.lucas@vu.nl, Erkki Silde: e.silde@vu.nl; VU University Amsterdam, FEWEB/FIN, de Boelelaan 1105, 1081 HV Amsterdam, Netherlands.

1 Introduction

Time-variation in correlations is an important feature of economic and financial data and a crucial ingredient of empirical analyses, such as the assessment of risk and the construction of investment portfolios. Available models for capturing the time-variation in correlations include, amongst many others, the BEKK model of Engle and Kroner (1995), the switching correlation models of Pelletier (2006), the DCC model of Engle (2002) with its adaptation by Aielli (2013), the DECO model of Engle and Kelly (2012), the dynamic copula models of Patton (2009) and Oh and Patton (2012), and the score driven models of Creal et al. (2011, 2013a) and Harvey (2013); see also the overviews of Bauwens et al. (2006) and Silvennoinen and Teräsvirta (2009).

Here, we focus on the stochastic properties of the recently proposed score driven models of Creal et al. (2011, 2013a) and Harvey (2013), which we refer to as the generalized autoregressive score (GAS) model. These models have been shown to be particularly useful when modeling fat-tailed or skewed data, such as often encountered in empirical finance. The dynamics of correlations and volatilities in these models are driven by the score of the error distribution. If the latter is fat-tailed, the score driven dynamics automatically correct for influential observations, see Creal et al. (2011). In this way, they share some similarities with models from the robust GARCH literature; see for example Boudt et al. (2013). The score driven approach used in the construction of GAS models, however, provides a much more general and unified framework for parameter dynamics that is applicable far beyond the volatility and correlation context; see Creal et al. (2013a,b) for a range of other examples. In addition, from a forecasting perspective GAS models have a similar performance to correctly specified state-space models, see Koopman et al. (2012).

Despite their proven empirical usefulness, the theoretical properties of GAS models are less well developed. The complication lies in the highly nonlinear dynamics of the time-varying parameter in typical GAS models. In this paper we contribute to our understanding of the stochastic properties of GAS models for dynamic correlations. We do so by studying which models and which parameter values generate stationary and ergodic (or SE from now on) time series processes. This offers an important characterization of the stochastic properties of GAS models. Furthermore, together with the existence of certain unconditional moments, it constitutes an important ingredient in proofs of consistency and asymptotic normality of extremum estimators; see e.g. Straumann and Mikosch (2006) for maximum likelihood estimation of nonlinear conditional volatility models, Francq and

Zakoian (2011) and Boussama et al. (2011) for the case of GARCH models, and Harvey (2013) for GAS volatility models. For each correlation model we consider, we define the parameter values that ensure the SE property and call this the 'SE region' of the parameter space. To establish SE regions, we follow the classical average contraction argument for stochastic recurrence relations as laid out in the sufficient conditions formulated by Bougerol (1993). Given these conditions, we compute numerically the SE regions for a range of empirically relevant models.

Our contribution is threefold. First, we are the first to derive SE regions for the class of score driven correlation models that have been suggested recently in the literature. We show that the SE sufficiency regions take a highly non-standard form, dissimilar to the familiar triangle and curved triangular shape for the GARCH model, see Nelson (1990). In an empirical example, we demonstrate that the conditions for nonlinear recurrence equations can be used to ensure stationarity of concrete models, applied on real data. This also extends the results in Blasques et al. (2012) for volatility and tail index models with univariate observations to the case of time-varying parameters and multivariate observations. Second, we show that the shape and size of the SE sufficiency region as derived from the conditions of Bougerol (1993) depends on the way the stochastic recurrence equation for the correlation is constructed from bivariate uncorrelated noise. In particular, we show that the choice of the square root of the correlation matrix in this construction has a non-trivial effect on the size of the SE sufficiency region. Third, we show analytically why the correlation case is markedly different from the volatility case. For the volatility case, Harvey (2013) shows that modeling the log-volatility renders the information matrix independent of the time-varying volatility. The resulting stochastic recurrence equation becomes linear, and we can use linear processes theory to study the SE properties in that case. A similar feature is generally not available for the dynamic correlation model: neither a reparameterization of the correlation, nor a scaling of the score steps, makes the stochastic recurrence equation a linear process. The reason is that unlike the volatility case, the GAS steps for the correlation model consist of two separate terms with different nonlinearities in the correlation parameter.

The remainder of this paper is organized as follows. In Section 2, we introduce the GAS model for dynamic bivariate correlations. In Section 3 we state the conditions for the SE sufficiency regions. In Section 4, we determine the SE regions numerically for a number of different models. We also show where typical empirical estimates are located with respect to the SE region's boundaries by investigating a time-varying correlation model for UK and Greek equity indices. We conclude in Section 5. The Appendix gathers the more technical results and derivations.

2 Generalized Autoregressive Score models for correlations

Consider a real-valued bivariate stochastic sequence of observations $\{y_t, t \in \mathbb{N}\}$ generated by a zero mean elliptical conditional distribution with time-varying correlation matrix $R(f_t)$,

$$y_t|f_t \overset{i.i.d.}{\sim} \frac{p(y_t' R(f_t)^{-1} y_t)}{|R(f_t)|^{1/2}}, \qquad R(f_t) = \begin{pmatrix} 1 & \rho(f_t) \\ \rho(f_t) & 1 \end{pmatrix}, \tag{1}$$

where $p : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ denotes a real-valued density generator function in the quadratic form $y_t' R(f_t)^{-1} y_t$, the sequence $\{f_t\}_{t\in\mathbb{N}}$ is a real-valued sequence for the time-varying parameter f_t, $\rho(f_t) \in (-1, 1)$ is the dynamic correlation parameter at time t, and $R(f_t)$ is the correlation matrix at time t. For example, if y_t is conditionally normal, we have $p(x) = (2\pi)^{-1}\exp(-x/2)$. We fully focus the exposition on the correlation case by restricting the variances in (1) to one. Time-varying variances in score driven models have already been dealt with in for example Creal et al. (2011) and Harvey (2013). The formulation in (1) can also be interpreted as a copula model, see the discussion in Patton (2009). Under the assumptions of stationary marginals and no volatility spillovers, stability conditions for the copula then lead to stability of the whole model. The class of elliptical models is also economically interesting, as it enables an analytic characterization of the resulting potfolio returns and the risk-return tradeoff; see for example Chamberlain (1983), Owen and Rabinovitch (1983), and Hamada and Valdez (2008).

Following Creal et al. (2011,2013a), the generalized autoregressive score (GAS) dynamics for the time-varying parameter f_t in (1) take the form

$$f_{t+1} = \omega + \beta f_t + \alpha s(f_t, y_t), \tag{2}$$

$$s(f_t, y_t) = S(f_t) q(y_t, f_t), \qquad q(y_t, f_t) = \frac{\partial}{\partial f} \log \frac{p(y_t' R(f)^{-1} y_t)}{|R(f)|^{1/2}}\Bigg|_{f=f_t}, \tag{3}$$

with initial condition $f_1 \in \mathcal{F}$, where $(\omega, \alpha, \beta) \in \mathbb{R}^3$ is a vector of time-invariant parameters, and S is a scaling function for the score of the conditional observation density. We define the parameter vector $\theta \in \Theta$ as $\theta = (\omega, \alpha, \beta)$, with $\Theta \subseteq \mathbb{R}^3$ denoting the parameter space.

The dynamic specification in (2) can easily be extended to include more lags of either $s(f_t, y_t)$ and/or f_t. For the case of the bivariate correlation model (1), we obtain

$$q(y_t, f_t) \;=\; \frac{\dot{\rho}(f_t)}{1-\rho(f_t)^2}\left(\dot{p}(y_t' R(f_t)^{-1} y_t)\left(2\rho(f_t) y_t' R(f_t)^{-1} y_t - y_t' \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right) y_t\right) + \rho(f_t)\right) \tag{4}$$

with $\dot{p}(x) = \partial \log p(x)/\partial x$ and $\dot{\rho}(f_t) = \partial\rho(f)/\partial f|_{f=f_t}$.

Each choice for the scaling function S in (3) gives rise to a new GAS model. An often used choice of S relates to the local curvature of the score as measured by the information matrix, for example

$$S(f) = (\mathcal{I}_t(f))^{-a}, \qquad \mathcal{I}_t(f) = \mathrm{E}_{t-1}[q(y_t, f) q(y_t, f)'], \tag{5}$$

where a is typically taken as 0, 1/2 or 1.

The parameter dynamics in (2) and (3) are intuitive. The time-varying parameter f_t is updated in the (scaled) direction of steepest ascent as measured by the scaled conditional log observation density at time t. For example, standard GARCH and ACD model are special cases of the GAS framework, see Creal et al. (2013a). The GAS set-up is very general and can also easily be applied outside the correlation context as long as a conditional observation density is available. For other examples, including many new models, we again refer to Creal et al. (2013a,b).

3 Conditions for stationarity and ergodicity

We follow the approach of Blasques et al. (2012), who consider a treatment of univariate GAS models. Our stationarity and ergodicity (SE) results build on the stochastic recurrence relations or iterated random functions approach; see Diaconis and Freedman (1999) and Wu and Shao (2004). In particular, we use the sufficient conditions of Bougerol (1993) and results in Straumann and Mikosch (2006) to establish exponentially fast almost sure convergence of the time series $(y_t, f_t)_{t\in\mathbb{N}}$ generated by (1)–(3) with initial condition $f_1 \in \mathcal{F}$ to a unique SE sequence $(\tilde{y}_t, \tilde{f}_t)_{t\in\mathbb{Z}}$.

Let $\mathcal{F} \subseteq \mathbb{R}$ and $\mathcal{Y} \subseteq \mathbb{R}^2$ denote the domains of f_t and y_t, respectively. We have that $\rho : \mathcal{F} \to (-1, 1)$ and $s : \mathcal{F} \times \mathcal{Y} \to \mathbb{R}$ and derivatives thereof are almost surely (a.s.) smooth in f_t. Also we require the measurability of s and of its derivatives with respect to u_t.

Using the model as specified in (1)–(4), we analyze the stochastic properties of y_t and f_t via the stochastic recurrence equation

$$f_{t+1} = \phi_t(f_t;\theta), \tag{6}$$

where

$$\phi_t(f;\theta) \;=\; \omega + \beta f + \alpha S(f)\, q(h(f)u_t, f), \qquad h(f)h(f)' = R(f), \tag{7}$$

and $\{u_t\}$ is an independent and identically distributed (i.i.d.) sequence with $y_t = h(f_t)u_t$. We notice two particular features of equations (6) and (7). First, the dynamics of $\{f_t\}$ are now written in terms of the innovation sequence $\{u_t\}$ rather than the observed data $\{y_t\}$ by substituting $h(f_t)u_t$ for y_t. As a result, when seen as a function of f, the shape of $q(h(f)u_t, f)$, for every u_t, is markedly different from that of $q(y_t, f)$, for every y_t. This additional dependence on f may either complicate or simplify the nonlinear dependence of f_{t+1} on f_t as embedded in (7). Second, the functional form of (7) is not uniquely defined. Each square root $h(f)$ of the correlation matrix $R(f)$ leads to an observationally equivalent model in y_t. The choice of $h(f)$, however, is not innocuous for determining the size and shape of the SE region, as we see later.

Continuity of ϕ_t in u_t for every t can be used to ensure that $\{\phi_t\}$ is an i.i.d. sequence of functions. Together with equations (6)–(7), it then follows directly from Bougerol (1993) and Straumann and Mikosch (2006) that there is a unique SE solution to (1)–(3) if ϕ_t is contracting on average, i.e., if the Lyapunov exponent of the mapping is negative. In particular, we obtain the desired SE result if

$$\mathrm{E}\log \sup_{f,f^*\in\mathcal{F}} \frac{|\phi_t(f;\theta) - \phi_t(f^*;\theta)|}{|f - f^*|} \leq \mathrm{E}\log \sup_{f\in\mathcal{F}} \left| \frac{\partial \phi_t(f;\theta)}{\partial f} \right| \;<\; 0; \tag{8}$$

see Bougerol (1993). In computing the supremum in condition (8), f is treated as a parameter rather than as the random variable f_t.

For the score driven dynamic correlation model of Section 2, we prove the following result in the Appendix.

Lemma 1. *Let Ψ be a class of functions such that for every $\psi \in \Psi$, $\psi \in \mathcal{C}^1([-1,1],\mathbb{R})$ with $\dot\psi(\rho) = \partial\psi(\rho)/\partial\rho = \mathcal{O}((1-\rho^2)^{-1/2})$. Assume that $\mathrm{E}|\dot p(u_t'u_t)u_{i,t}u_{j,t}| < \infty$ for $i,j \in \{1,2\}$, with $u_t = (u_{1,t}, u_{2,t})'$. The process $\{f_t\}_{t\in\mathbb{N}}$ generated by the dynamic correlation model*

(1)–(4) *converges exponentially fast almost surely*[1] *(e.a.s.) to a unique stationary and ergodic process* $\{\tilde{f}_t\}_{t\in\mathbb{Z}}$ *for any initial condition* $f_1 \in \mathcal{F}$ *if*

$$\inf_{\psi\in\Psi} \mathrm{E}\log\sup_{f\in\mathcal{F}}\left|\beta + \alpha\left(\frac{\partial}{\partial f}\left(\frac{S(f)\,\dot{\rho}(f)}{1-\rho(f)^2}\right)\right) g(\rho(f))k(u_t) + \alpha\frac{S(f)\,\dot{\rho}(f)^2}{1-\rho(f)^2}\dot{g}(\rho(f))k(u_t)\right| < 0, \tag{9}$$

where

$$g(\rho) = \Big(\rho\ ,\ \rho c_{2\psi}(\rho) - \sqrt{1-\rho^2}s_{2\psi}(\rho)\ ,\ \sqrt{1-\rho^2}c_{2\psi}(\rho) + \rho s_{2\psi}(\rho)\Big), \tag{10}$$

$$k(u_t) = \big(\dot{p}(u_t'u_t)u_t'u_t + 1\ ,\ \dot{p}(u_t'u_t)(u_{1,t}^2 - u_{2,t}^2)\ ,\ -2\dot{p}(u_t'u_t)u_{1,t}u_{2,t}\big)', \tag{11}$$

$\dot{g}(\rho) = \partial g(\rho)/\rho$, $c_{2\psi}(\rho) = \cos(2\psi(\rho))$, *and* $s_{2\psi}(\rho) = \sin(2\psi(\rho))$.

We note several features of the result stated in Lemma 1. First, the SE region depends directly only on the parameters α and β, on the functional forms of $S(f)$ and $q(h(f)u_t, f)$, and on the density of u_t. The depencence on the latter enters in two ways, namely through the expectations operator in (9) and through the functional form of $k(u_t)$ in (11). Also note that the expectations operator in (9) does not require the second moments of u_t to exist. Rather, we only require the expectation of $|\dot{p}(u_t'u_t)u_{i,t}u_{j,t}|$ for $i,j \in \{1,2\}$ to exist. This condition is much weaker, particularly for fat-tailed elliptical densities. For example, it is easily satisfied for the bivariate Cauchy distribution, even though neither the second, nor the first moment exists for this distribution. The continuity and boundedness properties of s can be verified immediately for parametric distributional forms, notably for the Student's t density in Section 4.1.[2] Therefore, condition (9) effectively forms a sufficient condition for the SE property of the model.

Second, equation (9) directly reveals that the correlation case is markedly different from the volatility case. For the volatility case, it is shown in Harvey (2013) and Blasques et al. (2012) that through a clever choise of parameterization h or scaling S the scaled score in recurrence relation (7) can be made independent of f_t The SE condition then reduces to the requirement that $|\beta| < 1$. In the volatility case the analogue of the function $g(\rho)$ is scalar valued. In the correlation case, (9) shows that through the trivariate nature of the functions $g(\rho)$ and $\dot{g}(\rho)$ the contraction condition consists of a number of different terms, each with a different nonlinear dependence on f. It is impossible to off-set all of

[1]A sequence $\{x_t\}$ converges exponentially fast almost surely to a sequence $\{\tilde{x}_t\}$ if for some constant $c > 1$ we have $c^t \cdot |x_t - \tilde{x}_t| \stackrel{a.s.}{\to} 0$ for $t \to \infty$.

[2]The functional forms for the updating equation for the particular case of the Student's t distribution are presented in the online appendix (http://bit.ly/12GFbro) accompanying this paper.

these simultaneously by a single choice of scaling function or parameterization. This makes the correlation model theoretically more interesting in its own right.

Third, the SE sufficient condition in (9) has an additional degree of flexibility provided by the choice of ψ. As follows from the proof of Lemma 1, the function ψ determines which square root $h(f)$ is used for the correlation matrix $R(f)$. For the purpose of guaranteeing a proper correlation matrix, define $\phi(\rho) = \arcsin(\rho) - \psi(\rho)$, and

$$h(f) = \begin{pmatrix} \cos(\phi(\rho(f))) & \sin(\phi(\rho(f))) \\ \sin(\psi(\rho(f))) & \cos(\psi(\rho(f))) \end{pmatrix}, \tag{12}$$

such that $h(f)h(f)' = R(f)$ for all $\psi \in \Psi$. Any choice of ψ and thus of h results in an observationally equivalent model for y_t. The dynamic properties of $\{f_t\}$ following from (8), however, depend on the precise ψ that is chosen. We therefore obtain a sufficient condition for SE if (8) is satisfied for any choice of $\psi \in \Psi$ satisfying the conditions formulated in Lemma 1. This yields the additional infimum in condition (9). A similar complication is absent in the volatility case; compare Blasques et al. (2012) and Harvey (2013).

Fourth, condition (9) simplifies for particular choices of parameterizations and scale functions. For example, if we use the familiar Fisher transformation $\rho(f) = \tanh(f_t)$, the correlation lies in the interval $(-1, +1)$ by construction for all values of $f_t \in \mathbb{R}$. The GAS dynamics automatically adjust to this parameterization via the use of the score in (3). The use of the Fisher transformation has the additional advantage that its derivative is given by $\dot{\rho}(f_t) = 1 - \rho(f_t)^2$. This simplifies the expressions in (9) considerably. If, moreover, we choose a scaling function that is independent of f_t, the entire middle term in (9) vanishes.

To conclude this section, we provide an analytic result on the optimal choice of the function ψ that can be obtained for the special setting of the familiar Fisher transformation $\rho(f) = \tanh(f_t)$ with unit scaling $S(f) \equiv 1$. Using Jensen, triangle and Cauchy-Schwarz inequalities, we then obtain a stricter sufficient condition for SE from (9) as

$$\begin{aligned} \inf_{\psi\in\Psi} \; \mathrm{E}\sup_{f\in\mathcal{F}} \left|\beta + \alpha(1-\rho(f)^2)\dot{g}(\rho(f))k(u_t)\right| \leq \\ |\beta| + |\alpha| \inf_{\psi\in\Psi} \; \mathrm{E}\sup_{f\in\mathcal{F}} \left|(1-\rho(f)^2)\dot{g}(\rho(f))k(u_t)\right| \leq \\ |\beta| + |\alpha| \; \mathrm{E}\|k(u_t)\| \cdot \inf_{\psi\in\Psi} \sup_{f\in\mathcal{F}} \left\|(1-\rho(f)^2)\; \dot{g}(\rho(f))\right\| < 1, \end{aligned} \tag{13}$$

where $\|\cdot\|$ denotes the standard Euclidean norm. Instead of the Cauchy-Schwarz inequality, we could also use a second triangle inequality to obtain the alternative stricter sufficient condition

$$
\begin{aligned}
|\beta| + |\alpha| \ \inf_{\psi\in\Psi} \ \mathrm{E} \sup_{f\in\mathcal{F}} \left|(1-\rho(f)^2)\dot{g}(\rho(f))k(u_t)\right| \leq \\
|\beta| + |\alpha| \ \inf_{\psi\in\Psi} \sum_{i=1}^{3} \sup_{f\in\mathcal{F}} \left|(1-\rho(f)^2) \ \dot{g}_i((\rho(f)))\right| \cdot \mathrm{E}\,|k_i(u_t)| < 1,
\end{aligned}
\tag{14}
$$

where $\dot{g}_i$ and k_i are the ith elements of $\dot{g}$ and k, respectively. Using either of the more stringent SE conditions (13) or (14), we obtain the following result.

Lemma 2. *Under the assumptions stated in Lemma 1, setting $\psi(\rho) = c_\psi \arcsin(\rho)$ with $c_\psi = 1/2$ reaches the functional lower bound for the sufficient condition stated in either equation (13) or (14). The condition then reduces to $|\beta| + |\alpha| \ \mathrm{E}\|k(u_t)\| < 1$ for condition (13) and $|\beta| + |\alpha| \ \mathrm{E}|k_1(u_t)| < 1$ for condition (14), respectively, where $k_1(u_t)$ is the first element of $k(u_t)$. The link function becomes the symmetric matrix root*

$$
h(f) = \begin{pmatrix} \cos(\arcsin(\rho(f))/2) & \sin(\arcsin(\rho(f))/2) \\ \sin(\arcsin(\rho(f))/2) & \cos(\arcsin(\rho(f))/2) \end{pmatrix}.
$$

The result in Lemma 2 shows that we uniformly obtain the largest SE region for the stricter conditions (13) or (14) for the symmetric matrix root h in (7). The choice of h in setting up the dynamic equation (6) is thus far from innocuous and directly influences the size and shape of the SE region. We illustrate our results from this section numerically in the next section for a number of relevant dynamic correlation models.

4 Numerical and empirical results

4.1 Numerical results

Alternative choices for p, S, ρ, and h give rise to different models with different SE regions. For each particular choice we check for every pair (α, β) whether the sufficient condition (9) is satisfied. We plot the SE region in the (α, β)-plane by numerically identifying, for every given α in a grid, the corresponding maximum and minimum values of β that satisfy (9).

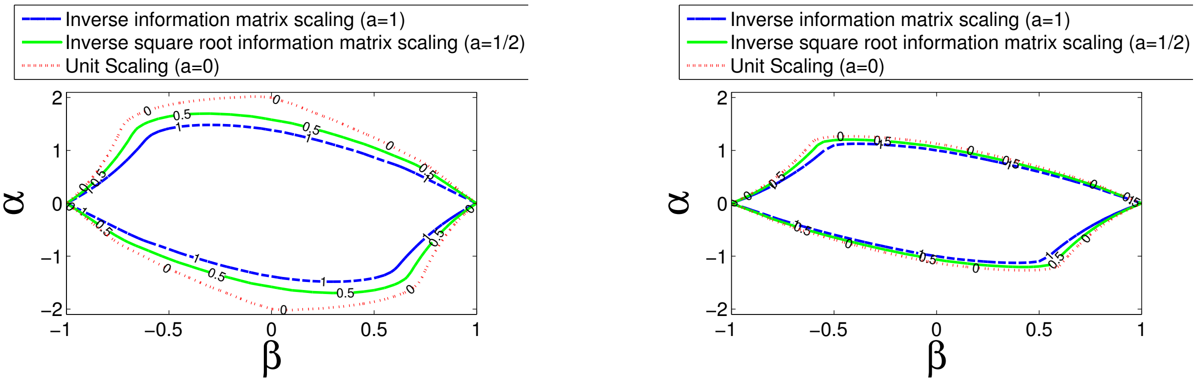
To fix ideas, we consider the class of Student's t densities for u_t as in Creal et al. (2011). The Fisher transformation $\rho(f_t) = \tanh(f_t)$ ensures propers value for the correlation parameter. As indicated in Section 3, this also simplifies the evaluation of the SE condition in Lemma 1. For the scaling function, we adopt the three choices based on the information matrix as presented in equation (5).

Next, we investigate the sensitivity of the SE region to the choice of matrix root $h(\cdot)$. For this, we consider two prominent alternatives, both described by $\psi(\rho) = c_\psi \arcsin(\rho)$ for $c_\psi \in \mathbb{R}$. The first alternative is the symmetric matrix root of Lemma 2 with $c_\psi = 1/2$. The second is the familiar (lower triangular) Choleski decomposition, which is obtained by setting $c_\psi = 1$.

To numerically evaluate the sufficient condition for SE as formulated in (9), we need to solve an optimization problem within an integration procedure. Due to the univariate state equation, the integral can be evaluated via a quadrature rule. Local optima are avoided by evaluating the function over a wide grid and by noting that $(\partial/\partial f)^i s(f, y_t) \to 0$ as $|f| \to \infty$ for all $i > 1$. The numerically cumbersome optimization required by condition (9) for every data generating process can be circumvented by storing maximally positive and negative values of $S(f)\, q(h(f)u_t, f)$ and recycling these for evaluation at different points in the (α, β)-plane. We can further halve the computation time by noting that in our setting $|\partial\phi_t(f;\theta)/\partial f| = |\partial\phi_t(f;-\theta)/\partial f|$.

In the left panel of Figure 1, we present the results for the normal distribution and the symmetric root $h(f)$. The figure contains three different regions, each one corresponding to a different form of scaling in equation (5). Points inside each region are combinations of (α, β) for which the sufficient condition (9) is met. The shape of the sufficient SE region is anti-symmetric around the origin due to the absolute signs in (9). The region also shows a non-monotonic curvature, particularly in the second and fourth quadrant. These are due to the use of absolute values, the change in the location of the supremum in (9), and a shift in the relevant region of integration if the derivative of $S(f)q(h(f)u_t, f)$ changes sign.

An interesting feature in Figure 1 is the behavior of the region for square root inverse information matrix scaling, $a = 1/2$ in (5). First note that $a = 1/2$ has the property that the direct updating via $s(f_t, y_t)$ is invariant with respect to reparametrizations of f_t and with respect to data transformations. Furthermore, the steps $S(f)q(y_t, f)$ in (4) are by construction martingale differences with unit variance; see also Creal et al. (2013). This implies that $\{f_t\}_{t\in\mathbb{N}}$ converges to a covariance stationary process as long as $|\beta| < 1$. The region in Figure 1 shows that $|\beta| < 1$ is necessary, but not sufficient for (9) to be satisfied. This relates directly to discussions in the GARCH literature, where in the univariate setting covariance stationarity is a more restrictive condition than strict stationarity, but



(a) Symmetric root ($c_\psi = 1/2$) (b) Choleski root ($c_\psi = 1$)

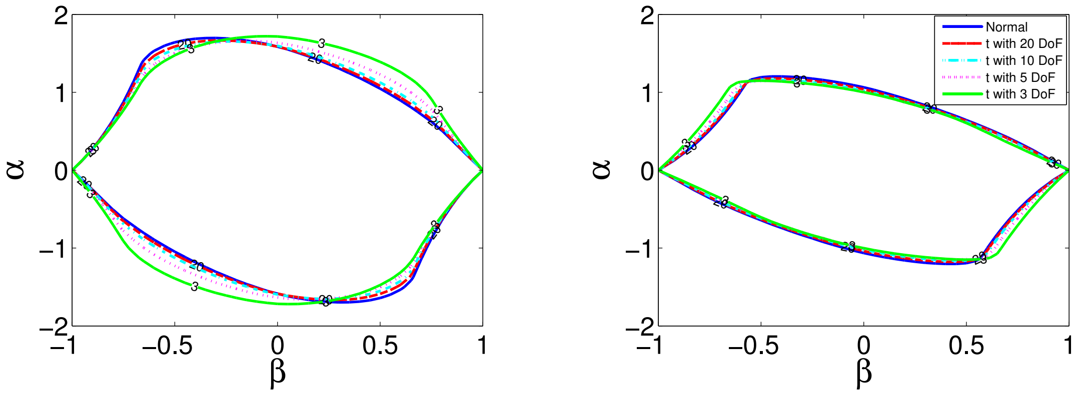
Figure 1: Stationarity and Ergodicity (SE) sufficiency regions for the normal distribution and different scaling choices $S(f) = (\mathcal{I}_t(f))^{-a}$ for $a = 0, 1/2, 1$

Notes: The two panels contain different regions obtained by parametrizing the matrix roots $h(f)$ with $\psi(\rho) = c_\psi \arcsin(\rho)$. The left panel panel contains the results for the symmetric matrix root ($c_\psi = 1/2$) and the right panel corresponds to the Choleski decomposition ($c_\psi = 1$).

the relation between the two remains an open question in a multivariate context; see for example Boussama et al. (2011).

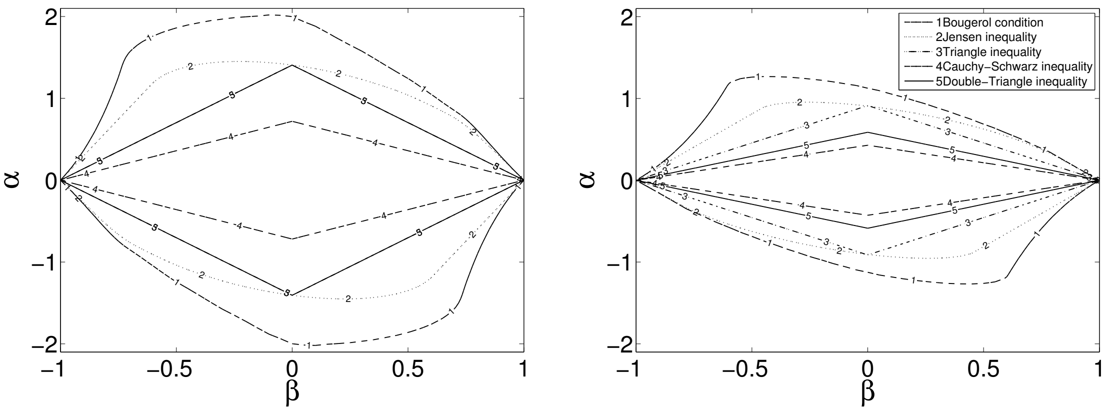
The right hand panel in Figure 1 shows the different SE regions for a different choice of matrix root $h(f)$, namely the Choleski decomposition. It is clear that the sufficiency regions in the (α, β)-plane are smaller than the corresponding regions for the symmetric root. As models constructed with a symmetric root and a Choleski root are observationally equivalent, we can take the larger regions as sufficient regions for SE to hold; see also Lemma 2. The differences make clear that the choice of root is important for determining the size of the region either analytically or numerically.

In Figure 2, we show the results for the Student's t distribution for different degrees of freedom using square root information matrix scaling ($a = 1/2$). We see that the SE region heavily depends on the degrees of freedom parameter. In the empirically relevant first quadrant, for the Choleski decomposition, fatter tails imply that only smaller values of $|\alpha|$ guarantee SE for given values of $|\beta| < 1$. For the symmetric root decomposition, on the other hand, the opposite is true. In Section 4.2, we demonstrate that this difference can crucially affect the applicability of our methodology in the empirically relevant subset of the parameter space. Part of the discrepancy is explained by the fact that changing the degrees of freedom parameter directly influences the magnitude of the steps $S(f)q(y_t, f)$; see also Creal et al. (2011).



(a) Symmetric root ($c_\psi = 1/2$) (b) Choleski root ($c_\psi = 1$)

Figure 2: SE regions for different Student's t degrees of freedom (DoF) and $a = 1/2$



(a) Symmetric root ($c_\psi = 1/2$) (b) Choleski root ($c_\psi = 1$)

Figure 3: Stationarity and Ergodicity (SE) sufficiency regions for the normal distribution using unit scaling ($S(f) \equiv 1$) and the stricter inequalities in equation (13).

To conclude the numerical evaluation of the sufficient conditions, we highlight the results stated in Lemma 2. Figure 3 plots the results for $\psi(\rho) = c_\psi \arcsin(\rho)$. The left panel gives the result for the symmetric matrix root $c_\psi = 1/2$. The right panel is for the Choleski decomposition, $c_\psi = 1$. Each panel presents 5 different regions. The outer region is based on the numerical evaluation of the original condition (9), with the infimum over ψ replaced by the choice $\psi(\rho) = \arcsin(\rho)/2$. The next region is obtained a numerical evaluation of (9) after applying Jensen's inequality, interchanging the expectations and the log operator. The next region follows after applying the triangle inequality, see the second line of equation (13). The final two regions are obtained after applying the Cauchy-Schwarz, or a second triangle inequality; see equations (13) and (14).

First we note that all regions are wider for the symmetric root case compared to their

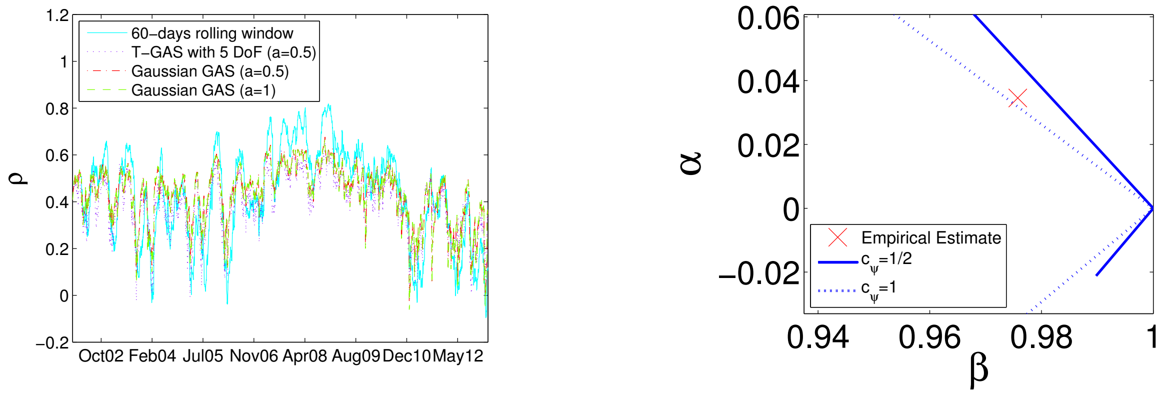
Choleski based counterparts. This is in line with Lemma 2. This not only holds for the stricter contraction conditions in (13) and (14), but also for the less strict original contraction conditions. A second thing to note is that for the symmetric root case the region based on the double triangle inequality coincides with the region based on a single application of the triangle inequality. The same does not hold for the Choleski root. It suggests that the use of the bound in equation (14) might be the easiest and most useful one to apply to empirical models.

4.2 Empirical illustration

In this section we study the time-varying correlation between the London and Athens stock exchange. We take daily returns of the FTSE 100 and the Athex Composite over the period January 1, 2002 to March 2, 2013. We are particularly interested in whether there are indications that the correlation between these two markets changed since the onset of the European sovereign debt crisis. To focus on the correlation part of the model, we first filter both series using the GJR-GARCH model. Using the filtered residuals, we estimate the copula part of the model using Gaussian GAS dynamics as an illustration. The left graph in Figure 4 shows the dynamic correlations for the filtered series along with a 60-day rolling window correlation estimate as a non-parametric benchmark. Though the correlation is generally stable, there are also clear signs of short term memory in the series and periodic swings. Interestingly, the correlation patterns suggest that the Greek sovereign debt crisis, if anything, has reduced the correlation between the two indices.

The estimated GAS correlation model turns out to be highly persistent. The estimated values of α and β are indicated by the cross mark in the right panel of Figure 4. The value of β is very close to 1, which is typical for GAS scale models based on daily data. The figure also shows the SE region boundaries based on the Choleski and the symmetric root specification. Clearly, we confirm the importance of the choice of ψ in verifying the SE properties. The symmetric root based region can be used to show the SE nature of the data. In contrast, using the Choleski based region one would fail to obtain this result. Again we stress that all of these regions are only based on sufficient conditions, and that the actual regions may be wider. The empirical results do illustrate, however, that the stationarity and ergodicity properties of score driven correlation models can be studied also for real data.

To further corroborate the sensitivity of the SE region for alternative empirical models,



(a) Correlation paths (b) Gaussian estimates ($a = 1$)

Figure 4: Empirical estimation

Notes: The left panel displays filtered correlations between the FTSE 100 (UK) and Athex Composite (Greece) equity index returns. The right panel puts the empirical estimates obtained by unconstrained estimation into the to the zoomed in SE region perspective.

Table 1: Estimation results

	Gaussian			Student's t with 5 DoF		
	$a = 0$	$a = 0.5$	$a = 1$	$a = 0$	$a = 0.5$	$a = 1$
ω	0.011	0.011	0.011	0.010	0.009	0.009
	(0.014)	(0.018)	(0.014)	(0.005)	(0.004)	(0.005)
α	0.027	0.031	0.035	0.041	0.038	0.036
	(0.015)	(0.022)	(0.018)	(0.010)	(0.008)	(0.008)
β	0.977	0.976	0.976	0.966	0.967	0.968
	(0.028)	(0.038)	(0.028)	(0.014)	(0.011)	(0.014)
Log-likelihood	−7904.1	−7903.8	−7903.7	−7861.4	−7861.5	−7861.7
Inside SE region?						
. . . For $c_\psi = 1$	yes	yes	no	yes	yes	no
. . . For $c_\psi = 1/2$	yes	yes	yes	yes	yes	yes

we estimated a range of GAS models for the filtered equity returns. The models different in their choice of the scaling function for the score, i.e., the choice of a, and in the choice of the error distribution. We consider $a \in \{0, 0.5, 1\}$ and a Gaussian and a Student's t distribution with 5 degrees of freedom. The results are presented in Table 1. For all models, the estimate of β is large and close to one, implying that there is strong persistence in the correlation dynamics at the daily frequency.

In the current context with all models having the same number of parameters, the log-likelihood value can be interpreted as information theoretic model selection criteria, see Akaike (1974). Switching to a Student's t specification with 5 degrees of freedom, the likelihood increases by more than 40 points, which indicates a strong preference for a fat-tailed specification. The differences between the models within a distributional class are small and the choice of information matrix scaling does not change the qualitative nature of the paths.

The two rows at the bottom of the table indicate whether the estimated parameters lie inside the SE region as computed using the techniques in Section 4.1. The results are identical for the Gaussian and the Student's t model. The time series processes implied by the estimated empirical models are stationary and ergodic for unit ($a = 0$) and inverse square root information matrix ($a = 0.5$) scaling. This holds irrespective of whether the symmetric root ($c_\psi = 1/2$) or Choleski root ($c_\psi = 1$) is used to verify the average contraction condition. For inverse information matrix scaling, however, we cannot ensure SE properties for the Choleski root, while we can for the symmetric root; see also the right hand panel in Figure 4. This again highlights that the use of different constructive devices such as different matrix roots is also relevant empirically for the verification of sufficient conditions for SE in a multivariate setting.

5 Concluding Remarks

We have derived sufficient regions for stationarity and ergodicity for a new class of score driven dynamic correlation models. The regions exhibit a highly non-standard shape. Moreover, we showed that the shape and size of the SE regions depends on the type of matrix root that is chosen in checking the sufficient conditions of Bougerol (1993). The numerical results were supported by an empirical investigation of the time varying correlation between UK and Greek stock markets.

There are a number of possible interesting extensions of our current results. First, one can try to generalize the results to the fully multivariate (rather than bivariate) setting of score driven correlation models proposed in Creal et al. (2011). The challenge here is to limit the number of parameters describing the dynamics of time-varying volatilities and correlations. Second, one can try to relax the uniform bounds on the average contraction propery. Third, one can try to apply the current stationarity and ergodicity results in a proof of the asymptotic properties of the maximum likleihood estimator for GAS correlation models, such as consistency and asymptotic normality. This would require us to also establish the existence of nonlinear moments of f_t. We leave this for future work.

References

Aielli, G. P. (2013). Dynamic conditional correlation: on properties and estimation. *Journal of Business & Economic Statistics* (forthcoming).

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19*(6).

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics 21*(1), 79–109.

Blasques, F., S. J. Koopman, and A. Lucas (2012). Stationarity and ergodicity of univariate Generalized Autoregressive Score processes. *Tinbergen Institute Discussion Papers 12-059/4*.

Boudt, K., J. Danielsson, and S. Laurent (2013). Robust forecasting of dynamic conditional correlation garch models. *International Journal of Forecasting 29*(2), 244–257.

Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization 31*(4), 942–959.

Boussama, F., F. Fuchs, and R. Stelzer (2011). Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and their Applications 121*(10), 2331–2360.

Chamberlain, G. (1983). A characterization of the distributions that imply mean–variance utility functions. *Journal of Economic Theory 29*(1), 185–201.

Creal, D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics 29*(4), 552–563.

Creal, D., S. J. Koopman, and A. Lucas (2013a). General autoregressive score models with applications. *Journal of Applied Econometrics*, forthcoming.

Creal, D., S. J. Koopman, A. Lucas, and B. Schwaab (2013b). Observation-driven mixed measurement dynamic factor models. *Review of Economics and Statistics*, forthcoming.

Diaconis, P. and D. Freedman (1999). Iterated random functions. *SIAM review*, 45–76.

Engle, R. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics 17*(5), 425–446.

Engle, R. F. and B. Kelly (2012). Dynamic equicorrelation. *Journal of Business and Economic Statistics 30*, 212–228.

Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized arch. *Econometric theory 11*(01), 122–150.

Francq, C. and J. Zakoian (2011). *GARCH models: structure, statistical inference and financial applications*. Wiley.

Hamada, M. and E. A. Valdez (2008). Capm and option pricing with elliptically contoured distributions. *Journal of Risk and Insurance 75*(2), 387–409.

Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails*. Cambridge University Press.

Koopman, S. J., A. Lucas, and M. Scharth (2012). Predicting time-varying parameters with parameter-driven and observation-driven models. *Tinbergen Institute Discussion Papers 12-020/4*.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH (1, 1) model. *Econometric theory 6*(03), 318–334.

Oh, D. H. and A. J. Patton (2012). Modelling dependence in high dimensions with factor copulas. *Manuscript, Duke University*.

Owen, J. and R. Rabinovitch (1983). On the class of elliptical distributions and their applications to the theory of portfolio choice. *The Journal of Finance 38*(3), 745–752.

Patton, A. J. (2009). Copula–based models for financial time series. *Handbook of Financial Time Series*, 767–785.

Pelletier, D. (2006). Regime switching for dynamic correlations. *Journal of Econometrics 131*(1), 445–473.

Silvennoinen, A. and T. Teräsvirta (2009). Multivariate GARCH models. In *Handbook of Financial Time Series*, pp. 201–229. Springer.

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

Wu, W. and X. Shao (2004). Limit theorems for iterated random functions. *Journal of Applied Probability 41*(2), 425–436.

A Proofs

We first state Theorem 3.1 of Bougerol (1993). Denote by $\log \Lambda(\phi_0)$ the term inside the expectation on the left hand side of (8).

Theorem 1 (Bougerol (1993, Theorem 3.1))**.** *Let $\{\phi_t\}$ be a stationary and ergodic sequence of endomorphic Lipschitz maps. Assume*

1. *There exists a $f \in \mathcal{F}$ and distrance measure d such that* $\mathrm{E}[\log^+ d(\phi_0(f), f)] < \infty$*;*
2. $\mathrm{E}[\log^+ \Lambda(\phi_0)] < \infty$;
3. $\mathrm{E}[\log \Lambda(\phi_0^{(r)})] < 0$, *where $\phi_0^{(r)}$ denotes the r-fold backward iterates.*

Then the stochastic recurrence equation (6) admits a stationary ergodic solution $\{f_t\}$.

Proof of Lemma 1:

The SE property of $\{f_t\}$ follows from the measurability with respect to $\{u_t\}$. The Lipschitz property is obtained from the boundedness of the terms in equation (A6) below. Condition 1 is then ensured by the definition of the GAS transition equation and the assumed moments in Lemma 1, as we can write $\mathrm{E}[\log^+ d(\phi_0(f), f)] \leq \mathrm{E}|\phi_0(f) - f| = \mathrm{E}|\omega + (\beta - 1)f + \alpha S(f)q(h(f)u_0, f)| \leq |\omega| + |(\beta - 1)f| + \alpha \mathrm{E}|S(f)q(h(f)u_0, f)|$. As requirement 2 is implied by 3, we can now turn our main interest towards the study of the latter, non-trivial, condition 3.

We write ϕ and ψ for $\phi(\rho)$ and $\psi(\rho)$, respectively. Define the shorthand notation $c_w = c_w(\rho) = \cos(w(\rho))$ with $w : [-1, 1] \to \mathbb{R}$, and similarly $s_w = s_w(\rho) = \sin(w(\rho))$. Each matrix root h of the correlation matrix can be written as

$$h(f) = \begin{pmatrix} c_\phi(\rho(f)) & s_\phi(\rho(f)) \\ s_\psi(\rho(f)) & c_\psi(\rho(f)) \end{pmatrix}. \tag{A1}$$

Using (A1), we obtain

$$\begin{pmatrix} c_\phi & s_\phi \\ s_\psi & c_\psi \end{pmatrix} \begin{pmatrix} c_\phi & s_\psi \\ s_\phi & c_\psi \end{pmatrix} = \begin{pmatrix} 1 & s_{\phi+\psi} \\ s_{\phi+\psi} & 1 \end{pmatrix},$$

such that we require $\sin(\psi + \phi) = \rho$ or $\phi(\rho) = \arcsin(\rho) - \psi(\rho)$ for some arbitrary function $\psi(\rho)$.

It follows that $s_\phi = \rho c_\psi - \sqrt{1-\rho^2} s_\psi$, and $c_\phi = \sqrt{1-\rho^2} c_\psi + \rho s_\psi$. From this we obtain

$$
\begin{aligned}
&\begin{pmatrix} c_\phi & s_\psi \\ s_\phi & c_\psi \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_\phi & s_\phi \\ s_\psi & c_\psi \end{pmatrix} = \\
&\qquad \begin{pmatrix} 2c_\phi s_\psi & s_\phi s_\psi + c_\phi c_\psi \\ s_\phi s_\psi + c_\phi c_\psi & 2s_\phi c_\psi \end{pmatrix} = \begin{pmatrix} 2c_\phi s_\psi & c_{\phi-\psi} \\ c_{\phi-\psi} & 2s_\phi c_\psi \end{pmatrix} = \\
&\qquad \begin{pmatrix} -\rho c_{2\psi} + \sqrt{1-\rho^2} s_{2\psi} + \rho & \sqrt{1-\rho^2} c_{2\psi} + \rho s_{2\psi} \\ \sqrt{1-\rho^2} c_{2\psi} + \rho s_{2\psi} & \rho c_{2\psi} - \sqrt{1-\rho^2} s_{2\psi} + \rho \end{pmatrix} =: H(\rho) + \rho \mathrm{I}.
\end{aligned}
$$

Using $y_t = h(f)u_t$, we can rewrite (4) as

$$
\begin{aligned}
&(1-\rho(f)^2) q(h(f)u_t, f)/\dot{\rho}(f) = \\
&\qquad 2\dot{p}(u_t'u_t)\rho(f)u_t'u_t - \dot{p}(u_t'u_t)u_t' \left(H(\rho(f)) + \rho(f)\mathrm{I}\right) u_t + \rho(f) = \\
&\qquad \dot{p}(u_t'u_t)\rho(f)u_t'u_t - \dot{p}(u_t'u_t)u_t' H(\rho(f)) u_t + \rho(f) = \\
&\qquad \rho(f)\left(\dot{p}(u_t'u_t)u_t'u_t + 1\right) - \dot{p}(u_t'u_t)u_t' H(\rho(f)) u_t = g(\rho)k(u_t),
\end{aligned} \tag{A2}
$$

with

$$
\begin{aligned}
g(\rho) &= \left(\rho \ , \ \rho c_{2\psi} - \sqrt{1-\rho^2} s_{2\psi} \ , \ \sqrt{1-\rho^2} c_{2\psi} + \rho s_{2\psi}\right), \\
k(u_t) &= \left(\dot{p}(u_t'u_t)u_t'u_t + 1 \ , \ \dot{p}(u_t'u_t)(u_{1,t}^2 - u_{2,t}^2) \ , \ -2\dot{p}(u_t'u_t)u_{1,t}u_{2,t}\right)',
\end{aligned}
$$

and $u_t = (u_{1,t}, u_{2,t})'$. Defining $\dot{g}(\rho) = \partial g(\rho)/\rho$ as the derivative of $g(\rho)$, it holds that

$$
\begin{aligned}
\dot{g}(\rho) \ &= \ \left(1 \ , \ c_{2\psi}(\rho) + \rho \cdot (1-\rho^2)^{-1/2} s_{2\psi}(\rho) \ , \ -\rho \cdot (1-\rho^2)^{-1/2} c_{2\psi}(\rho) + s_{2\psi}(\rho)\right) + \\
&\qquad 2\dot{\psi}(\rho)\left(0 \ , \ -\rho s_{2\psi}(\rho) - \sqrt{1-\rho^2} c_{2\psi}(\rho) \ , \ -\sqrt{1-\rho^2} s_{2\psi}(\rho) + \rho c_{2\psi}(\rho)\right), \qquad \text{(A3)} \\
&= \ (1,0,0) + \left((1-\rho^2)^{-1/2} - 2\dot{\psi}(\rho)\right) \cdot \\
&\qquad \left(0 \ , \ \sqrt{1-\rho^2} c_{2\psi}(\rho) + \rho \cdot s_{2\psi}(\rho) \ , \ \sqrt{1-\rho^2} s_{2\psi}(\rho) - \rho \cdot c_{2\psi}(\rho)\right). \qquad \text{(A4)}
\end{aligned}
$$

The definitions in (7) and (A2) then imply that (8) can be written as

$$
\frac{\partial \phi_t(f;\theta)}{\partial f} = \beta + \alpha \left(\frac{\partial}{\partial f}\left(\frac{S(f)\ \dot{\rho}(f)}{1-\rho(f)^2}\right)\right) g(\rho(f))k(u_t) + \alpha \frac{S(f)\ \dot{\rho}(f)}{1-\rho(f)^2}\frac{\partial g(\rho(f))}{\partial f} k(u_t). \tag{A5}
$$

■

Proof of Lemma 2: Using (A4), we can rewrite $\|\dot{g}(\rho)\|^2$ as

$$1+\left|\frac{1}{\sqrt{1-\rho^2}}-2\dot{\psi}(\rho)\right|^2\cdot\left(c_{2\psi}(\rho)^2+s_{2\psi}(\rho)^2\right)=1+\left|\frac{1}{\sqrt{1-\rho^2}}-2\dot{\psi}(\rho)\right|^2\geq 1. \tag{A6}$$

For $\psi(\rho)=\arcsin(\rho)/2$ the second term vanishes and we obtain the functional lower bound $(1-\rho^2)\cdot\|\dot{g}(\rho)\|^2=1-\rho^2$, which reaches its supremum of 1 at $\rho=0$. The rest of the result follows directly from the definition of $\phi(\rho)=\arcsin(\rho)-\psi(\rho)=\arcsin(\rho)/2$. For computational reasons, it may be useful to note that

$$\|k(u_t)\|^2=2\dot{p}(u_t'u_t)^2(u_t'u_t)^2+2\dot{p}(u_t'u_t)(u_t'u_t)+1.$$

which only depends on the quadratic form $u_t'u_t$.

An analogous line of reasoning holds for condition (14) based on the double triangle inequality.

■