TI 2020-035/III
Tinbergen Institute Discussion Paper

The Jacobian of the exponential function

Jan R. Magnus[1]
Henk G.J. Pijls[2]
Enrique Sentana[3]

[1] Department of Econometrics and Data Science, Vrije Universiteit Amsterdam and Tinbergen Institute,
[2] Korteweg-de Vries Institute for Mathematics, University of Amsterdam
[3] CEMFI, Madrid

The Jacobian of the exponential function

June 16, 2020

Jan R. Magnus
Department of Econometrics and Data Science,
Vrije Universiteit Amsterdam and Tinbergen Institute

Henk G. J. Pijls
Korteweg-de Vries Institute for Mathematics, University of Amsterdam

Enrique Sentana
CEMFI

Abstract: We derive closed-form expressions for the Jacobian of the matrix exponential function for both diagonalizable and defective matrices. The results are applied to two cases of interest in macroeconometrics: a continuous-time macro model and the parametrization of rotation matrices governing impulse response functions in structural vector autoregressions.

JEL Classification: C65, C32, C63.

Keywords: Matrix differential calculus, Orthogonal matrix, Continuous-time Markov chain, Ornstein-Uhlenbeck process.

Corresponding author: Enrique Sentana, CEMFI, Casado del Alisal 5, 28014 Madrid, Spain. E-mail: `sentana@cemfi.es`

Declarations of interest: None.

1 Introduction

The exponential function e^x is one of the most important functions in mathematics. Its history goes back to the brothers Jacob and Johann Bernoulli in the late 17th century, while the matrix exponential e^X was not introduced until the late 19th century by Sylvester, Laguerre, and Peano.

The matrix exponential plays an important role in the solution of systems of ordinary differential equations (Bellman, 1970), multivariate Ornstein-Uhlenbeck processes (Bergstrom, 1984 and Section 8 below), and continuous-time Markov chains defined over a discrete state space (Cerdà-Alabern, 2013). The matrix exponential is also used in modelling positive definiteness (Linton, 1993; Kawakatsu, 2006) and orthogonality (Section 9 below), as e^X is positive definite when X is symmetric and orthogonal when X is skew-symmetric.

The derivative of e^x is the function itself, but this is no longer true for the matrix exponential (unless the matrix is diagonal). We can obtain the derivative (Jacobian) directly from the power series, or as a block of the exponential in an augmented matrix, or as an integral. We shall review these three approaches, but they all involve either infinite sums or integrals, and the numerical methods required for computing the Jacobian are not trivial (Chen and Zadrozny, 2001; Tsai and Chan, 2003; Fung, 2004).

The purpose of this paper is to provide a closed-form expression which is easy to compute, is applicable to both defective and nondefective real matrices, and has no restrictions on the number of parameters that characterize X.

We have organized the paper as follows. In Section 2 we discuss and review the matrix exponential function. Three expressions for its Jacobian (Propositions 1–3) are presented in Section 3 together with some background and history. Our main result is Theorem 1 in Section 4. In Sections 5 and 6 we apply the theorem to defective and nondefective matrices and discuss structural restrictions such as symmetry and skew-symmetry. In Section 7 we derive the Hessian matrix (Proposition 4). Two applications in macroeconometrics demonstrate the usefulness of our results: a continuous-time multivariate Ornstein-Uhlenbeck process for stock variables observed at equidistant points in time (Section 8) and a structural vector autoregression with non-Gaussian shocks (Section 9). In both cases, we explain how to use our main result to obtain the loglikelihood scores and information matrix in closed form. Section 10 concludes. There are two appendices. Appendix A provides proofs of the four propositions and Appendix B provides the proof of the theorem in three lemmas. As a byproduct of the proof, Lemma 2 presents an alternative expression for the characteristic (and moment-generating) function of the beta distribution, which is valid for integer values of its two shape parameters.

2 The exponential function

Let A be a real matrix of order $n \times n$. The exponential function, denoted by $\exp(A)$ or e^A, is defined as

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} = I_n + \sum_{k=0}^{\infty} \frac{A^{k+1}}{(k+1)!}, \tag{1}$$

and it exists for all A because the norm of a finite-dimensional matrix is finite so that the infinite sum converges absolutely. We mention two well-known properties. First, we have

$$e^{(A+B)t} = e^{At}e^{Bt} \text{ for all } t \iff A \text{ and } B \text{ commute,}$$

so that $e^{A+B} = e^A e^B$ when A and B commute, but not in general. Second, as a special case, we have $e^{A(s+t)} = e^{As}e^{At}$, and hence, upon setting $s = -t$,

$$e^{-At}e^{At} = I_n,$$

so that e^{At} is nonsingular and its inverse is e^{-At}.

Let us introduce the $n \times n$ 'shift' matrix

$$E_n = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix},$$

which is nilpotent of index n, that is $E_n^n = 0$, and has various other properties of interest; see Abadir and Magnus (2005, Section 7.5). The Jordan decomposition theorem states that there exists a nonsingular matrix T such that $T^{-1}AT = J$, where

$$J = \operatorname{diag}(J_1, \dots, J_m), \qquad J_i = \lambda_i I_{n_i} + E_{n_i}. \tag{2}$$

The matrix J thus contains m Jordan blocks J_i, where the λ's need not be distinct and $n_1 + \dots + n_m = n$. Since I_n and E_n commute, we have

$$\exp(J_i) = \exp(\lambda_i I_{n_i}) \exp(E_{n_i}) = e^{\lambda_i} \sum_{k=0}^{n_i - 1} \frac{1}{k!} E_{n_i}^k \tag{3}$$

and

$$e^A = Te^J T^{-1}, \qquad e^J = \operatorname{diag}(e^{J_1}, \dots, e^{J_m}). \tag{4}$$

3 First differential

We are interested in the derivative of $F(X) = \exp(X)$. The simplest case is $X(t) = At$, where t is a scalar and A is a matrix of constants. Then,

$$de^{At} = Ae^{At}\,dt = e^{At}A\,dt, \tag{5}$$

as can be verified directly from the definition.

The general case is less trivial. Without making any assumptions about the structure of X, the differential of X^{k+1} is

$$dX^{k+1} = (dX)X^k + X(dX)X^{k-1} + \cdots + X^k(dX),$$

and hence the differential of F is

$$dF = \sum_{k=0}^{\infty} \frac{dX^{k+1}}{(k+1)!} = \sum_{k=0}^{\infty} \frac{C_{k+1}}{(k+1)!}, \qquad C_{k+1} = \sum_{j=0}^{k} X^j(dX)X^{k-j};$$

see Magnus and Neudecker (2019, Miscellaneous Exercise 8.9, p. 188). To obtain the Jacobian we vectorize F and X. This gives

$$d\,\mathrm{vec}\,F = \sum_{k=0}^{\infty} \frac{1}{(k+1)!}\,\mathrm{vec}\,C_{k+1} = \sum_{k=0}^{\infty} \frac{\nabla_{k+1}(X)}{(k+1)!}\,d\,\mathrm{vec}\,X.$$

Thus, we have proved the following result.

Proposition 1. *The Jacobian of the exponential function $F(X) = \exp(X)$ is given by*

$$\nabla(X) = \frac{\partial\,\mathrm{vec}\,F}{\partial(\mathrm{vec}\,X)'} = \sum_{k=0}^{\infty} \frac{\nabla_{k+1}(X)}{(k+1)!},$$

where

$$\nabla_{k+1}(X) = \sum_{j=0}^{k} \left((X')^{k-j} \otimes X^j\right).$$

The Jacobian can also be obtained as the appropriate submatrix of an augmented matrix, following ideas in Van Loan (1978, pp. 395–396). Since

$$\begin{pmatrix} A & C \\ 0 & B \end{pmatrix}^{k+1} = \begin{pmatrix} A^{k+1} & \Gamma_{k+1} \\ 0 & B^{k+1} \end{pmatrix}, \qquad \Gamma_{k+1} = \sum_{j=0}^{k} A^j C B^{k-j},$$

we obtain

$$\exp\begin{pmatrix} A & C \\ 0 & B \end{pmatrix} = \begin{pmatrix} e^A & \Gamma \\ 0 & e^B \end{pmatrix}, \qquad \Gamma = \sum_{k=0}^{\infty} \frac{\Gamma_{k+1}}{(k+1)!}, \tag{6}$$

which holds for any square matrices A, B, and C of the same order.

Proposition 2. *We have*

$$\exp\begin{pmatrix} X & dX \\ 0 & X \end{pmatrix} = \begin{pmatrix} e^X & de^X \\ 0 & e^X \end{pmatrix}$$

and

$$\exp\begin{pmatrix} X' \otimes I_n & I_n \otimes I_n \\ 0 & I_n \otimes X \end{pmatrix} = \begin{pmatrix} (e^X)' \otimes I_n & \nabla(X) \\ 0 & I_n \otimes e^X \end{pmatrix}.$$

The two results are obtained by appropriate choices of A, B, and C in (6). For the first equation we choose $A = B = X$ and $C = dX$, and use fact that

$$\Gamma = \sum_{k=0}^{\infty} \frac{C_{k+1}}{(k+1)!} = de^X;$$

see Mathias (1996, Theorem 2.1). The result holds, in fact, much more generally; see Naifeld and Havel (1995). For the second equation we choose $A = X' \otimes I_n$, $B = I_n \otimes X$, and $C = I_n \otimes I_n$; see Chen and Zadrozny (2001, Eq. 2.6). The second equation provides the Jacobian as the appropriate submatrix of the augmented exponential. In contrast, the first equation provides matrices of partial derivatives. Letting $X = X(t)$, the partial derivatives of $\exp(X(t))$ can thus be found from

$$\exp\begin{pmatrix} X & \partial X(t)/\partial t_i \\ 0 & X \end{pmatrix} = \begin{pmatrix} e^X & \partial e^X(t)/\partial t_i \\ 0 & e^X \end{pmatrix}. \tag{7}$$

The somewhat trivial result (5) has a direct consequence which is rather less trivial. Differentiating $F(t) = e^{(A+B)t} - e^{At}$ gives

$$dF(t) = (A+B)e^{(A+B)t}\,dt - Ae^{At}\,dt = AF(t)\,dt + Be^{(A+B)t}\,dt,$$

and hence

$$d\left(e^{-At}F(t)\right) = -Ae^{-At}F(t)\,dt + e^{-At}\,dF(t) = e^{-At}Be^{(A+B)t}\,dt.$$

This leads to

$$e^{(A+B)t} - e^{At} = \int_0^t e^{A(t-s)}Be^{(A+B)s}\,ds, \tag{8}$$

and hence to our third representation.

Proposition 3. *We have*

$$\begin{aligned}\nabla(X) = \frac{\partial \operatorname{vec} F}{\partial(\operatorname{vec} X)'} &= (I_n \otimes e^X) \int_0^1 (e^{Xs})' \otimes e^{-Xs} \, ds \\ &= (I_n \otimes e^X) \int_0^1 e^{(X' \otimes I_n - I_n \otimes X)s} \, ds \\ &= (I_n \otimes e^X) \sum_{k=0}^{\infty} \frac{1}{(k+1)!} (X' \otimes I_n - I_n \otimes X)^k.\end{aligned}$$

The first equality has been known for a long time, at least since Karplus and Schwinger (1948); see also Snider (1964), Wilcox (1967), and Bellman (1970, p. 175). The third equality provides a link with the corresponding formula for Lie algebras; see Tuynman (1995) and Hall (2015, Theorem 5.4), among others.

4 Main result

Propositions 1–3 provide expressions for the Jacobian of $F(X) = e^X$, but their computation involves integrals or infinite sums. We now formulate the Jacobian in a more transparent form which is easy to compute and does not involve infinite sums or integrals. This is our main result.

Theorem 1. *Let $X = TJT^{-1}$ be expressed in Jordan form. The Jacobian of the exponential function $F(X) = \exp(X)$ is*

$$\nabla(X) = \frac{\partial \operatorname{vec} F}{\partial(\operatorname{vec} X)'} = S \Delta S^{-1},$$

where

$$S = (T')^{-1} \otimes T, \qquad \Delta = \operatorname{diag}(\Delta_{11}, \Delta_{12}, \ldots, \Delta_{mm}),$$

and

$$\Delta_{uv} = \sum_{t=0}^{n_u - 1} \sum_{s=0}^{n_v - 1} \theta_{ts} (E'_{n_u})^t \otimes E^s_{n_v}.$$

The coefficients θ_{ts} take the form

$$\theta_{ts} = \begin{cases} \dfrac{e^{\lambda_v}}{(s+t+1)!} & (\lambda_u = \lambda_v), \\ \dfrac{e^{\lambda_v}}{(s+t+1)!} \displaystyle\sum_{i=0}^{t} \alpha_i(s,t) R_{s+i+1} & (\lambda_u \neq \lambda_v), \end{cases}$$

where

$$\alpha_i(s,t) = (-1)^i \binom{s+i}{i}\binom{s+t+1}{t-i}, \qquad R_{n+1}(w) = \frac{e^w - \sum_{j=0}^{n} w^j/j!}{w^{n+1}/(n+1)!},$$

and $w = \lambda_u - \lambda_v$*. The* α_i *are the coefficients of a (Gauss) hypergeometric function and satisfy* $\sum_{i=0}^{t} \alpha_i = 1$*.*

Remark 1. When w approaches zero, then R_{n+1} approaches one, which can be seen by writing R_{n+1} as

$$R_{n+1}(w) = 1 + \frac{w}{(n+2)} + \frac{w^2}{(n+2)(n+3)} + \cdots .$$

So the derivative is continuous at $w = 0$. This representation also shows that $R_{n+1}(w) = M(1, n+2, w)$ where M denotes Kummer's confluent hypergeometric function, and relates R_{n+1} to the incomplete gamma function (see also Lemma 1 in Appendix B).

Remark 2. We can compute R_{n+1} either from its definition, or from the power series under Remark 1 (when w is close to zero), or from the recursion

$$R_{n+1} = \frac{(n+1)(R_n - 1)}{w}, \qquad R_1 = \frac{e^w - 1}{w}.$$

Remark 3. In the definition of $S = (T')^{-1} \otimes T$, we require the ordinary transpose and *not* the complex conjugate. The rule $\operatorname{vec} ABC = (C' \otimes A) \operatorname{vec} B$ also holds for complex matrices and should not be replaced by $\operatorname{vec} ABC = (C^* \otimes A) \operatorname{vec} B$. This is because the rule reflects a rearrangement of the elements rather than a matrix product.

Remark 4. There are m Jordan blocks $J_1, \dots, J_m$, and we have to consider each pair (J_u, J_v). To illustrate, we present the case where both J_u and J_v have dimension 2 ($n_u = n_v = 2$) assuming that $w = \lambda_u - \lambda_v \neq 0$. This gives

$$\Delta_{uv} = \begin{pmatrix} \theta_{00} & \theta_{01} & 0 & 0 \\ 0 & \theta_{00} & 0 & 0 \\ \theta_{10} & \theta_{11} & \theta_{00} & \theta_{01} \\ 0 & \theta_{10} & 0 & \theta_{00} \end{pmatrix}$$

and

$$\theta_{00} = e^{\lambda_v} R_1(w), \qquad \theta_{01} = e^{\lambda_v} R_2(w)/2,$$
$$\theta_{10} = e^{\lambda_v}\left(2R_1(w) - R_2(w)\right)/2, \qquad \theta_{11} = e^{\lambda_v}\left(3R_2(w) - 2R_3(w)\right)/6.$$

Remark 5. Some concepts in matrix algebra (rank, dimension of a Jordan block) are integer-valued and therefore discontinuous. Since our theorem involves the Jordan decomposition, one may wonder whether the decomposition affects the continuity and differentiability of the exponential function, and whether the Jacobian is continuous at singularities where the composition of Jordan blocks changes. A simple example suffices to justify our procedure. Let A_ϵ be a 2×2 matrix, which can be diagonalized when $\epsilon \neq 0$ but not when $\epsilon = 0$. Specifically we have, for $\epsilon \neq 0$,

$$A_\epsilon = \begin{pmatrix} \epsilon & 0 \\ 1 & 0 \end{pmatrix} = T_\epsilon J_\epsilon T_\epsilon^{-1} = \begin{pmatrix} 0 & \epsilon \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \epsilon \end{pmatrix} \begin{pmatrix} -1/\epsilon & 1 \\ 1/\epsilon & 0 \end{pmatrix},$$

whose exponential is given by

$$e^{A_\epsilon} = T_\epsilon e^{J_\epsilon} T_\epsilon^{-1} = \begin{pmatrix} 0 & \epsilon \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^\epsilon \end{pmatrix} \begin{pmatrix} -1/\epsilon & 1 \\ 1/\epsilon & 0 \end{pmatrix} = \begin{pmatrix} e^\epsilon & 0 \\ (e^\epsilon - 1)/\epsilon & 1 \end{pmatrix}.$$

For $\epsilon = 0$ the matrix A_0 cannot be diagonalized and its Jordan decomposition is

$$A_0 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = T_0 J_0 T_0^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

with exponential

$$e^{A_0} = T_0 e^{J_0} T_0^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

We see that T_ϵ does not converge to T_0, that J_ϵ does not converge to J_0, and that e^{J_ϵ} does not converge to e^{J_0}. However, e^{A_ϵ} *does* converge to e^{A_0}, and $d\exp(A_\epsilon)$ *does* converge to $d\exp(A_0)$, which can be verified using Theorem 1.

To see what happens, define

$$S_\epsilon = (T'_\epsilon)^{-1} \otimes T_\epsilon, \qquad S_\epsilon^{-1} = T'_\epsilon \otimes T_\epsilon^{-1},$$

so that

$$S_\epsilon \left(e^{J'_\epsilon(1-s)} \otimes e^{J_\epsilon s} \right) S_\epsilon^{-1} = e^{A'_\epsilon(1-s)} \otimes e^{A_\epsilon s}. \tag{9}$$

Although S_ϵ and S_ϵ^{-1} have a singularity at $\epsilon = 0$, the left-hand side of (9) is regular near $\epsilon = 0$ because the right-hand side is regular. If we integrate it from 0 to 1 we obtain $d\exp(A_\epsilon)$ (Proposition 3), which is therefore also regular near $\epsilon = 0$. Then taking the limit for $\epsilon \to 0$ and interchanging limit and integral we see that $d\exp(A_\epsilon)$ converges to $d\exp(A_0)$.

The function exp is infinitely many times differentiable because each element is a power series in n^2 variables. The matrices T_ϵ, J_ϵ, e^{J_ϵ}, and S_ϵ have a singularity at $\epsilon = 0$, but the singularity in the left-hand side of (9) is removable, so that there are no discontinuities in the Jacobian and Theorem 1 is valid in the neighborhood of singularities.

5 Defective and nondefective matrices

An $n \times n$ matrix is *defective* if and only if it does not have n linearly independent eigenvectors, and is therefore not diagonalizable. A defective matrix always has fewer than n distinct eigenvalues. A real $n \times n$ matrix is *normal* if and only if $X'X = XX'$. A normal matrix is necessarily nondefective because it is diagonalizable. But a nonnormal matrix can be either defective or nondefective, as can be seen from the matrices

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}.$$

Neither A nor B is normal, but A is defective while B is not.

For nondefective (and in particular normal) matrices we have the following corollary to Theorem 1.

Corollary 1. *In the special case where X is nondefective, there exists a nonsingular matrix T such that $T^{-1}XT = \Lambda$, where $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. The Jacobian of the exponential function $F(X) = \exp(X)$ is then*

$$\nabla(X) = \frac{\partial \operatorname{vec} F}{\partial (\operatorname{vec} X)'} = S \Delta S^{-1},$$

where

$$S = (T')^{-1} \otimes T, \qquad \Delta = \operatorname{diag}(\delta_{11}, \delta_{12}, \dots, \delta_{nn}),$$

and

$$\delta_{ij} = \begin{cases} e^{\lambda_i} & (\lambda_i = \lambda_j), \\ \dfrac{e^{\lambda_i} - e^{\lambda_j}}{\lambda_i - \lambda_j} & (\lambda_i \neq \lambda_j). \end{cases}$$

Proof: In the special case of nondefective X, all Jordan block are of dimension one. The only relevant coefficient is then θ_{00} which takes the form $\theta_{00} = \delta_{ij}$, since $\alpha_0(0,0) = 1$, $R_1(w) = (e^w - 1)/w$, and $w = \lambda_i - \lambda_j$.

The special case of symmetry was solved by Linton (1995) and McCrorie (1995), but the extension to general nondefective matrices does not seem to have been recorded.

The corollary provides the derivative of $\exp(X)$ when X is nondefective at the point X_0 where the derivative is taken, but possibly defective in a neighborhood of X_0 so that perturbations are unrestricted. But when X is *structurally* nondefective, that is nondefective at X_0 *and* in a neighborhood of X_0, then we have to take this constraint into account. The next section deals with this case.

6 Restrictions on X

When $X = X(t)$ where t is a vector of fewer than n^2 parameters, then X is structurally restricted, and this restriction has to be taken into account. Since

$$d \operatorname{vec} X = \frac{\partial \operatorname{vec} X(t)}{\partial t'} \, dt,$$

the chain rule gives

$$\frac{\partial \operatorname{vec} \exp(X)}{\partial t'} = \nabla(X) \frac{\partial \operatorname{vec} X(t)}{\partial t'}, \qquad \nabla(X) = \frac{\partial \operatorname{vec} \exp(X)}{\partial (\operatorname{vec} X)'}.$$

Let us consider two restrictions that are of particular importance: symmetry and skew-symmetry. Both restrictions are *linear* so that the matrix $\partial \operatorname{vec} X(t)/\partial t'$ does not depend on t.

When X is structurally symmetric, that is, when $X' = X$ at X_0 and in a neighborhood of X_0, then we need to employ the duplication matrix D_n and the vech() operator with the property that

$$D_n \operatorname{vech}(X) = \operatorname{vec} X$$

for every symmetric X; see Magnus (1988, Chapter 4). The derivative of $\exp(X)$ is then given by

$$\frac{\partial \operatorname{vec} \exp(X)}{\partial (\operatorname{vech}(X))'} = \nabla(X) D_n. \tag{10}$$

Similarly, when X is structurally skew-symmetric, that is, when $X' = -X$ at X_0 and in a neighborhood of X_0, then we need to employ the matrix $\widetilde{D}_n$ and the $\tilde{v}()$ operator with the property that

$$\widetilde{D}_n \tilde{v}(X) = \operatorname{vec} X$$

for every skew-symmetric X; see Magnus (1988, Chapter 6). The derivative is now

$$\frac{\partial \operatorname{vec} \exp(X)}{\partial (\tilde{v}(X))'} = \nabla(X) \widetilde{D}_n. \tag{11}$$

Symmetric and skew-symmetric matrices are both normal, that is, they satisfy the restrictions $X'X = XX'$. This implies that the perturbations are also restricted because

$$(dX)'X + X'(dX) - (dX)X' - X(dX)' = 0,$$

so that

$$(I_{n^2} + K_n)(I_n \otimes X' - X \otimes I_n) d \operatorname{vec} X = 0, \tag{12}$$

where K_n is the $n^2 \times n^2$ commutation matrix such that $K_n \operatorname{vec} A = \operatorname{vec} A'$ for any $n \times n$ matrix A; see Magnus (1988, Chapter 3). This restriction applies to all structurally normal matrices. In the case of symmetry the derivative satisfies the restriction because, for any symmetric X,

$$(I_{n^2} + K_n)(I_n \otimes X' - X \otimes I_n) D_n = 0.$$

Similarly, in the case of skew-symmetry we have, for any skew-symmetric X,

$$(I_{n^2} + K_n)(I_n \otimes X' - X \otimes I_n) \widetilde{D}_n = 0.$$

7 Second differential

Although less elegant, it is also possible to obtain higher-order derivatives of the exponential matrix function. For the case of a single parameter this was discussed in Mathias (1996, Theorem 4), and for the symmetric case by Baba (2003). Let's consider the general case for the second-order derivative.

Proposition 4. *The Hessian of the st-th element of the exponential function $F(X) = \exp(X)$ is given by*

$$H_{st} = \frac{\partial^2 F_{st}}{(\partial \operatorname{vec} X)(\partial \operatorname{vec} X)'} = \sum_{k=0}^{\infty} \frac{K_n Q_{k+2}^{(s,t)} + (Q_{k+2}^{(s,t)})' K_n}{(k+2)!},$$

where K_n is the commutation matrix,

$$Q_{k+2}^{(s,t)} = \sum_{h+i+j=k} (X^j E_{ts} X^h)' \otimes X^i,$$

and E_{ts} denotes the $n \times n$ matrix with one in the ts-th position and zeros elsewhere.

In the case of symmetry, skew-symmetry or another linear structure restriction, we need to adjust the Hessian matrix. For example, when X is structurally symmetric, the Hessian matrices with respect to $\operatorname{vech}(X)$ become $D_n' H_{st} D_n$.

8 Discretized Ornstein-Uhlenbeck process

Consider a multivariate version of the Ornstein-Uhlenbeck stochastic process characterized by the following system of linear stochastic differential equations with constant coefficients:

$$dy(t) = Ax(t)\,dt + \Sigma^{1/2}\,dW(t), \tag{13}$$

where $W(t)$ is a continuous-time Wiener process such that $\mathrm{E}\,dW(t) = 0$ and $\mathrm{E}\,dW(t)\,dW'(t) = I_n\,dt$, and the real part of each eigenvalue of A is negative to guarantee stationarity of the process.

When all the elements of y_t are stock variables, Bergstrom (1984) showed that (13) generates discrete observations which, regardless of the discretization interval $h \in \mathbb{R}^+$, follow the VAR(1) model

$$y_t = e^{Ah}y_{t-h} + \eta_t^{(h)} \qquad (t = h, 2h, \dots), \tag{14}$$

where the Gaussian error term $\eta_t^{(h)} = \int_{t-h}^{t} e^{A(t-s)}\Sigma^{1/2}\,dW(s)$ satisfies

$$\mathrm{E}\,\eta_t^{(h)} = 0, \qquad \mathrm{E}(\eta_t^{(h)})(\eta_t^{(h)})' = \int_0^h e^{As}\Sigma e^{A's}\,ds,$$

and

$$\mathrm{E}(\eta_t^{(h)})(\eta_{t-r}^{(h)})' = 0 \qquad (r \geq h).$$

Let θ denote the vector of underlying structural parameters that characterize the continuous-time model (13) through the matrices $A(\theta)$ and $\Sigma(\theta)$. We can then exploit the discretized version (14) to estimate θ from a sample of T discrete equidistant observations on y_t. To simplify the expressions we set $h = 1$ without loss of generality. Given that the conditional distribution of the discrete-time innovations is Gaussian, we can efficiently estimate θ by maximum likelihood under the maintained assumption of identifiability, which we revisit below. (If $W(t)$ is not Gaussian, the estimation procedure should be understood as Gaussian pseudo-maximum likelihood.) To do so, it is convenient to obtain analytical expressions for the derivatives of the conditional mean and variance functions

$$\mu_t(\theta) = e^{A(\theta)}y_{t-1}, \qquad \Omega(\theta) = \int_0^1 e^{A(\theta)s}\Sigma(\theta)e^{A(\theta)'s}\,ds$$

with respect to θ.

Regarding μ_t, we have $d\mu_t = (de^A)y_{t-1}$, and hence

$$\frac{\partial \mu_t}{\partial \theta'} = (y'_{t-1} \otimes I_n)\nabla(A)\frac{\partial\,\mathrm{vec}\,A}{\partial \theta'}, \tag{15}$$

where $\nabla(A)$ denotes the derivative of $\operatorname{vec} e^A$ with respect to $\operatorname{vec} A$ given in Theorem 1.

Regarding Ω, let $F_s = e^{A(\theta)s}\Sigma(\theta)e^{A(\theta)'s}$ so that

$$dF_s = (de^{As})\Sigma e^{A's} + e^{As}(d\Sigma)e^{A's} + e^{As}\Sigma(de^{A's})$$

and

$$\begin{aligned} d\operatorname{vec} F_s &= (e^{As}\Sigma \otimes I_n)\, d\operatorname{vec} e^{As} + (e^{As} \otimes e^{As})\, d\operatorname{vec}\Sigma + (I_n \otimes e^{As}\Sigma)\, d\operatorname{vec} e^{A's} \\ &= (I_{n^2} + K_n)(e^{As}\Sigma \otimes I_n)\, d\operatorname{vec} e^{As} + (e^{As} \otimes e^{As})\, d\operatorname{vec}\Sigma, \end{aligned}$$

where K_n is the commutation matrix. Then,

$$\frac{\partial \operatorname{vech}(\Omega)}{\partial(\operatorname{vec} A)'} = 2D_n^+ \left(\int_0^1 s(e^{As}\Sigma \otimes I_n)\nabla(As)\, ds \right) \tag{16}$$

and

$$\frac{\partial \operatorname{vech}(\Omega)}{\partial(\operatorname{vech}(\Sigma))'} = D_n^+ \left(\int_0^1 (e^{As} \otimes e^{As})\, ds \right) D_n, \tag{17}$$

where D_n is the duplication matrix. The derivatives with respect to θ then follow from the chain rule,

$$\frac{\partial \operatorname{vech}(\Omega)}{\partial\theta'} = \frac{\partial \operatorname{vech}(\Omega)}{\partial(\operatorname{vec} A)'}\frac{\partial \operatorname{vec} A}{\partial\theta'} + \frac{\partial \operatorname{vech}(\Omega)}{\partial(\operatorname{vech}(\Sigma))'}\frac{\partial \operatorname{vech}(\Sigma)}{\partial\theta'}.$$

Alternative expressions for the derivatives can be obtained by noting, as in Phillips (1973), that

$$\Omega = \int_0^1 e^{As}\Sigma e^{A's}\, ds \iff e^A\Sigma e^{A'} - \Sigma = A\Omega + \Omega A', \tag{18}$$

the so-called discrete-time Lyapunov equation. This gives

$$\begin{aligned} &(de^A)\Sigma e^{A'} + e^A(d\Sigma)e^{A'} + e^A\Sigma(de^A)' - d\Sigma \\ &\qquad = (dA)\Omega + A(d\Omega) + (d\Omega)A' + \Omega(dA)', \end{aligned}$$

and upon vectorizing,

$$\begin{aligned} (I_n \otimes A + A \otimes I_n)\ d\operatorname{vec}\Omega &= (e^A \otimes e^A - I_n \otimes I_n)\, d\operatorname{vec}\Sigma \\ &\quad + (I_{n^2} + K_n)(e^A\Sigma \otimes I_n) d\operatorname{vec} e^A - (I_{n^2} + K_n)(\Omega \otimes I_n)\, d\operatorname{vec} A. \end{aligned}$$

Taking the symmetry of Ω and Σ into account, we obtain

$$\begin{aligned} D_n^+ (I_n \otimes A + A \otimes I_n) D_n d\operatorname{vech}(\Omega) &= D_n^+(e^A \otimes e^A - I_n \otimes I_n)D_n d\operatorname{vech}(\Sigma) \\ &\quad + 2D_n^+(e^A\Sigma \otimes I_n)d\operatorname{vec} e^A - 2D_n^+(\Omega \otimes I_n)d\operatorname{vec} A. \end{aligned}$$

The matrix $I_n \otimes A + A \otimes I_n$ is nonsingular if and only if A is nonsingular and its eigenvalues λ_i satisfy $\lambda_i + \lambda_j \neq 0$ for all $i \neq j$ (Magnus, 1988, Theorem 4.12). This is the case in model (13) because we have assumed that $\Re(\lambda_i(A)) < 0$ for all i. Then,

$$\begin{aligned}\frac{\partial \operatorname{vech}(\Omega)}{\partial(\operatorname{vec} A)'} &= 2D_n^+ \left(I_n \otimes A + A \otimes I_n\right)^{-1} D_n D_n^+(e^A \Sigma \otimes I_n)\nabla(A) \\ &\quad - 2D_n^+ \left(I_n \otimes A + A \otimes I_n\right)^{-1} D_n D_n^+(\Omega \otimes I_n) \qquad (19)\end{aligned}$$

and

$$\frac{\partial \operatorname{vech}(\Omega)}{\partial(\operatorname{vech}(\Sigma))'} = D_n^+ \left(I_n \otimes A + A \otimes I_n\right)^{-1} D_n D_n^+(e^A \otimes e^A - I_n \otimes I_n)D_n, \quad (20)$$

which does not involve any integral.

Given that the mapping between Ω and Σ is bijective when Σ is unrestricted, we can estimate the model in terms of A and Ω without loss of efficiency, which considerably simplifies the calculations, especially if we take into account that Ω can be concentrated out of the log-likelihood function (see again Bergstrom, 1984). Given A and Ω, we can solve Σ by writing (18) as

$$D_n^+(e^A \otimes e^A - I_n \otimes I_n)D_n \operatorname{vech}(\Sigma) = D_n^+(I_n \otimes A + A \otimes I_n)D_n \operatorname{vech}(\Omega).$$

This will guarantee that Σ is symmetric, but not that it is positive (semi)-definite, unless $A\Omega + \Omega A'$ is positive (semi)definite; see also Hansen and Sargent (1983).

An important advantage of the analytical expressions for the Jacobian of the exponential in Theorem 1 is that we do not need to compute the exact discretization of the Ornstein-Uhlenbeck process.

Without any restrictions on the matrices A and Σ, the so-called aliasing problem may prevent the global identification of θ from the discretized continuous time process (14); see e.g. Phillips (1973) or Hansen and Sargent (1983). Theorem 1 in McCrorie (2003) states that the parameters A and Σ are identifiable from (14) if the eigenvalues of the matrix

$$M = \begin{pmatrix} -A & \Sigma \\ 0 & A' \end{pmatrix}$$

are strictly real and no Jordan block of M belonging to any eigenvalue appears more than once.

To illustrate this result, let us consider a bivariate example in which $y_2(t)$ does not Granger cause $y_1(t)$ at any discrete horizon, and the instantaneous variance matrix of the shocks is unrestricted. Proposition 21 in Comte and Renault (1996) states that this will happen when A is upper triangular, intuitively because e^{Ah} inherits the upper triangularity from A. McCrorie's conditions are now satisfied when $a_{11} \neq a_{22}$, in which case A is diagonalizable, but also when $a_{11} = a_{22}$, in which case it is defective. The strength of Theorem 1 is that it can be employed to compute the required derivatives in either case.

9 Rotation matrices and structural vector autoregressions

Consider the n-variate structural vector autoregressive process

$$y_t = Ay_{t-1} + C\xi_t,$$

where $\xi_t|I_{t-1} \sim i.i.d.(0, I_N)$ and C is an unrestricted matrix of impact multipliers. Let $\epsilon_t = C\xi_t$ denote the reduced-form innovations, so that $\epsilon_t|I_{t-1} \sim i.i.d.(0, \Sigma)$ with $\Sigma = CC'$. A Gaussian pseudo-loglikelihood function can identify Σ but not C, which implies that the structural shocks ξ_t and their loadings in C are only identified up to an orthogonal transformation. Specifically, we can use the QR decomposition to relate the matrix C to the Cholesky decomposition of $\Sigma = \Sigma_L \Sigma_L'$ as $C' = Q'\Sigma_L'$, where Q is an $n \times n$ orthogonal matrix, which we can model as a function of $n(n-1)/2$ free parameters ω by assuming that $|Q| = 1$. This assumption involves no loss of generality because if $|Q| = -1$ then we can always change the sign of the ith structural shock and its impact multipliers in the ith column of C as long as we also modify the shape parameters of the distribution of ξ_{it} to alter the sign of all its nonzero odd moments.

Nevertheless, statistical identification of both the parameters in ω and the structural shocks in ξ (up to permutations and sign changes) is possible assuming (i) cross-sectional independence of the n shocks, and (ii) a non-Gaussian distribution for at least $n-1$ of them; see Lanne et al. (2017) for a proof, and Brunnermeier et al. (2019) for a recent example of the increased popularity of Svar models with non-Gaussian shocks. Thus, if we exploit the non-Gaussianity of the structural shocks, we can estimate not only the parameters $a = \operatorname{vec} A$ and $\sigma_L = \operatorname{vech}(\Sigma_L)$, but also ω.

To obtain analytical expressions for the score and the conditional information matrix, we require the derivatives of the conditional mean $\mu_t = Ay_{t-1}$ and the conditional variance $\Sigma_t = CC' = \Sigma_L QQ'\Sigma_L'$, and this raises the question of how we should model the orthogonal matrix Q.

To answer this question, we first note that orthogonal matrices have determinant ± 1. The subgroup whose determinant is $+1$ is called the 'special' orthogonal group of rotations. For $n = 2$ there is only one free parameter and any orthogonal matrix takes the form of either

$$A = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \qquad \text{or} \qquad B = \begin{pmatrix} \cos\beta & \sin\beta \\ \sin\beta & -\cos\beta \end{pmatrix}.$$

The matrix A has determinant 1 and is a rotation matrix, while B has determinant -1 and is a reflection matrix.

For $n = 3$, there are three free parameters and any rotation matrix is a product of the Givens matrices

$$A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}, \qquad A_2 = \begin{pmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{pmatrix},$$

and

$$A_3 = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix};$$

see Golub and Van Loan (2013, Section 5.1.8). For an application of this approach to multivariate GARCH models; see van der Weide (2002).

The order in which we multiply the matrices matters, so $A_3A_2A_1$ is just one of six possible rotation matrices that can be constructed from these matrices. The derivative of the resulting orthogonal matrix with respect to α, β, and γ can now be easily constructed.

There are, however, two problems with modelling the special orthogonal group through rotation matrices. The first problem is what navigators call a 'gimbal lock'. For example, when $\beta = \pi/2$ we can only identify $\alpha + \gamma$ from $A = A_3A_2A_1$, but neither parameter separately. The second problem is that parameterizing rotation matrices in terms of the angles of $n(n-1)/2$ Givens matrices becomes rather cumbersome when n increases.

A second way to model orthogonal matrices is through the Cayley transform given by

$$Q = (I_n - H)(I_n + H)^{-1}, \qquad H = (I_n - Q)(I_n + Q)^{-1}, \tag{21}$$

where H is skew-symmetric (Bellman, 1970, p. 92). This gives

$$\begin{aligned} dQ &= -(dH)(I_n + H)^{-1} - (I_n - H)(I_n + H)^{-1}(dH)(I_n + H)^{-1} \\ &= -\frac{1}{2}(I_n + Q)(dH)(I_n + Q), \end{aligned}$$

and hence

$$\frac{\partial \operatorname{vec} Q}{\partial(\tilde{v}(H))'} = -\frac{1}{2}\left((I_n + Q') \otimes (I_n + Q)\right) \widetilde{D}_n. \tag{22}$$

This approach also has drawbacks because some rotation matrices can only be obtained by letting some elements of the underlying skew-symmetric matrix go to infinity. For example, for $Q = -I_2$ and letting

$$H = \begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix}$$

we have

$$(I_n - H)(I_n + H)^{-1} = \frac{1}{1+\omega^2}\begin{pmatrix} 1-\omega^2 & -2\omega \\ 2\omega & 1-\omega^2 \end{pmatrix},$$

and this only approaches $-I_2$ when $\omega \to \pm\infty$.

The third and arguably most suitable approach to modelling orthogonality is also based on the connection between orthogonal and skew-symmetric matrices. This connection results from the fact that $Q'Q = I_n$ implies that $(dQ)'Q + Q'dQ = 0$, and hence that $Q'dQ$ is skew-symmetric. The Lie algebra of an orthogonal matrix group thus consists of skew-symmetric matrices. Put differently, the matrix exponential of any skew-symmetric matrix is a (special) orthogonal matrix because $H + H' = 0$ implies that

$$I_n = e^0 = e^{H'+H} = (e^H)'(e^H).$$

For $Q = e^H$ we thus obtain

$$dQ = de^H = \frac{\partial \operatorname{vec} e^H}{\partial(\operatorname{vec} H)'}\, d \operatorname{vec} H = \nabla(H)\widetilde{D}_n\, d\tilde{v}(H),$$

and hence

$$\frac{\partial \operatorname{vec} Q}{\partial(\tilde{v}(H))'} = \nabla(H)\widetilde{D}_n, \tag{23}$$

where $\nabla(H)$ is given in Corollary 1 in closed form.

10 Conclusions

The purpose of this paper was to present a closed-form expression for the Jacobian of the exponential function, applicable for both diagonalizable and defective matrices, and to discuss some applications. It may be possible to obtain a similarly attractive result for the Hessian (instead of Proposition 4), but this is perhaps a topic for future research.

We mention two further issues. First, if $Y = \exp(X)$, then $X = \log(Y)$ is the logarithm of Y. Differentiating both sides of $X = \log(\exp(X))$, we find

$$\frac{\partial \operatorname{vec} \log(Y)}{\partial (\operatorname{vec} Y)'} \frac{\partial \operatorname{vec} \exp(X)}{\partial (\operatorname{vec} X)'} = I_{n^2},$$

and hence the Jacobian of the logarithm is the inverse of the Jacobian of the exponential. Some care is, however, required because not all matrices have a logarithm and those matrices that do have a logarithm may have more than one (Bellman, 1970, Section 11.20). A necessary condition for a matrix Y to have a logarithm is that Y is nonsingular. For complex matrices, this condition is also sufficient, but a real matrix Y has a real logarithm if and only if it is nonsingular and each Jordan block belonging to a negative eigenvalue occurs an even number of times.

Second, we have assumed that the matrix X is real, although its eigenvalues and eigenvectors will in general be complex. Our results are, however, also valid for complex matrices. In particular Proposition 3 and Theorem 1 remain valid without modification. The derivative now becomes the complex derivative with respect to the complex matrix Z, and $\exp(Z)$ and $d\exp(Z)$ are analytic in n^2 complex variables.

Appendix A: Proof of the propositions

Proof of Proposition 1: See text.

Proof of Proposition 2: See text.

Proof of Proposition 3: Setting $t = 1$ in (8) gives

$$e^{X+dX} - e^X = \int_0^1 e^{X(1-s)}(dX)e^{(X+dX)s}\, ds,$$

so that

$$de^X = e^X \int_0^1 e^{-Xs}(dX)e^{Xs}\, ds,$$

using the fact that

$$(dX)e^{(X+dX)s} = (dX)e^{Xs} + O((dX)^2).$$

This gives the first expression. The second expression follows from the fact that the matrices $A = X' \otimes I_n$ and $B = I_n \otimes X$ commute, so that

$$\begin{aligned} e^{(X'\otimes I_n - I_n \otimes X)s} &= e^{(A-B)s} = e^{As}e^{-Bs} \\ &= \left((e^{Xs})' \otimes I_n\right)\left(I_n \otimes e^{-Xs}\right) = (e^{Xs})' \otimes e^{-Xs}. \end{aligned}$$

To prove the third expression we note that

$$\int_0^1 e^{(A-B)s}\, ds = \sum_{k=0}^{\infty} \frac{(A-B)^k}{k!} \int_0^1 s^k ds = \sum_{k=0}^{\infty} \frac{(A-B)^k}{(k+1)!}.$$

Proof of Proposition 4: We have

$$\begin{aligned} d^2X^2 &= 2(dX)(dX), \\ d^2X^3 &= 2\left((dX)(dX)X + (dX)X(dX) + X(dX)(dX)\right), \end{aligned}$$

and, in general,

$$d^2X^{k+2} = 2 \sum_{h+i+j=k} X^h(dX)X^i(dX)X^j.$$

Let e_s and e_t be elementary $n \times 1$ vectors, that is, e_s has one in its s-th position and zeros elsewhere, and e_t has one in its t-th position and zeros elsewhere. Then $E_{st} = e_s e_t'$. Now consider the st-th element of d^2X^{k+2}:

$$\begin{aligned} \left(d^2X^{k+2}\right)_{st} &= 2 \sum_{h+i+j=k} e_s' X^h(dX)X^i(dX)X^j e_t \\ &= 2 \sum_{h+i+j=k} \operatorname{tr} X^j E_{ts} X^h(dX)X^i(dX) \\ &= 2\,(d\operatorname{vec} X)' K_n Q_{k+2}^{(s,t)} d\operatorname{vec} X, \end{aligned}$$

where we have used the fact that

$$\operatorname{tr} A(dX)B(dX) = (d\operatorname{vec} X)' K_n (A' \otimes B)\, d\operatorname{vec} X.$$

Hence, the second differential of the st-th element of $F(X) = \exp(X)$ is given by

$$d^2F_{st} = \sum_{k=0}^{\infty} \frac{2}{(k+2)!} (d\operatorname{vec} X)' K_n Q_{k+2}^{(s,t)} d\operatorname{vec} X,$$

and the Hessian follows.

Appendix B: Proof of Theorem 1

The proof of the theorem rests on following three lemmas.

Lemma 1. *For any integer $p \geq 0$ and any w (real or complex), we have*

$$w^{p+1} \int_0^1 r^p e^{-wr}\, dr = p! \left(1 - e^{-w} \sum_{j=0}^{p} \frac{w^j}{j!}\right).$$

Proof: Let $a_p(w) = \int_0^1 r^p e^{-wr}\, dr$. Partial integration gives the recursion

$$wa_p(w) = pa_{p-1}(w) - e^{-w}, \qquad a_0(w) = (1 - e^{-w})/w,$$

and the result follows by induction. Note the close relationship with the (lower) incomplete gamma function

$$\gamma(p, w) = \int_0^w t^{p-1} e^{-t}\, dt \qquad (\Re(p) > 0),$$

where p and w may be complex and the real part of p is positive. In the special case where p is a positive integer this can also be written as

$$\gamma(p, w) = (p-1)! \left(1 - e^{-w} \sum_{j=0}^{p-1} \frac{w^j}{j!}\right) \qquad (p \geq 1);$$

see DLMF (2020, Eq. 8.4.7).

Lemma 2. *Let x follow a beta distribution*

$$f(x; p, q) = \frac{1}{B(p, q)} x^{p-1}(1 - x)^{q-1}$$

where $p \geq 1$ and $q \geq 1$ are integers and $0 \leq x \leq 1$. Then, for any w (real or complex),

$$\mathrm{E}(e^{-wx}) = e^{-w} \sum_{i=0}^{q-1} \alpha_i(p-1, q-1) R_{p+i}(w),$$

where $\alpha_i(s, t)$ and $R_{n+1}(w)$ are defined in the theorem.

Proof: Using Lemma 1 we obtain

$$\begin{aligned}
\int_0^1 e^{-wx} x^{p-1}(1-x)^{q-1}\, dx &= \sum_{i=0}^{q-1} (-1)^i \binom{q-1}{i} \int_0^1 e^{-wx} x^{p+i-1}\, dx \\
&= \sum_{i=0}^{q-1} (-1)^i \binom{q-1}{i} w^{-(p+i)} (p+i-1)! \left(1 - e^{-w} \sum_{j=0}^{p+i-1} \frac{w^j}{j!}\right) \\
&= e^{-w} \sum_{i=0}^{q-1} (-1)^i \binom{q-1}{i} \frac{R_{p+i}(w)}{p+i} \\
&= B(p, q) e^{-w} \sum_{i=0}^{q-1} \alpha_i(p-1, q-1) R_{p+i}(w).
\end{aligned}$$

Here we note that the moment-generating and characteristic functions of the beta distribution with integer-valued parameters follow as special cases by setting $w = -t$ and $w = -it$, respectively.

Lemma 3. *Let $\alpha_i(s,t)$ be as defined in the theorem. Then* $\sum_{i=0}^{t}\alpha_i(s,t)=1$.

Proof: The result follows from the Chu-Vandermonde identity (Askey, 1975, p. 60), but can also be proved by observing that

$$\int_0^1 r^s(1-r)^t\,dr=\frac{s!\,t!}{(s+t+1)!}$$

from the definition of the beta distribution, and also

$$\int_0^1 r^s(1-r)^t\,dr=\sum_{i=0}^{t}(-1)^i\binom{t}{i}\int_0^1 r^{s+i}\,dr=\sum_{i=0}^{t}\frac{(-1)^i}{s+i+1}\binom{t}{i}.$$

Hence,

$$1=\sum_{i=0}^{t}\frac{(-1)^i}{s+i+1}\frac{(s+t+1)!}{s!\,t!}\binom{t}{i}=\sum_{i=0}^{t}(-1)^i\binom{s+i}{i}\binom{s+t+1}{t-i}.$$

Proof of the theorem: Based on the three lemmas we now prove the theorem. Our starting point is

$$\nabla(X)=\frac{\partial\,\mathrm{vec}\,F}{\partial(\mathrm{vec}\,X)'}=\int_0^1 e^{X'r}\otimes e^{X(1-r)}\,dr=\int_0^1 e^{X'(1-r)}\otimes e^{Xr}\,dr,$$

as given in Proposition 3. Let

$$T^{-1}XT=J=\mathrm{diag}(J_1,\ldots,J_m),\qquad J_i=\lambda_i I_{n_i}+E_{n_i}$$

be the Jordan decomposition. Then, as in the derivation of (3),

$$e^{J_v r}=e^{\lambda_v r}\sum_{s=0}^{n_v-1}\frac{r^s}{s!}E_{n_v}^s$$

and

$$e^{J_u'(1-r)}=e^{\lambda_u(1-r)}\sum_{t=0}^{n_u-1}\frac{(1-r)^t}{t!}(E_{n_u}')^t,$$

so that

$$e^{J_u'(1-r)}\otimes e^{J_v r}=\sum_{t=0}^{n_u-1}\sum_{s=0}^{n_v-1}\frac{e^{\lambda_u(1-r)}(1-r)^t}{t!}\,\frac{e^{\lambda_v r}r^s}{s!}\,(E_{n_u}')^t\otimes(E_{n_v})^s.$$

Hence, the Jacobian is $\nabla(X) = S\Delta S^{-1}$, where

$$S = (T')^{-1} \otimes T, \qquad \Delta = \operatorname{diag}(\Delta_{11}, \Delta_{12}, \ldots, \Delta_{mm}),$$

and

$$\Delta_{uv} = \sum_{t=0}^{n_u-1} \sum_{s=0}^{n_v-1} \theta_{ts} (E'_{n_u})^t \otimes (E_{n_v})^s$$

with

$$\theta_{ts} = \int_0^1 \frac{e^{\lambda_u(1-r)}(1-r)^t}{t!} \, \frac{e^{\lambda_v r} r^s}{s!} \, dr$$

To complete the proof we need to show that this expression for θ_{ts} equals the expression for θ_{ts} in the theorem. Let $w = \lambda_u - \lambda_v$. Then, by Lemma 2,

$$e^{-\lambda_v}\theta_{ts} = \frac{e^w}{s!\,t!} \int_0^1 e^{-wr} r^s (1-r)^t \, dr = \frac{1}{(s+t+1)!} \sum_{i=0}^{t} \alpha_i(s,t) R_{s+i+1}(w).$$

This completes the proof.

Achnowledgements

We are grateful to Tom Koornwinder for help on the Chu-Vandermonde equality. Sentana gratefully acknowledges financial support from the Spanish Ministry of Economy, Industry and Competitiveness through grant ECO 2017-89689. This research did not receive any other grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Abadir, K. M. and J. R. Magnus (2005). *Matrix Algebra.* Cambridge University Press.

Askey, R. (1975). *Orthogonal Polynomials and Special Functions.* SIAM.

Baba, Y. (2003). The derivation of the Hessian matrix of $\exp(A)$ with a symmetric real matrix A. *Soka Economic Studies*, 33, 35–45.

Bellman, R. (1970). *Introduction to Matrix Analysis*, 2nd edition. Mc-Graw Hill.

Bergstrom, A. R. (1984). Continuous time stochastic models and issues of aggregation over time, in: *Handbook of Econometrics*, Vol. 2 (Z. Griliches and M. D. Intriligator, eds). North Holland.

Brunnermeier, M., D. Palia, K. A. Sastry, and C. A. Sims (2019). Feedbacks: financial markets and economic activity. Mimeo, Princeton University.

Cerdà-Alabern, L. (2013). Closed form transient solution of continuous time Markov chains through uniformization. Proceedings of the 7th International Conference on Performance Evaluation Methodologies and Tools, pp. 263–272.

Chen, B. and P. A. Zadrozny (2001). Analytic derivatives of the matrix exponential for estimation of linear continuous-time models. *Journal of Economic Dynamics & Control*, 25, 1867–1879.

Comte, F. and E. Renault (1996). Noncausality in continuous time models. *Econometric Theory*, 12, 215–256.

DLMF (2020). *NIST Digital Library of Mathematical Functions*. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds. http://dlmf.nist.gov/, Release 1.0.26.

Fung, T. C. (2004). Computation of the matrix exponential and its derivatives by scaling and squaring *International Journal for Numerical Methods in Engineering*, 59, 1273–1286.

Golub, G. H. and C. F. Van Loan (2013). *Matrix Computations*, 4th ed. Johns Hopkins.

Hall, B. C. (2015). *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*, 2nd edition. Springer.

Hansen, L. P. and T. J. Sargent (1983). The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica*, 51, 377–387.

Karplus, R. and J. Schwinger (1948). A note on saturation in microwave spectroscopy. *Physical Review*, 73, 1020–1026.

Kawakatsu, H. (2006). Matrix exponential GARCH. *Journal of Econometrics*, 134, 95–128.

Lanne, M., M. Meitz, and P. Saikkonen (2017). Identification and estimation of non-Gaussian structural vector autoregressions. *Journal of Econometrics*, 196, 288–304.

Linton, O. (1993). An alternative exponential Garch model with multivariate extension. Mimeo, Nuffield College, Oxford, UK.

Linton, O. (1995). Differentiation of an exponential matrix function. Solution to Problem no. 94.3.2. *Econometric Theory*, 11, 1182–1183.

Magnus, J. R. (1988). *Linear Structures.* Oxford University Press.

Magnus, J. R. and H. Neudecker (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd edition. John Wiley & Sons.

Mathias, R. (1996). A chain rule for matrix functions and applications. *SIAM Journal on Matrix Analysis and Applications*, 17, 610–620.

McCrorie, J. R. (1995). Differentiation of an exponential matrix function. Solution to Problem no. 94.3.2. *Econometric Theory*, 11, 1183–1185.

McCrorie, J. R. (2003). The problem of aliasing in identifying finite parameter continuous time stochastic models. *Acta Applicandae Mathematica*, 79, 9–16.

Naifeld, I. and T. F. Havel (1995). Derivatives of the matrix exponential and their computation. *Advances in Applied Mathematics*, 16, 321–375.

Phillips, P. C. B. (1973). The problem of identification in finite parameter continuous time models. *Journal of Econometrics*, 1, 351–362.

Snider, R. F. (1964). Perturbation variation methods for a quantum Boltzmann equation. *Journal of Mathematical Physics*, 5, 1580–1587.

Tsai, H. and K. S. Chan (2003). A note on parameter differentiation of matrix exponentials, with applications to continuous-time modelling. *Bernoulli*, 9, 895–919.

Tuynman, G. M. (1995). The derivation of the exponential map of matrices. *The American Mathematical Monthly*, 102, 818–820.

van der Weide, R. (2002). Go-Garch: A multivariate generalized orthogonal Garch model. *Journal of Applied Econometrics*, 17, 549–564.

Van Loan, C. F. (1978). Computing integrals involving the matrix exponential. *IEEE Transactions on Automatic Control*, 23, 395–404.

Wilcox, R. M. (1967). Exponential operators and parameter differentiation in quantum physics. *Journal of Mathematical Physics*, 8, 962–982.