Pseudo-Maximum Likelihood and Lie Groups of Linear Transformations

Christian Gouriéroux and Alain Monfort and Jean-Michel Zakoian

9 June 2017

Pseudo-Maximum Likelihood and Lie Groups of Linear Transformations

C., Gouriéroux [(1)], A., Monfort [(2)], and J.M., Zakoian [(3)]

June 2017

Abstract

Newey, Steigerwald (1997) considered a univariate conditionally heteroscedastic model, with independent and identically distributed errors. They showed that the parameters characterizing the serial dependence are consistently estimated by any pseudo maximum likelihood approach, whenever two additional parameters, one for location, one for scale, are appropriately introduced in the model. Our paper extends their result to a more general multivariate framework. We show the consistency of any pseudo maximum likelihood method for multivariate models based on Lie groups of (linear, affine) transformations when these groups commute, or at least satisfy a property of closure under commutation. We explain how to introduce appropriately the additional parameters which capture all the bias due to the misspecification of the error distribution. We also derive the asymptotic distribution of the PML estimators.

Keywords : Pseudo Maximum Likelihood, Lie Group, Transformation Model, GARCH Model, Infinitesimal Generator, Rotation, Computer Vision, Machine Learning, Volatility Matrices.

[1] CREST and University of Toronto.

[2] CREST.

[3] CREST and University of Lille.

[4] We gratefully acknowledge financial support of the chair ACPR : "Regulation and Systemic Risk", and of the chair LCL : "New Challenges for New Data".

1 Introduction

A pseudo maximum likelihood[5] (PML) estimator is a maximum likelihood estimator based on a misspecified parametric model. In general, a PML estimator does not converge to the true value of the parameter of interest due to misspecification. However, in some cases, the convergence can be satisfied. A well-known example is the estimation of a regression function by Gaussian PML leading to the nonlinear least squares consistent method [see Bollerslev, Wooldridge (1992) for the analogue in conditionally heteroscedastic models] and more generally the PML methods based on linear exponential pseudo-families [Gouriéroux, Monfort, Trognon (1984)]. In such situations the PML approaches are semi-parametric methods and are alternatives to other semi-parametric methods, such as the methods of moments [Hansen (1982)].

Typically these PML approaches allow for consistent estimation of θ_0 in a regression model[6]:

$$y_t = a(x_t; \theta_0) + u_t, \tag{1.1}$$

where the error is conditionally zero-mean:

$$E_0(u_t|x_t) = 0.$$

However, other consistent PML approaches can be constructed for the regression model (1.1), when the errors are independently, identically distributed (i.i.d.) with not necessarily a zero mean [Newey, Steigerwald (1997)]. Under this alternative assumption on the error term, they prove the consistency of the PML estimator of θ_0, up to the intercept parameter, for any selected pseudo distribution of u_t, not necessarily a pseudo distribution in the linear exponential family.

The objective of our paper is to extend the Newey, Steigerwald consistency result to more general econometric models, constructed from Lie groups[7] of linear transformations. In particular, we explain how to introduce "intercept" parameters in order to capture all the misspecfication biases and to obtain the consistency of the other parameters characterizing the sensitivities with respect to the explanatory variables.

In Section 2 we introduce single index models, where the effect of the explanatory variables is contained in a scalar parametric function of the explanatory variables. In Section 3 we consider PML estimators of such models and we analyze their asymptotic properties. The case of multi-index models is developed in Section 4. Transformation models for normalized data are introduced in Section 5. Section 6 concludes. Proofs are gathered in appendices.

2 Single Index Model

2.1 The econometric model

We consider observations of a period-t n-dimensional random vector $\boldsymbol{y}_t, t = 1, \ldots, T$, satisfying:

$$\boldsymbol{y}_t = \exp[-\{a(\boldsymbol{x}_t; \boldsymbol{\beta}_0) + \alpha_0\}\boldsymbol{C}]\boldsymbol{u}_t, \tag{2.1}$$

[5] or quasi-maximum likelihood (QML) estimator [see e.g. Newey, Steigerwald (1997)]. We avoid this terminology, which could be source of misunderstanding, as it overlaps a slightly different alternative [Wedderburn (1974), McCullagh (1983)].

[6] Consistent PML approaches have also been developed for median and quantile regression models [see e.g. Gouriéroux and Monfort (Chap. 8, 1995), Gouriéroux, Monfort and Renault (2016)].

[7] A Lie group is a differentiable manifold with a group structure. Any group of square matrices is automatically a Lie group [see Boothby (1975)].

where the n-dimensional errors $\boldsymbol{u}_t$ are i.i.d. with unknown common probability density function (p.d.f.) g_0 (with respect to a dominating measure), $\boldsymbol{x}_t$ are period-t explanatory variables such that $\boldsymbol{x}_t$ and $\boldsymbol{u}_t$ are independent, a is a known scalar index function, $\boldsymbol{C}$ a known (n,n) matrix, $\alpha_0, \boldsymbol{\beta}_0$ are the true values of parameters, belonging to parameter sets Θ_α and $\Theta_{\boldsymbol{\beta}}$, respectively, with $\dim \alpha_0 = 1, \dim \boldsymbol{\beta}_0 = p$.[8] In particular we assume that the index function is well-specified. The exponential of a matrix $\boldsymbol{C}$ is defined by:

$$\exp \boldsymbol{C} = \sum_{h=0}^{\infty} \frac{\boldsymbol{C}^h}{h!},$$

where the absolute convergence holds for any multiplicative norm on the set of real (n,n) matrices. The econometric model (2.1) is related to the family of transformations such that:

$$\boldsymbol{y} = \exp(-a\boldsymbol{C})\boldsymbol{u}, \quad \text{say.} \tag{2.2}$$

The scalar parameter a in (2.2) has been replaced in (2.1) by a (nonlinear) regression (index) function $a(x;\boldsymbol{\beta})$, with an additive "intercept" α.

The family of transformations (2.2) parameterized by a can be interpreted as follows: we apply to the (multivariate) shock $\boldsymbol{u}$ a linear transformation indexed by a. The set of linear transformations from $\boldsymbol{y}$ to $\boldsymbol{u}$: $\mathcal{S} = \{\exp(a\boldsymbol{C}), a \in \mathbb{R}\}$ is a commutative Lie group.[9] These transformations are one-to-one, commute and the function: $a \to \exp(a\boldsymbol{C})$ is differentiable at $a = 0$. Its derivative at $a = 0$ is the infinitesimal generator of the Lie group and is equal to $\boldsymbol{C}$.[10] The function $a \to \exp(a\boldsymbol{C})$ defines a curve called a geodesic (or orbit), for which the "velocity" a is constant. Regression (2.1) corresponds to a situation with time varying velocity.

As seen in the examples below, depending of the model, this velocity can be interpreted as a location, scale, or angle parameter.

2.2 Examples

(i) One-dimensional model

The variables: $\exp(-a)u$ define a family of transformations on $\mathbb{R}$, with scale parameter a. This family can be used to define conditionally heteroscedastic models:

$$y_t = \exp[-\{a(\boldsymbol{x}_t;\boldsymbol{\beta}) + \alpha\}]u_t,$$

where $\boldsymbol{x}_t$ can include exogenous variables as well as lagged endogenous variables.

(ii) Bidimensional model associated with the group of rotations

The most standard Lie group on $\mathbb{R}^2$ is the group of rotations. The infinitesimal generator is $\boldsymbol{C} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $\exp(a\boldsymbol{C}) = \begin{pmatrix} \cos a & -\sin a \\ \sin a & \cos a \end{pmatrix}$.

[8] For clarity, we denote matrices by bold uppercase quantities and vectors by bold lowercase quantities.

[9] or Abelian Lie group, since $\exp(a\boldsymbol{C})\exp(b\boldsymbol{C}) = \exp\{(a+b)\boldsymbol{C}\}$, and $\{\exp(a\boldsymbol{C})\}^{-1} = \exp(-a\boldsymbol{C})$. Any one-dimensional Abelian group of (n,n) matrices has this exponential form. See for instance Godement (2004), Chapter 3, Theorem 2.

[10] $\boldsymbol{C}$ is the generator of the tangent space at 0 of the Lie group, that is, of the associated Lie algebra.

Thus the associated econometric model is:

$$\boldsymbol{y}_t = \begin{pmatrix} \cos[a(\boldsymbol{x}_t;\boldsymbol{\beta})+\alpha] & \sin[a(\boldsymbol{x}_t;\boldsymbol{\beta})+\alpha] \\ -\sin[a(\boldsymbol{x}_t;\boldsymbol{\beta})+\alpha] & cos[a(\boldsymbol{x}_t;\boldsymbol{\beta})+\alpha] \end{pmatrix} \boldsymbol{u}_t := R[a(\boldsymbol{x}_t;\boldsymbol{\beta})+\alpha]\boldsymbol{u}_t, \text{ say.}$$

It defines a bivariate conditionally heteroscedastic model with conditional volatility-covolatility matrix given by:

$$V(\boldsymbol{y}_t|\boldsymbol{x}_t) = R[a(\boldsymbol{x}_t;\boldsymbol{\beta})+\alpha]\Omega R'[a(\boldsymbol{x}_t;\boldsymbol{\beta})+\alpha],$$

with $\boldsymbol{\Omega} = V(\boldsymbol{u}_t)$. If the components of $\boldsymbol{y}_t$ are asset (excess) returns, we get a dynamic model in which the maximum (resp. minimum) reachable risk measured by the largest (resp. smallest) eigenvalue of $\boldsymbol{\Omega}$ is path invariant, but the direction of the most (resp. least) risky portfolio allocation is path dependent. This modelling is the opposite of standard multivariate ARCH models where the directions are fixed whereas the eigenvalues are path dependent.

iii) Orthogonal transformations

The example above is easily extended to orthogonal transformations of any dimension n. The transformations $\exp(a\boldsymbol{C}), a$ varying, are orthogonal if and only if,

$$\forall a, \quad [\exp(a\boldsymbol{C})]' = [\exp(a\boldsymbol{C})]^{-1} \Longleftrightarrow \forall a, \quad \exp(a\boldsymbol{C}') = \exp(-a\boldsymbol{C}) \Longleftrightarrow \boldsymbol{C}' = -\boldsymbol{C}.$$

Thus we have just to select an antisymmetric infinitesimal generator $\boldsymbol{C}$.

iv) Additive model

For special infinitesimal generators, parameter a can be interpreted as a location parameter. For expository purpose let us assume $n = 2$ and consider the infinitesimal generator:

$$\boldsymbol{C} = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}.$$

Matrix $\boldsymbol{C}$ is nilpotent: $\boldsymbol{C}^2 = 0$, and $\exp(-a\boldsymbol{C}) = \boldsymbol{I}_2 - a\boldsymbol{C} = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$.

Let us also choose $\boldsymbol{u}_t = (u_{1t}, 1)'$, i.e. let us fix to 1 the second component of $\boldsymbol{u}_t$. Then we have:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \exp(-a\boldsymbol{C}) \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \Longleftrightarrow \begin{cases} y_{1t} = u_{1t} + au_{2t} = u_{1t} + a, \\ y_{2t} = u_{2t} = 1. \end{cases}$$

Thus a is the location parameter for the distribution of y_{1t}. In this example the second component $y_{2t} = 1$ is not informative about parameter a.

v) Nilpotent infinitesimal generator

More generally, if the infinitesimal generator is such that $\boldsymbol{C}^q = 0$, we have: $\exp(a\boldsymbol{C}) = \sum_{h=0}^{q-1} \frac{\boldsymbol{C}^h a^h}{h!}$.
The observations $\boldsymbol{y}_t$ are linear combinations of the shocks $\boldsymbol{u}_t$, with coefficients that are polynomials in parameter a.

vi) Dynamics of probability distribution

Specifications (2.1), (2.2) can also be used to define dynamic models for probability distributions on a finite state space. In this framework both the endogenous variable $\boldsymbol{y}$ and the vector of shocks $\boldsymbol{u}$ define probability distributions, that is, they satisfy:

$$u_i > 0, \quad y_i > 0, \quad i = 1, \ldots, n, \qquad \sum_{i=1}^{n} u_i = 1, \quad \sum_{i=1}^{n} y_i = 1.$$

To pass from $\boldsymbol{u}$ to $\boldsymbol{y}$, we can choose as generator $\boldsymbol{C}$ the infinitesimal generator matrix of a Markov chain, that is a matrix such that:

$$c_{ii} < 0, \quad i = 1, \ldots, n, \qquad c_{ij} > 0, \quad i \neq j, \quad i, j = 1, \ldots, n,$$

with $\sum_{j=1}^{n} c_{ij} = 0, i = 1, \ldots, n$.

In the continuous time Markov chain interpretation, the generator matrix defines the intensities of jumping from one state to another one (for $i \neq j$), and the exponential transform $\exp(a\boldsymbol{C})$ is the transition matrix at horizon $a, a > 0$ [see e.g. Norris (1997), Karlin and Taylor (1998)].

3 Pseudo Maximum Likelihood Estimator

We start this section with general asymptotic results which can be applied when an econometric model (such as Model (2.1)) is related to a generic model (such as Model (2.2)).

3.1 General asymptotic results for PML estimators

We assume that there exists:

$$\boldsymbol{\lambda}_0^* = \arg\max_{\boldsymbol{\lambda} \in \Lambda} E_0 l(\boldsymbol{\lambda}; \boldsymbol{u}_t), \tag{3.1}$$

a maximizer of the asymptotic pseudo log-likelihood of some generic model, where $\boldsymbol{\lambda}$ is a parameter belonging to a compact set $\Lambda \subset \mathbb{R}^s$, $(\boldsymbol{u}_t)$ is an i.i.d. sequence of random vectors belonging to $\mathbb{R}^n$, and l denotes the log-likelihood for one observation. We also assume that there exists a maximizer of the asymptotic pseudo log-likelihood of the econometric model,

$$\boldsymbol{\theta}_0^* = \arg\max_{\boldsymbol{\theta} \in \Theta} E_0 l\{\boldsymbol{\lambda}_t(\boldsymbol{\theta}); \boldsymbol{u}_t\}, \tag{3.2}$$

where $\boldsymbol{\theta}$ is a parameter belonging to a compact set $\Theta \subset \mathbb{R}^r$, and that $\boldsymbol{\lambda}_t : \mathbb{R}^r \mapsto \mathbb{R}^s$ is a twice differentiable random function. In these expressions, and in the sequel, E_0 denotes the expectation w.r.t. the true distribution of the variables involved ($\boldsymbol{u}_t$ in (3.1), $\boldsymbol{\lambda}_t(\boldsymbol{\theta})$ and $\boldsymbol{u}_t$ in (3.2)).

Assumption A0 [Fundamental assumption] The function $\boldsymbol{\lambda}_t$ is such that:

i) $\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0^*) = \boldsymbol{\lambda}_0^*$,

ii) the matrices $\dfrac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}}, \quad \dfrac{\partial^2 \boldsymbol{\lambda}_t'(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta} \partial \theta_j}, \quad$ for $j = 1, \ldots, r$, are independent of $\boldsymbol{u}_t$.

Usually, a log-likelihood function is written as a function of parameters and observed endogenous and exogenous variables. The quantity $l(\boldsymbol{\lambda}_t(\boldsymbol{\theta}); \boldsymbol{u}_t)$ is the expression of the log-likelihood for the $t-$th observation, after expressing the endogenous variables $\boldsymbol{y}_t$ in terms of exogenous variables $\boldsymbol{x}_t$ and innovations $\boldsymbol{u}_t$. The stochastic function $\boldsymbol{\lambda}_t(\boldsymbol{\theta})$ only depends on the explanatory variables $\boldsymbol{x}_t$. Therefore, Assumption **A0** ii) will be satisfied when the innovations are independent from the explanatory variables. Assumption **A0** i), which relates the pseudo-true values of the parameters in the generic and econometric models, is crucial for studying the convergence of the PML estimator of $\boldsymbol{\theta}$. We will show that this condition is satisfied in all our examples.

Let us also assume that $(\lambda_t(\boldsymbol{\theta}), \boldsymbol{u}_t)$ is a strictly stationary sequence, for any $\boldsymbol{\theta} \in \Theta$, that l is second-order differentiable with respect to $\boldsymbol{\lambda}$, and that the following $s \times s$ matrices exist:

$$\boldsymbol{I}^0 = \mathrm{Var}_0 \left\{ \frac{\partial l}{\partial \boldsymbol{\lambda}} (\boldsymbol{\lambda}_0^*; \boldsymbol{u}_t) \right\}, \quad \boldsymbol{J}^0 = E_0 \left\{ -\frac{\partial^2 l}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} \{\boldsymbol{\lambda}_0^*; \boldsymbol{u}_t\} \right\}.$$

Our first result relates matrices involved in the asymptotic distribution of the PML estimators in the econometric and generic models. Let $\ell_t(\boldsymbol{\theta}) = l\{\boldsymbol{\lambda}_t(\boldsymbol{\theta}); \boldsymbol{u}_t\}$.

Proposition 3.1 *Under Assumption* **A0***, and assuming that the vectors* $\frac{\partial \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}_j}$ $(j = 1, \dots, r)$ *have a finite variance-covariance matrix, we have*

$$\boldsymbol{I} := Var_0 \left\{ \frac{\partial \ell_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}} \right\} = E_0 \left\{ \frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}} \boldsymbol{I}^0 \frac{\partial \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}'} \right\}, \tag{3.3}$$

and

$$\boldsymbol{J} := E_0 \left\{ -\frac{\partial^2 \ell_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\} = E_0 \left\{ \frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}} \boldsymbol{J}^0 \frac{\partial \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}'} \right\}. \tag{3.4}$$

Moreover, $\boldsymbol{I}$ *and* $\boldsymbol{J}$ *are nonsingular whenever* $\boldsymbol{I}^0$ *and* $\boldsymbol{J}^0$ *are nonsingular and the matrix*

$$\left(\frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}} \right) \textit{is full row rank}, \quad a.s. \tag{3.5}$$

Proof: See Appendix B.1.

Our next result gives sufficient conditions for the existence of $\boldsymbol{\theta}_0^*$ satisfying (3.2).

Proposition 3.2 *Suppose that:*

1) for any $\boldsymbol{\theta} \in \Theta$*, we have* $\boldsymbol{\lambda}_t(\boldsymbol{\theta}) \in \Lambda, a.s.$

2) there exists $\boldsymbol{\theta}_0^*$ *such that* $\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0^*) = \boldsymbol{\lambda}_0^*$*, where* $\boldsymbol{\lambda}_0^*$ *satisfies (3.1),*

3) for any $\boldsymbol{\theta} \in \Theta$*, the random vectors* $\boldsymbol{\lambda}_t(\boldsymbol{\theta})$ *and* $\boldsymbol{u}_t$ *are independent.*

Then $\boldsymbol{\theta}_0^*$ *satisfies (3.2).*

Proof: We have

$$\begin{aligned} E_0 l\{\boldsymbol{\lambda}_t(\boldsymbol{\theta}); \boldsymbol{u}_t\} &= E_0(E_0[l\{\boldsymbol{\lambda}_t(\boldsymbol{\theta}); \boldsymbol{u}_t\} | \boldsymbol{\lambda}_t(\boldsymbol{\theta})]) \\ &\leq E_0[E_0\{l(\boldsymbol{\lambda}_0^*; \boldsymbol{u}_t) | \boldsymbol{\lambda}_t(\boldsymbol{\theta})\}] = E_0 l(\boldsymbol{\lambda}_0^*; \boldsymbol{u}_t) = E_0 l\{\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0^*); \boldsymbol{u}_t\}, \end{aligned}$$

by the independence between $\boldsymbol{\lambda}_t(\boldsymbol{\theta})$ and $\boldsymbol{u}_t$ for any $\boldsymbol{\theta}$, and by the definition of $\boldsymbol{\lambda}_0^*$. Thus, the property is established. □

We now turn to the asymptotic properties of a PML estimator of the form

$$\hat{\boldsymbol{\theta}}_T = \arg\max_{\boldsymbol{\theta}\in\Theta} \sum_{t=1}^{T} l\{\boldsymbol{\lambda}_t(\boldsymbol{\theta}); \boldsymbol{u}_t\}.$$

Proposition 3.3 *Under Assumption* **A0***, the assumptions of Propositions 3.1 and 3.2, and other regularity conditions (see Appendix A.1), the PML estimator $\hat{\boldsymbol{\theta}}_T$ converges a.s. to $\boldsymbol{\theta}_0^*$ and is asymptotically normal:*

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0^*\right) \xrightarrow{d} N\left(0, \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right).$$

Proof: See Appendix B.2.

It should be noted that $\boldsymbol{\theta}_0^*$ does not necessarily coincide with the "true value" of the econometric model. This will be illustrated in the forthcoming sections.

3.2 The generic model

Let us first consider observations $\tilde{\boldsymbol{y}}_t, t = 1, \ldots, T$, satisfying the generic model (2.2) with a true value $a_0 \in \Theta_\alpha$, a true error distribution g_0, and a pseudo distribution g for the errors u_t. The pseudo log-likelihood for one observation is:

$$l(a, \tilde{\boldsymbol{y}}) \quad = \quad \log g[\exp(a\boldsymbol{C})\tilde{\boldsymbol{y}}] + \log\det[\exp(a\boldsymbol{C})] = \log g[\exp(a\boldsymbol{C})\tilde{\boldsymbol{y}}] + a\mathrm{Tr}(\boldsymbol{C}), \tag{3.6}$$

since:

$$\log\det[\exp(a\boldsymbol{C})] = \log\prod_{i=1}^{n}\exp(a\lambda_i) = a\sum_{i=1}^{n}\lambda_i = a\mathrm{Tr}(\boldsymbol{C}),$$

where $\lambda_i, i = 1, \ldots, n$, are the eigenvalues of $\boldsymbol{C}$ [see Arsigny et al. (2007), Prop. 2.4].

The pseudo maximum likelihood estimator of a in the generic model is solution of:

$$\hat{a}_T = \arg\max_{a\in\Theta_\alpha} \frac{1}{T}\sum_{t=1}^{T} l(a; \tilde{\boldsymbol{y}}_t).$$

Under standard regularity conditions, it converges to a pseudo true value a_0^* solution of the asymptotic optimization problem:

$$a_0^* = \arg\max_{a\in\Theta_\alpha} E_0 l(a; \tilde{\boldsymbol{y}}) = \arg\max_{a\in\Theta_\alpha} E_0 l(a - a_0; \boldsymbol{u}), \tag{3.7}$$

where E_0 is the expectation w.r.t. the true distribution of $\tilde{\boldsymbol{y}}$ (resp. $\boldsymbol{u}$).

This pseudo true value[11] depends on the true value a_0, and on the pseudo and true distributions of u. Note that:

$$\overline{a}_0 := a_0^* - a_0$$

[11] This pseudo true value is a notion of mean on the manifold for the cost function corresponding to the pseudo log-likelihood [Karcher (1977)].

maximizes $E_0 l(a;u)$ and only depends on the true distribution of u. We introduce an identification condition in this generic pseudo-model.

Assumption A1 [Identification of the pseudo-true value in the generic model]: There is a unique solution a_0^* to the asymptotic optimization (3.7), where $\tilde{\boldsymbol{y}}$ satisfies (2.2) with true distribution g_0 for error u and true value a_0.

3.3 The pseudo-likelihood and the pseudo-true values

Let us now consider observations $(\boldsymbol{x}_t, \boldsymbol{y}_t), t = 1, \ldots, T$, satisfying the econometric model (2.1) and a pseudo-distribution g for the errors. The pseudo log-likelihood function is:

$$L_T(\boldsymbol{\beta}, \alpha) = \sum_{t=1}^{T} \{\log g[\exp[(a(\boldsymbol{x}_t; \boldsymbol{\beta}) + \alpha)\boldsymbol{C}]\boldsymbol{y}_t] + [a(\boldsymbol{x}_t; \boldsymbol{\beta}) + \alpha]\mathrm{Tr}(\boldsymbol{C})\} := \sum_{t=1}^{T} l\{a(\boldsymbol{x}_t, \boldsymbol{\beta}) + \alpha; \boldsymbol{y}_t\}.$$

The pseudo maximum likelihood estimator of $(\boldsymbol{\beta}_0, \alpha_0)$ in Model (2.1) is solution of:

$$(\hat{\boldsymbol{\beta}}_T, \hat{\alpha}_T) = \arg\max_{\boldsymbol{\beta}, \alpha} L_T(\boldsymbol{\beta}, \alpha).$$

Under standard regularity conditions to be detailed below, the standardized pseudo log-likelihood function tends almost surely (a.s.) to its asymptotic theoretical counterpart:

$$\frac{1}{T} L_T(\boldsymbol{\beta}, \alpha) \xrightarrow[T \to \infty]{} \tilde{L}_\infty(\boldsymbol{\beta}, \alpha) := E_0 l[a(x; \boldsymbol{\beta}) + \alpha, y].$$

Then we introduce a second identification condition.

Assumption A2 [Identification of the pseudo-true value in the econometric model]: There exists a unique solution to the asymptotic optimization problem:

$$(\boldsymbol{\beta}_0^*, \alpha_0^*) = \arg\max_{\boldsymbol{\beta} \in \Theta_{\boldsymbol{\beta}}, \alpha \in \Theta_\alpha} \tilde{L}_\infty(\boldsymbol{\beta}, \alpha).$$

This assumption can be replaced by the following more primitive identification condition:

Assumption A2': for any $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, there exists no constant K such that:

$$a(\boldsymbol{x}_t; \boldsymbol{\beta}_0) - a(\boldsymbol{x}_t; \boldsymbol{\beta}) = K, \quad a.s.$$

We also introduce the following standard assumption.

Assumption A3 [Stationarity and independence]: The process $(\boldsymbol{u}_t, \boldsymbol{x}_t)$ is a strictly stationary and ergodic process. The sequence $(\boldsymbol{u}_t)$ is i.i.d., and $\boldsymbol{u}_t$ is independent of $\boldsymbol{x}_t$.

In particular, $\boldsymbol{x}_t$ may be equal to the past $(\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \ldots)$ of $\boldsymbol{y}_t$.

3.4 Consistency of the PML estimator

We have the following result:

Proposition 3.4 *Under Assumptions* **A1-A3** *(or* **A1-A2'-A3***), and other regularity conditions (see Appendix A.2), the PML estimator of* $\boldsymbol{\beta}$ *converges (a.s.) to the true value* $\boldsymbol{\beta}_0^* = \boldsymbol{\beta}_0$*, whereas the PML estimator of* α *converges (a.s) to* $\alpha_0^* = \alpha_0 + \overline{a}_0$*.*

Proof: We will show that Assumption **A0** is satisfied, allowing us to derive the consistency result. By the ergodic theorem, the pseudo log-likelihood converges almost surely to the limit criterion. We are looking for the solution of:

$$\underset{\boldsymbol{\beta}\in\Theta_{\boldsymbol{\beta}},\alpha\in\Theta_\alpha}{\arg\max}\; E_0 l[a(\boldsymbol{x}_t;\boldsymbol{\beta})+\alpha;\boldsymbol{y}_t] = \underset{\boldsymbol{\beta}\in\Theta_{\boldsymbol{\beta}},\alpha\in\Theta_\alpha}{\arg\max}\; E_0 l[a(\boldsymbol{x}_t;\boldsymbol{\beta})-a(\boldsymbol{x}_t;\boldsymbol{\beta}_0)+\alpha-\alpha_0;\boldsymbol{u}_t], \tag{3.8}$$

where $l(\cdot)$ is defined in (3.6). Using the notations of Section 3.1, we thus have

$$\boldsymbol{\lambda}_t(\boldsymbol{\theta}) = \boldsymbol{\lambda}(\boldsymbol{x}_t;\boldsymbol{\theta}) = a(\boldsymbol{x}_t;\boldsymbol{\beta}) - a(\boldsymbol{x}_t;\boldsymbol{\beta}_0) + \alpha - \alpha_0, \tag{3.9}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}',\alpha)'$. We also have $\boldsymbol{\lambda}_0^* = \overline{a}_0 = a_0^* - a_0$, by (3.7). Thus

$$\boldsymbol{\lambda}(\boldsymbol{x}_t;\boldsymbol{\theta}_0^*) = \boldsymbol{\lambda}_0^* \quad \text{for} \quad \boldsymbol{\beta}_0^* = \boldsymbol{\beta}_0 \quad \text{and} \quad \alpha_0^* = \alpha_0 + \overline{a}_0.$$

By Proposition 3.2, the limit criterion in (3.8) is maximized at $\boldsymbol{\theta}_0^* = (\boldsymbol{\beta}_0^{*'},\alpha_0^*)'$. To prove the uniqueness of the argmax, suppose there exists $\boldsymbol{\theta}$ such that

$$a(\boldsymbol{x}_t;\boldsymbol{\beta}) - a(\boldsymbol{x}_t;\boldsymbol{\beta}_0) + \alpha - \alpha_0 = \overline{a}_0, \quad a.s.$$

By **A2'**, this entails $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, and thus $\alpha = \alpha_0^*$. The consistency of the PML estimator follows by applying Proposition 3.3. □

Proposition 3.4 shows that the asymptotic bias due to the misspecification of the error distribution is fully captured by the velocity "intercept" parameter α, and this result is valid for any selected pseudo-distribution g. This proposition extends the similar results derived in Newey, Steigerwald (1997). In the context of GARCH models, Fan, Qi, Xiu (2014) proposed a three-step procedure for correcting the bias due to the use of a non-Gaussian pseudo-likelihood function. Alternatively, the asymptotic bias in the PML estimation of GARCH models can be removed by imposing appropriate identifiability assumptions on the innovations distribution [see Berkes and Horváth (2004), Francq, Lepage and Zakoian (2011)].

Proposition 3.4 applies to the econometric model (2.1) with a velocity intercept parameter α. When such a velocity intercept parameter is not in the initial model, an additional intercept has to be artificially introduced to capture the asymptotic bias and allow for consistent estimation of parameter $\boldsymbol{\beta}$.

Since the consistency result for $\hat{\boldsymbol{\beta}}_T$ is valid for any pseudo distribution, we still have consistency if the pseudo-distribution is chosen in a parametric family $g(u;\nu)$, say, and the pseudo-likelihood is maximized with respect to $\alpha, \boldsymbol{\beta}$, and ν. This allows for choosing pseudo-families of distributions with special features expected on the innovations, such as asymmetries [see e.g. Calzolari et al. (2000), Harvey, Siddique (1999) for the use of Student-t family in ARCH type models].

Example 2.2.i) (continued). Let us consider the two following univariate ARCH(1)-type models:

$$
\begin{aligned}
y_t &= \exp[-(\beta y_{t-1} + \alpha)] u_t, \\
\text{and} \quad y_t &= \left(\gamma_1 y_{t-1}^2 + \gamma_2\right)^{1/2} u_t, \quad \gamma_1 > 0, \gamma_2 > 0.
\end{aligned}
$$

In the first ARCH type specification, the model is directly written under the form (2.1) with a velocity intercept and Proposition 3.4 directly applies. The second ARCH specification, which is a standard ARCH(1), can be rewritten as:

$$
y_t = \gamma_2^{1/2} \left\{(\gamma_1/\gamma_2) y_{t-1}^2 + 1\right\}^{1/2} u_t \quad = \quad \exp\left[\frac{1}{2}\log\{(\gamma_1/\gamma_2) y_{t-1}^2 + 1\} + \frac{1}{2}\log\gamma_2\right] u_t.
$$

This second specification involves a velocity intercept $\alpha = -\frac{1}{2}\log\gamma_2$ and a parameter $\beta = \gamma_1/\gamma_2$. Proposition 3.4 can be applied after an appropriate change of parameters. Therefore, γ_2, γ_1 are consistently estimated by PML up to a multiplicative factor.

To understand the consistency result of Proposition 3.4, let us consider the asymptotic first-order conditions (FOC) for a differentiable pseudo p.d.f. g. Let us denote:

$$
\gamma(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\beta}, \alpha) = \frac{\partial \log g}{\partial \boldsymbol{u}'}\{\exp([a(\boldsymbol{x};\boldsymbol{\beta}) + \alpha]\boldsymbol{C})\boldsymbol{y}\}\boldsymbol{C}\exp([a(\boldsymbol{x};\boldsymbol{\beta}) + \alpha]\boldsymbol{C})\boldsymbol{y} + \mathrm{Tr}(\boldsymbol{C}), \tag{3.10}
$$

$$
\tilde{\gamma}_0(\boldsymbol{u}_t; \overline{a}_0) = \gamma(\boldsymbol{x}_t, \boldsymbol{y}_t; \boldsymbol{\beta}_0, \alpha_0^*) = \frac{\partial \log g}{\partial \boldsymbol{u}'}[\exp[\overline{a}_0\boldsymbol{C}]\boldsymbol{u}_t]\boldsymbol{C}\exp[\overline{a}_0\boldsymbol{C}]\boldsymbol{u}_t + \mathrm{Tr}(\boldsymbol{C}), \tag{3.11}
$$

noting that $\overline{a}_0 = \alpha_0^* - \alpha_0$.

Corollary 3.1 *Under the assumptions of Proposition 1, and if the pseudo density g is differentiable, the asymptotic FOC for $\boldsymbol{\beta}_0$ can be written as*

$$
Cov_0[\tilde{\gamma}_0(\boldsymbol{u}_t; \overline{a}_0), \frac{\partial a}{\partial \boldsymbol{\beta}}(\boldsymbol{x}_t; \boldsymbol{\beta}_0)] = 0. \tag{3.12}
$$

Moreover,

$$
\alpha_0^* = \alpha_0 \quad \Longleftrightarrow \quad Tr\left\{\boldsymbol{C}\left(\boldsymbol{I}_n + E_0\left[\boldsymbol{u}_t \frac{\partial \log g}{\partial \boldsymbol{u}'}(\boldsymbol{u}_t)\right]\right)\right\} = 0,
$$

where $\boldsymbol{I}_n$ is the n-dimensional identity matrix.

Proof: See Appendix B.3.

When g is the standard Gaussian density (Gaussian PML), the latter equality reduces to $Tr\left\{\boldsymbol{C}\left[E_0\left(\boldsymbol{u}_t\boldsymbol{u}_t'\right) - \boldsymbol{I}_n\right]\right\} = 0$.

The asymptotic FOC for $\boldsymbol{\beta}_0$ implies that the PML estimator $\hat{\boldsymbol{\beta}}_T$ is a covariance estimator based on covariance restrictions of the type (3.12) between functions of $\boldsymbol{u}$ to be estimated and known functions of $\boldsymbol{x}$. Such an estimator is called a covariance estimator [see e.g. Gouriéroux and Jasiak (2016)].

3.5 Asymptotic distribution of the PML estimator

The asymptotic theory for PML estimators was initially developed in the i.i.d. setting (see e.g. White (1994), Gouriéroux, Monfort (1995), chap. 24). In our framework, by assuming independence between $\boldsymbol{u}_t$ and the past of $\boldsymbol{x}_t$, the asymptotic distribution of the PML estimator follows from the property of martingale difference of the pseudo score.

Proposition 3.5 *Under the assumptions of Proposition 3.4 and other regularity conditions (see Appendix A.3), the PML estimator is asymptotically normal:*

$$\sqrt{T}\begin{pmatrix} \hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0 \\ \hat{\alpha}_T - \alpha_0^* \end{pmatrix} \xrightarrow{d} N\left(0, \frac{i^*}{(j^*)^2}\boldsymbol{\Sigma}^{-1}\right),$$

where:

$$\boldsymbol{\Sigma} = E_0\left[\begin{pmatrix} \frac{\partial a(\boldsymbol{x}_t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \\ 1 \end{pmatrix}\begin{pmatrix} \frac{\partial a(\boldsymbol{x}_t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \\ 1 \end{pmatrix}'\right], \quad i^* = V_0[\tilde{\gamma}_0(\boldsymbol{u}_t;\overline{a}_0)],$$

$$j^* = E_0\left\{-\left(\frac{\partial \log g}{\partial \boldsymbol{u}'}[\exp\{\overline{a}_0\boldsymbol{C}\}\boldsymbol{u}_t]\right)\boldsymbol{C}^2\exp\{\overline{a}_0\boldsymbol{C}\}\boldsymbol{u}_t \right.$$
$$\left. - \boldsymbol{u}_t'\exp\{\overline{a}_0\boldsymbol{C}'\}\boldsymbol{C}'\left(\frac{\partial^2 \log g}{\partial \boldsymbol{u}\partial \boldsymbol{u}'}[\exp\{\overline{a}_0\boldsymbol{C}\}\boldsymbol{u}_t]\right)\boldsymbol{C}\exp\{\overline{a}_0\boldsymbol{C}\}\boldsymbol{u}_t\right\}.$$

In particular,

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0) \xrightarrow{d} N\left(0, \frac{i^*}{(j^*)^2}\left(V_0\left[\frac{\partial a(\boldsymbol{x}_t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}}\right]\right)^{-1}\right).$$

Proof: See Appendix B.4.

It is easily seen from the expressions of i^* and j^* that the ratio i^*/j^{*2} only depends on the true distribution g_0 of u and on the choice of the pseudo-distribution g (neither on the explanatory variables, nor on the pattern of index function a, nor on the true intercept value α_0). This ratio is the asymptotic variance of the PML estimator of a_0 in the generic model. It depends on moments of the adjusted error terms:

$$\boldsymbol{v}_t = \exp[\overline{a}_0\boldsymbol{C}]\boldsymbol{u}_t.$$

These adjusted error terms are consistently estimated by the adjusted residuals:

$$\hat{\boldsymbol{v}}_t = \exp([a(\boldsymbol{x}_t;\hat{\boldsymbol{\beta}}_T) + \hat{\alpha}_T]\boldsymbol{C})\boldsymbol{y}_t,$$

and i^*, j^* by their sample counterparts after substituting the residuals to the errors.

One extension of the present framework would consist in parameterizing the pseudo-density g by some parameter vector $\boldsymbol{\gamma}$, say. The asymptotic variances of the PML estimators, corresponding to different values of $\boldsymbol{\gamma}$, would all be proportional to $\boldsymbol{\Sigma}^{-1}$, with proportionality constants $i^*(\boldsymbol{\gamma})/\{j^*(\boldsymbol{\gamma})\}^2$, and obvious notation. The next step would consist in minimizing the later ratio with respect to $\boldsymbol{\gamma}$, as it was done in Francq, Lepage and Zakoian (2011) in the context of GARCH models estimated via non-Gaussian QML.

4 Multi-index model

The econometric model in Section 2 depends on a single index $a(x; \boldsymbol{\beta})$. We discuss in this section the extensions to multi-index models. We consider the case of commuting Lie groups of linear transformations and a semi-parametric transformation model, where the transformation is parameterized, but the error distribution is unconstrained. Then we discuss a case where the commutativity property is not satisfied, but where the Lie group satisfies a closure under commutation property. [12]

4.1 Multi-index model

The approach of Section 2 can be extended to models such as:

$$\boldsymbol{y}_t = \exp\left\{-\sum_{j=1}^{J}[a_j(\boldsymbol{x}_t, \boldsymbol{\beta}_0) + \alpha_{0j}]\boldsymbol{C}_j\right\}\boldsymbol{u}_t, \tag{4.1}$$

with several indexes a_j and commuting generators $\boldsymbol{C}_j$, $j = 1, \ldots, J$. In particular J is the dimension of the Lie group, equal to the dimension of the associated differentiable manifold[13].

The generic model is:

$$\tilde{\boldsymbol{y}}_t = \exp\left(-\sum_{j=1}^{J} a_{0j}\boldsymbol{C}_j\right)\boldsymbol{u}_t,$$

and the pseudo log-likelihood for one observation is:

$$l(\boldsymbol{a}, \tilde{\boldsymbol{y}}) \quad = \quad \log g\left[\exp\left(\sum_{j=1}^{J} a_j\boldsymbol{C}_j\right)\tilde{\boldsymbol{y}}\right] + \sum_{j=1}^{J} a_j \mathrm{Tr}(\boldsymbol{C}_j), \tag{4.2}$$

where $\boldsymbol{a} = (a_1, \ldots, a_J)'$. Let $\boldsymbol{a}_0 = (a_{01}, \ldots, a_{0J})'$. We still denote by Θ_α the parameter set for $\boldsymbol{a}$. The identification conditions become:

Assumption A1*: There is a unique solution $\boldsymbol{a}_0^* = (a_{0j}^*)_{j=1,\ldots,J}$ to the asymptotic optimization:

$$\boldsymbol{a}_0^* = \arg\max_{\boldsymbol{a}\in\Theta_\alpha} E_0 l(\boldsymbol{a}; \tilde{\boldsymbol{y}}) = \arg\max_{a\in\Theta_\alpha} E_0 l(\boldsymbol{a} - \boldsymbol{a}_0; \boldsymbol{u}),$$

with true distribution g_0 for the error $\boldsymbol{u}$.

Again, $\overline{\boldsymbol{a}}_0 := \boldsymbol{a}_0^* - \boldsymbol{a}_0$ only depends on the true distribution of u.

Assumption A2*: There is a unique solution to the optimization problem:

$$(\boldsymbol{\beta}_0^*, \boldsymbol{\alpha}_0^*) = \arg\max_{\boldsymbol{\beta}\in\Theta_{\boldsymbol{\beta}}, \boldsymbol{\alpha}\in\Theta_\alpha} \tilde{L}_\infty(\boldsymbol{\beta}, \boldsymbol{\alpha}),$$

where $\tilde{L}_\infty$ is the asymptotic pseudo-likelihood of Model (4.1), and $\boldsymbol{\alpha} = (\alpha_j)_{j=1,\ldots,J}$.

[12] See Gelfand et al. (1962), Boothby (1975), Hull (2015) for introductory courses on continuous Lie groups and closure under commutation.

[13] when the generators $\boldsymbol{C}_j$ are linearly independent

4.2 Consistency of the PML estimator

The pseudo log-likelihood function for Model (4.1) is:

$$L_T(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{t=1}^{T} l\{\boldsymbol{a}(\boldsymbol{x}_t, \boldsymbol{\beta}) + \boldsymbol{\alpha}; \boldsymbol{y}_t\},$$

where $l(\cdot)$ is defined in (4.2) and $\boldsymbol{a}(x;\boldsymbol{\beta}) = [a_1(x;\boldsymbol{\beta}), \ldots, a_J(x;\boldsymbol{\beta})]'$. The pseudo maximum likelihood estimator of $(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)$ is solution of:

$$(\hat{\boldsymbol{\beta}}_T, \hat{\boldsymbol{\alpha}}_T) = \arg\max_{\boldsymbol{\beta},\boldsymbol{\alpha}} L_T(\boldsymbol{\beta}, \boldsymbol{\alpha}).$$

Proposition 4.1 *Under Assumptions* **A1*, A2*** *(or* **A2'***) and the regularity conditions of Proposition 3.4, when the generators $\boldsymbol{C}_j, j = 1, \ldots, J$ commute in Model (4.1), the PML estimator of $\boldsymbol{\beta}_0$ converges (a.s.) to the true value $\boldsymbol{\beta}_0^* = \boldsymbol{\beta}_0$ whereas the PML estimator of $\boldsymbol{\alpha}$ converges (a.s) to $\boldsymbol{\alpha}_0^* = \boldsymbol{\alpha}_0 + \overline{\boldsymbol{a}}_0$.*

Proof: See Appendix B.5.

When the initial econometric model does not include "intercept" parameters, the "initial model":

$$\boldsymbol{y}_t = \exp\left\{-\sum_{j=1}^{J} a_j(\boldsymbol{x}_t;\boldsymbol{\beta}_0)\boldsymbol{C}_j\right\}\boldsymbol{u}_t,$$

is extended by substituting for $\boldsymbol{u}_t$ the transformation of the generic model to get:

$$\boldsymbol{y}_t \;=\; \exp\left\{-\sum_{j=1}^{J} a_j(\boldsymbol{x}_t;\boldsymbol{\beta}_0)\boldsymbol{C}_j\right\}\exp\left\{-\sum_{j=1}^{J}\alpha_{0j}\boldsymbol{C}_j\right\}\boldsymbol{u}_t = \exp\left\{-\sum_{j=1}^{J}(a_j(\boldsymbol{x}_t;\boldsymbol{\beta}_0)+\alpha_{0j})\boldsymbol{C}_j\right\}\boldsymbol{u}_t,$$

since the $\boldsymbol{C}_j, j = 1, \ldots, J$ commute.[14]

As in the single index model, the PML estimator of $\boldsymbol{\beta}$ has the interpretation of a covariance estimator. Indeed the pseudo-true value $\boldsymbol{\beta}_0^*$ satisfies the condition:

$$\sum_{j=1}^{J} Cov_0\left[\gamma_j(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\beta}_0^*,\alpha_0^*), \frac{\partial a_j(x;\boldsymbol{\beta}_0^*)}{\partial\boldsymbol{\beta}}\right] = 0, \tag{4.3}$$

where:

$$\gamma_j(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\beta}_0^*,\alpha_0^*) = \frac{\partial \log g}{\partial \boldsymbol{u}'}\left[\exp\left\{\sum_{k=1}^{J}(a_k(\boldsymbol{x};\boldsymbol{\beta}_0^*)+\alpha_{0k}^*)\boldsymbol{C}_k\right\}\boldsymbol{y}\right]\boldsymbol{C}_j\exp\left\{\sum_{k=1}^{J}(a_k(\boldsymbol{x};\boldsymbol{\beta}_0^*)+\alpha_{0k}^*)\boldsymbol{C}_k\right\}\boldsymbol{y}.$$

When $\boldsymbol{\beta}_0^* = \boldsymbol{\beta}_0$, we get by using (4.1):

$$\gamma_j(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\beta}_0,\alpha_0^*) = \frac{\partial \log g}{\partial \boldsymbol{u}'}\left[\exp\left\{\sum_{k=1}^{J}(\alpha_{0k}^*-\alpha_{0k})\boldsymbol{C}_k\right\}\boldsymbol{u}\right]\boldsymbol{C}_j\exp\left\{\sum_{k=1}^{J}(\alpha_{0k}^*-\alpha_{0k})\boldsymbol{C}_k\right\}\boldsymbol{u},$$

which depends on $\boldsymbol{u}$ only. Therefore the covariance restrictions (5.2) are automatically satisfied by the independence between $\boldsymbol{x}$ and $\boldsymbol{u}$.

[14] Indeed, if two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ commute, $\exp(\boldsymbol{A})\exp(\boldsymbol{B}) = \exp(\boldsymbol{A}+\boldsymbol{B})$.

Proposition 4.2 *Under the assumptions of Proposition 4.1 and other regularity conditions (see Appendix A.4), the PML estimator is asymptotically normal:*

$$\sqrt{T}\begin{pmatrix} \hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0 \\ \\ \hat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_0^* \end{pmatrix} \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\right),$$

where:

$$\boldsymbol{B} = E_0\left[\begin{pmatrix} \frac{\partial \boldsymbol{a}'(\boldsymbol{x}_t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \\ \boldsymbol{I}_J \end{pmatrix} \boldsymbol{K} \begin{pmatrix} \frac{\partial \boldsymbol{a}'(\boldsymbol{x}_t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \\ \boldsymbol{I}_J \end{pmatrix}'\right], \qquad \boldsymbol{A} = E_0\left[\begin{pmatrix} \frac{\partial \boldsymbol{a}'(\boldsymbol{x}_t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \\ \boldsymbol{I}_J \end{pmatrix} \boldsymbol{L} \begin{pmatrix} \frac{\partial \boldsymbol{a}'(\boldsymbol{x}_t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \\ \boldsymbol{I}_J \end{pmatrix}'\right],$$

where the $J \times J$ matrices $\boldsymbol{K}$ and $\boldsymbol{L}$ only depend on the true and pseudo densities of $\boldsymbol{u}_t$ and are displayed in the proof.

Proof: See Appendix B.6.

4.3 Examples

i) Multivariate regression model and CCC model

Consistent PML approaches can be applied to multivariate regression models:

$$\boldsymbol{y}_t^* = \boldsymbol{a}(\boldsymbol{x}_t;\boldsymbol{\beta}_0) + \boldsymbol{\alpha}_0 + \boldsymbol{u}_t^*, \tag{4.4}$$

where $\boldsymbol{y}^*, \boldsymbol{u}^*, \boldsymbol{a}, \boldsymbol{\alpha}_0$ have dimension J and the $\boldsymbol{u}_t^{*'}s$ are i.i.d. Indeed this model can equivalently be written as:

$$\boldsymbol{y}_t = \exp\{\text{diag}[-(\boldsymbol{a}(\boldsymbol{x}_t;\boldsymbol{\beta}_0) + \boldsymbol{\alpha}_0)]\}\boldsymbol{u}_t, \tag{4.5}$$

where $y_{j,t} = \exp(-y_{j,t}^*), u_{j,t} = \exp(-u_{j,t}^*)$, and the diagonal matrix has entries $\exp[-\{a_j(\boldsymbol{x}_t;\boldsymbol{\beta}_0) + \alpha_{0j}\}]$ on the diagonal. Model (4.5) is a special case of multi-index model (4.1) with $\boldsymbol{C}_j = \text{diag}(\boldsymbol{e}_j), j = 1,\ldots,J$, and $\boldsymbol{e}_j$ is the canonical vector with j^{th} component equal to 1, and zero components, otherwise. Propositions 4.1-4.2 can be applied to (4.4), since the errors $\boldsymbol{u}_t$ are i.i.d. as are the errors $\boldsymbol{u}_t^*$, and since the pseudo-likelihoods written on (4.4) and (4.5), respectively, differ by a Jacobian, which is independent of the parameters.

Model (4.5) can also be directly considered as a multivariate model with conditional heteroscedasticity. Since the variance-covariance matrix of u is not constrained, we get a pure Constant Conditional Correlation (CCC) model [Bollerslev (1990), Tse (2000)]. Thus the PML approach is consistent for any pseudo joint distribution of u, once J scale parameters $\sigma_j = \exp\alpha_j, j = 1,\ldots,J$, are added.

ii) A mix rotation/homothety

In the bidimensional case $n = 2$ the only generators which commute with the rotation generator $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ are of the form $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$, that are linear combinations of the rotation generator and the identity. This leads to an econometric model with two index functions a_1, a_2 such that:

$$\boldsymbol{y}_t \equiv \begin{pmatrix} \cos[a_1(\boldsymbol{x}_t;\theta)] & \sin[a_1(\boldsymbol{x}_t;\theta)] \\ \\ -\sin[a_1(\boldsymbol{x}_t;\theta)] & \cos[a_1(\boldsymbol{x}_t;\theta)] \end{pmatrix} \exp[-a_2(\boldsymbol{x}_t;\theta)]\boldsymbol{u}_t. \tag{4.6}$$

4.4 A Multivariate Extension of Newey-Steigerwald

Two characteristics of the multi-index model (4.1) explain the consistency of the PML estimator of the $\boldsymbol{\beta}$ coefficient: i) the Lie group property of each component, that provides the simple form of the Jacobian in the expression of the pseudo-likelihood, and ii) the fact that these groups commute. By considering a multivariate extension of Newey-Steigerwald, we will see that the commutativity property can be weakened. The bias adjustment in Newey, Steigerwald (1997) is written for a one-dimensional model and involves two Lie groups, that are the group of homotheties and the group of translations, respectively. These operators are:

$$S_1(u;m) = u + m, \qquad S_2(u;\sigma) = \sigma u. \tag{4.7}$$

We immediately note that:

$$S_1 o S_2(u;m,\sigma) = \sigma u + m \tag{4.8}$$

differs from:

$$S_2 o S_1(u;m,\sigma) = \sigma(u+m). \tag{4.9}$$

Thus, these two operators do not commute. The group of translations is not a group of linear transformations, but has no effect on the Jacobian of any pseudo-likelihood.

Let us now describe the different models underlying the Newey-Steigerwald framework.

i) The initial econometric model

is defined, for $\boldsymbol{m}(\boldsymbol{x}_t;\boldsymbol{\beta}_0) \in \mathbb{R}^n$, by:

$$\boldsymbol{y}_t = \boldsymbol{m}(\boldsymbol{x}_t;\boldsymbol{\beta}_0) + \exp[-a(\boldsymbol{x}_t;\boldsymbol{\beta}_0)\boldsymbol{C}]\boldsymbol{u}_t. \tag{4.10}$$

This is a multi-index model, with indexes $\boldsymbol{m}(\boldsymbol{x}_t;\boldsymbol{\beta}_0)$ and $a(\boldsymbol{x}_t;\boldsymbol{\beta}_0)$.

ii) The enlarged econometric model

is defined by:

$$\boldsymbol{y}_t = \boldsymbol{m}(\boldsymbol{x}_t;\boldsymbol{\beta}_0) + \exp[-a(\boldsymbol{x}_t;\boldsymbol{\beta}_0)\boldsymbol{C}][\boldsymbol{\mu}_0 + \exp(-\alpha_0\boldsymbol{C})\boldsymbol{u}_t], \tag{4.11}$$

$\boldsymbol{\mu}_0 \in \mathbb{R}^n$ and $\alpha_0 \in \mathbb{R}$ are the parameters introduced to capture the bias on $\boldsymbol{\beta}_0 \in \mathbb{R}^p$.

iii) The generic pseudo-model

is defined by:

$$\tilde{\boldsymbol{y}}_t = \boldsymbol{m}_0 + \exp(-a_0\boldsymbol{C})\boldsymbol{u}_t. \tag{4.12}$$

The pseudo log-likelihood for one observation in the generic model is:

$$\begin{aligned} l(a,\boldsymbol{m};\tilde{\boldsymbol{y}}) &= \log g\{\exp(a\boldsymbol{C})(\tilde{\boldsymbol{y}}-\boldsymbol{m})\} + a\mathrm{Tr}(\boldsymbol{C}) \\ &= \log g[\exp\{(a-a_0)\boldsymbol{C}\}\boldsymbol{u} + \exp(a\boldsymbol{C})(\boldsymbol{m}_0-\boldsymbol{m})] + a\mathrm{Tr}(\boldsymbol{C}). \end{aligned} \tag{4.13}$$

For parameter sets $\Theta_\alpha \subset \mathbb{R}$ and $\Theta_m \subset \mathbb{R}^n$ let

$$(a_0^*,\boldsymbol{m}_0^*) = \underset{a\in\Theta_\alpha,\boldsymbol{m}\in\Theta_m}{\arg\max}\; E_0 l(a,\boldsymbol{m};\tilde{\boldsymbol{y}}) = \underset{a\in\Theta_\alpha,\boldsymbol{m}\in\Theta_m}{\arg\max}\; E_0 l\{a-a_0,\exp(a_0\boldsymbol{C})(\boldsymbol{m}-\boldsymbol{m}_0);\boldsymbol{u}\}. \tag{4.14}$$

Note that $\overline{a}_0 := a_0^* - a_0, \overline{\boldsymbol{m}}_0 := \boldsymbol{m}_0^* - \boldsymbol{m}_0$ only depend on the true distribution of u.

Assumption A1 [Identification of the pseudo-true value in the generic model]:** There is a unique solution $(a_0^*, \boldsymbol{m}_0^*)$ to the asymptotic optimization (4.14), where $\tilde{\boldsymbol{y}}$ satisfies (4.12) with the true distribution g_0 for error $\boldsymbol{u}$.

The pseudo log-likelihood function for Model (4.11) is:

$$L_T(\boldsymbol{\beta}, \alpha, \boldsymbol{\mu}) = \sum_{t=1}^{T} l\{a(\boldsymbol{x}_t, \boldsymbol{\beta}) + \alpha, \boldsymbol{m}(\boldsymbol{x}_t; \boldsymbol{\beta}) + \exp[-a(\boldsymbol{x}_t; \boldsymbol{\beta})\boldsymbol{C}]\boldsymbol{\mu}; \boldsymbol{y}_t\},$$

where $l(\cdot)$ is defined in (4.13). The PML estimator of $(\boldsymbol{\beta}_0, \alpha_0, \boldsymbol{\mu}_0)$ is solution of:

$$(\hat{\boldsymbol{\beta}}_T, \hat{\alpha}_T, \hat{\boldsymbol{\mu}}_T) = \arg\max_{\boldsymbol{\beta}, \alpha, \boldsymbol{\mu}} L_T(\boldsymbol{\beta}, \alpha, \boldsymbol{\mu}).$$

Proposition 4.3 *Under Assumptions* **A1****, **A2'**, **A3** *and other regularity conditions (see Appendix A.5), the PML estimator of* $\boldsymbol{\beta}$ *converges (a.s.) to the true value* $\boldsymbol{\beta}_0^* = \boldsymbol{\beta}_0$ *in Model (4.11), whereas the PML estimator of* α *and* $\boldsymbol{\mu}$ *converge (a.s) to* $\alpha_0^* = \alpha_0 + a_0^* - a_0$ *and* $\boldsymbol{\mu}_0^* = \boldsymbol{\mu}_0 + \exp[(a_0 - \alpha_0)\boldsymbol{C}](\boldsymbol{m}_0^* - \boldsymbol{m}_0)$*, respectively.*

Proof: See Appendix B.7.

Proposition 4.4 *Under the assumptions of Proposition 4.3 and other regularity conditions (see Appendix A.6), the PML estimator is asymptotically normal:*

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0^*\right) \xrightarrow{d} N\left(0, \boldsymbol{J}^{-1}\boldsymbol{I}\boldsymbol{J}^{-1}\right),$$

where $\hat{\boldsymbol{\theta}}_T = (\hat{\boldsymbol{\beta}}_T', \hat{\alpha}_T, \hat{\boldsymbol{\mu}}_T)'$, $\boldsymbol{\theta}_0^* = (\boldsymbol{\beta}_0', \alpha_0^*, \boldsymbol{\mu}_0^{*'})'$ *and the* $(n+p+1) \times (n+p+1)$ *matrices* $\boldsymbol{I}$ *and* $\boldsymbol{J}$ *are as in Proposition 3.3.*[15]

Proof: See Appendix B.8.

As noted in Newey and Steigerwald (1997) in the univariate case :

i) two parameters have to be introduced to capture the misspecification bias, μ for the "location" effect, and α for the "scale" effect.

ii) to properly adjust for the location regression $m(\boldsymbol{x}_t; \boldsymbol{\beta})$, the adjustment $+\exp[-a(\boldsymbol{x}_t; \boldsymbol{\beta})\boldsymbol{C}]\mu$, has to account for the scale in an appropriate way.

As seen in the next subsection, the consistency of the PML still holds, since the translation group and the homothety group satisfy a closure under commutation property. Typically from the multivariate extension of (4.7)-(4.8), namely,

$$S_1(\boldsymbol{u}; \boldsymbol{m}) = \boldsymbol{u} + \boldsymbol{m}, \qquad S_2(\boldsymbol{u}; a) = \exp(-a\boldsymbol{C})\boldsymbol{u}. \tag{4.15}$$

we get:

$$S_2 o S_1(\boldsymbol{u}; \boldsymbol{m}, a) = S_1 o S_2(\boldsymbol{u}; \exp(-a\boldsymbol{C})\boldsymbol{m}, a).$$

Even if the two Lie groups do not commute, it is easy to commute the operators after applying an appropriate change of parameters.

[15] The expressions for $\boldsymbol{\lambda}_t(\boldsymbol{\theta})$, $l(\boldsymbol{\lambda}, \boldsymbol{u})$ and their derivatives is provided in the proof.

5 Transformation Models for Normalized Data

The consistency of the PML estimator of $\boldsymbol{\beta}$ in the model of Section 3 is due to the simple expression of the pseudo log-likelihood function:

$$\log g[\exp(a\boldsymbol{C})\boldsymbol{y}] + aTr\boldsymbol{C},$$

to the possibility to (partly) commute the Lie groups within the $\log g(.)$ and to apply the Jacobian formula. This Jacobian formula is valid for a transformation of Lie groups, whenever the transformation is a diffeomorphism. It is valid when the set of values of $\boldsymbol{u}$ (and $\boldsymbol{y}$) is $\mathbb{R}^n$, but also in more complicated cases when $\boldsymbol{u}$ and $\boldsymbol{y}$ are taking values in a same affine subspace [see examples iv) and vi)] or, as in Section 5.2 below, on the unit sphere [see Chirikjian (2011), Vol 2, Chap 11, for the differential geometry of Lie groups]. This is a consequence of the locally Euclidian structure of a manifold.

The consistency results will apply to normalized data with different normalizations. [16] Such normalizations are standard in computer vision, where the images using cameras induce observations obeying mathematical constraints. This arises when the variability due to the viewing angle of the camera has to be taken into account, e.g. when the pose of a human head is parameterized by three angles. Such appropriate normalizations are also required for the gait-bared activity recognition [see e.g. Liu and Sarkar (2006)].[17] We provide below examples of normalized data in Economics and Finance.

5.1 Budget shares

Let us denote $\boldsymbol{y}_t^* = (y_{t1}^*, \ldots, y_{tn}^*)'$ the expenses of a given household at period t, and $\boldsymbol{y}_t = (\boldsymbol{y}_{t1}, \ldots, \boldsymbol{y}_{tn})'$, with $y_{ti} = y_{ti}^* / \sum_{i=1}^{n} y_{ti}^*$ the corresponding budget shares. For such data the transformation model, on $\mathcal{M} = \{\boldsymbol{y} : y_i \geq 0, \forall i, \sum_{i=1}^{n} y_i = 1\}$, is:

$$\boldsymbol{y}_t = \exp[a^*(\boldsymbol{x}_t;\theta)\boldsymbol{C}]\boldsymbol{u}_t,$$

where the matrix generator $\boldsymbol{C}$ is such that:

$$c_{ij} \geq 0, \quad \forall i \neq j, \qquad c_{ii} = -\Sigma_{j \neq i} c_{ij}.$$

Then, the exponential matrix $\exp(a\boldsymbol{C})$ is a stochastic matrix (see also Section 2.2, vi)).

5.2 Zero-cost portfolio allocation

The structure of household expenses is generally analyzed by means of the budget shares instead of the expenses in order to eliminate the size effect. This standardization practice cannot be followed

[16] A part of the statistical literature studies regression models, where the conditional mean belongs to a submanifold, not the observations themselves [see Cheng and Wu (2013)]

[17] from a mathematical point of view, the normalization is defined from an equivalence relationship and the normalized data are elements in the quotient group of a Lie group by the equivalence class [see e.g. Befelgor and Werman (2006)].

in Finance when analyzing the portfolio allocations. Indeed the possibility of short sell allows for negative allocations and, typically, in zero-investment portfolios [18] the total budget is equal to zero. Therefore another normalization is usually considered to eliminate the size effect. If y_i^* denotes the investment in asset $i, i = 1, \dots, n$, possibly negative, it is usual to define the standardized allocation as:

$$y_i = y_i^* / ||\boldsymbol{y}^*||,$$

where[19] : $||\boldsymbol{y}^*||^2 = \sum_{i=1}^{n} y_i^{*2}$. In such a framework the observations $y_i, i = 1, \dots, n$ take values on the unit sphere: $\{\boldsymbol{y} : \sum_{i=1}^{n} y_i^2 = ||\boldsymbol{y}||^2 = 1\}$. Transformation models on the unit sphere (or unit hyper-spherical manifold on $\mathbb{R}^n$) are appropriate to study such data and are based on Lie subgroups of orthogonal transformations, the so-called special orthogonal group $\mathrm{SO}(n)$. Such transformation models on the unit sphere are frequently used in more general shape analyses [see e.g. Bhattacharya and Bhattacharya (2011)].

5.3 Observations of volatility matrices

The set of symmetric positive definite (SPD) matrices is also a Riemannian manifold denoted $\mathrm{Sym}^+(n)$ and our approach can be applied to SPD matrix values data such as observed volatility matrices [see Arsigny et al. (2007), Yuan et al. (2012), Huang et al. (2014), and the references therein for other SPD matrix valued data in the medical imaging literature].

Let us consider a SPD matrix $\boldsymbol{U}$ of size (n, n), another (n, n) matrix $\boldsymbol{B}$, and the application:

$$\boldsymbol{U} \mapsto \exp(a\boldsymbol{B})\boldsymbol{U}\exp(a\boldsymbol{B}') := \boldsymbol{Y}.$$

This defines a Lie group of linear transformations on $\mathrm{Sym}^+(n)$. To link this Lie group with our general specification, let us apply the vec operator. [20] We have:

$$\mathrm{vec}(\boldsymbol{Y}) = \mathrm{vec}\{\exp(a\boldsymbol{B})\boldsymbol{U}\exp(a\boldsymbol{B}')\} = \{\exp(a\boldsymbol{B}) \otimes \exp(a\boldsymbol{B})\}\mathrm{vec}(\boldsymbol{U}), \tag{5.1}$$

where $\otimes$ denotes the Kronecker product [see e.g. Henderson and Searle (1979)]. Let us now consider the behaviour of this transformation when a is close to zero. We get:

$$\{\exp(a\boldsymbol{B}) \otimes \exp(a\boldsymbol{B})\}\mathrm{vec}(\boldsymbol{U}) \sim \{(\boldsymbol{I} + a\boldsymbol{B}) \otimes (\boldsymbol{I} + a\boldsymbol{B})\}\mathrm{vec}(\boldsymbol{U}) \sim \mathrm{vec}(\boldsymbol{U}) + a\{\boldsymbol{B} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{B}\}\mathrm{vec}(\boldsymbol{U}).$$

Therefore, the infinitesimal generator is:

$$\boldsymbol{C} = \boldsymbol{B} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{B}, \tag{5.2}$$

and the transformation (5.1) can be equivalently written as

$$\mathrm{vec}(\boldsymbol{Y}) = \exp\{a(\boldsymbol{B} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{B})\}\mathrm{vec}(\boldsymbol{U}). \tag{5.3}$$

[18] Zero investment strategies are balancing short and long investments and are the basis of arbitrage. Indeed a zero-investment (or zero-cost) portfolio with positive gains is an arbitrage portfolio [see e.g. Alexander (2000) and the references therein]. Such strategies are typically followed by hedge funds reporting the management style Long/Short Equity (hedge).

[19] For a zero-investment portfolio $||\boldsymbol{y}^*||$ measures the magnitude of the leverage.

[20] or its analogue vech for symmetric matrices.

Therefore, the bias adjustment methodology of our paper will apply to Lie groups of such transformations, applied to observed realized volatility matrices, or to observed implied volatility matrices derived from observed derivative prices. Note that the transformation (5.3) for different $\boldsymbol{B}_1, \boldsymbol{B}_2$, say, can be combined if the condition of closure under commutation is satisfied. The approach can also be directly extended to observed 3-dimensional (resp. 4-dimensional) matrices gathering all 3-order cross moments (resp. 4-order cross moments).

6 Concluding Remarks

The objective of our paper was to extend the consistency result of the PML estimator by Newey and Steigerwald to a multivariate framework and to more complex linear transformations. We highlight the key role of Lie groups of linear transformations. At least three research developments seem to be promising.

i) We get a large number of consistent estimators of the sensitivity parameters by changing the pseudo-distributions of the error. Their comparison can be the basis of specification tests on the pattern of the index functions.

ii) The methodology, i.e. the choice of a pseudo-distribution and the introduction of the additional parameters for bias adjustment, is valid for any type of index function and any known transformation of the endogenous variables. Thus, it is valid for models of the type:

$$\boldsymbol{y}_t = H[a(\boldsymbol{x}_t; \boldsymbol{\beta})\boldsymbol{u}_t],$$

where H is a known one-to-one transformation and the distribution of $\boldsymbol{u}_t$ is left unspecified. By increasing the number of parameters in the index functions, we can approach nonparametric transformations of the $\boldsymbol{x}$ variables; therefore the bias adjustment by a fixed number of additional parameters, equal to the number of index functions, is likely to be also valid in the following situation : T is known, $a(\boldsymbol{x}_t)$ is unspecified (but constrained by the Lie group structure) as well as the distribution of the error [see e.g. Chiappori, Komunjer, Kristensen (2015), for the nonparametric estimation of transformation models, when H is also unknown, in the one-dimensional single index framework].

iii) Our paper deals with semi-parametric inference on manifolds. It differs from the statistical literature on manifold-valued observations, which is mainly static [see e.g. Bhattacharya and Bhattacharya (2015)], by the structural introduction of exogenous or lagged endogenous variables and by the focus on robust estimation methods. It might be fruitful to extend the literature in these directions, not only for applications to Finance and Economics, but also to improve the statistical analysis on manifolds currently performed in video analysis or neuroscience [see e.g. Turaga et al. (2010) for a recent survey]. Such applications involve a large number of data $\boldsymbol{y}$ and possibly also a rather large number of parameters $\boldsymbol{\theta}$. However, the number of additional intercept parameters introduced for bias adjustment is equal to the number of index functions, which is usually small.

Appendices

A Regularity conditions

A.1 For Proposition 3.3

Together with the assumptions of Propositions 3.1 and 3.2, we require:

Assumption A01: There is a unique $\boldsymbol{\lambda}_0^*$ satisfying (3.1).

Assumption A02: For any $\boldsymbol{\theta} \in \Theta$, the process $(\boldsymbol{\lambda}_t(\boldsymbol{\theta}), \boldsymbol{u}_t)$ is strictly stationary and ergodic.

Assumption A03: $E_0|l\{\{\boldsymbol{\lambda}_t(\boldsymbol{\theta}); \boldsymbol{u}_t\}\}| < \infty$, for any $\boldsymbol{\theta} \in \Theta$.

Assumption A04: $\boldsymbol{\lambda}_t(\boldsymbol{\theta}_0^*) = \boldsymbol{\lambda}_t(\boldsymbol{\theta}), \quad a.s. \quad \Rightarrow \quad \boldsymbol{\theta} = \boldsymbol{\theta}_0^*$.

Assumption A05: $\boldsymbol{u}_t$ is independent from the $\frac{\partial \boldsymbol{\lambda}'_{t-i}(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}}$, for $i \geq 0$. Moreover, the process $\left(\frac{\partial \boldsymbol{\lambda}'_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}}\right)$ is strictly stationary and ergodic.

Assumption A06: $\boldsymbol{\theta}_0^*$ belongs to the interior of Θ.

Assumption A07: For any $\boldsymbol{x}$, the random function $\boldsymbol{\theta} \to \boldsymbol{\lambda}_t(\boldsymbol{\theta})$ has continuous third-order derivatives. The function $l(\cdot; \boldsymbol{u})$ has continuous third-order derivatives, for any $\boldsymbol{u} \in \mathbb{R}^n$.

Assumption A08: $\boldsymbol{I}^0$ and $\boldsymbol{J}^0$ are nonsingular and rank $\left(\frac{\partial \boldsymbol{\lambda}'_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}}\right) = r, \quad a.s.$

Assumption A09: There exists a neighborhood $V(\boldsymbol{\theta}_0^*)$ of $\boldsymbol{\theta}_0^*$ such that, for $i, j = 1, \ldots, r$,

$$\text{for all } \boldsymbol{\theta} \in V(\boldsymbol{\theta}_0^*), \quad \text{the process } \left\{ \frac{\partial}{\partial \boldsymbol{\theta}'} \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell_t(\boldsymbol{\theta}) \right) \right\} \quad \text{is strictly stationary and ergodic, and,}$$

$$E_0 \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0^*)} \left\| \frac{\partial}{\partial \boldsymbol{\theta}'} \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell_t(\boldsymbol{\theta}) \right) \right\| < \infty.$$

A.2 For Proposition 3.4

Assumption A4: $E_0|l\{a(\boldsymbol{x}_t, \boldsymbol{\beta}) + \alpha; \boldsymbol{y}_t\}| < \infty$, for any $(\alpha, \boldsymbol{\beta}) \in \Theta_\alpha \times \Theta_{\boldsymbol{\beta}}$.

Assumption A5: $\boldsymbol{\beta}_0 \in \Theta_{\boldsymbol{\beta}}$ and $\alpha_0 + \overline{a}_0 \in \Theta_\alpha$. Moreover, $\Theta_{\boldsymbol{\beta}}$ and Θ_α are compact parameter sets.

A.3 For Proposition 3.5

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \alpha)' = (\theta_j)_{1 \leq j \leq p+1}$ and $\ell_t(\boldsymbol{\theta}) = l[a(\boldsymbol{x}_t; \boldsymbol{\beta}) + \alpha; \boldsymbol{y}_t]$. Together with Assumptions **A09** and **A1-A5**, the proposition requires the following assumptions:

Assumption A6: $\boldsymbol{u}_t$ is independent from the $\boldsymbol{x}_{t-i}$, for $i \geq 0$.

Assumption A7: $(\alpha_0^*, \boldsymbol{\beta}_0)$ belongs to the interior of $\Theta_\alpha \times \Theta_{\boldsymbol{\beta}}$.

Assumption A8: For any $\boldsymbol{x}$, the function $\boldsymbol{\beta} \to a(\boldsymbol{x};\boldsymbol{\beta})$ has continuous third-order derivatives. The pseudo-density function g is three times continuously differentiable.

Assumption A9: i^* and j^* exist, and $j^* \neq 0$.

Assumption A10: The matrix $\mathrm{V}_0\left(\frac{\partial a}{\partial \boldsymbol{\beta}}(\boldsymbol{x}_t,\boldsymbol{\beta}_0)\right)$ is positive definite.

A.4 For Proposition 4.2

The Assumptions for Proposition 3.5 are maintained, except **A9-A10** which are replaced by

Assumption A9*: The matrices $\boldsymbol{K}$ and $\boldsymbol{L}$ are positive definite.

Assumption A10*: For at least one $j \in \{1,\ldots,J\}$, the matrix $\mathrm{V}_0\left(\frac{\partial a_j}{\partial \boldsymbol{\beta}}(x_t,\boldsymbol{\beta}_0)\right)$ is positive definite.

A.5 For Proposition 4.3

Assumption A11: $E_0|l\{a(\boldsymbol{x}_t;\boldsymbol{\beta})+\alpha, \boldsymbol{m}(\boldsymbol{x}_t;\boldsymbol{\beta})+\exp\{-a(\boldsymbol{x}_t;\boldsymbol{\beta})\boldsymbol{C}\}\boldsymbol{\mu};\boldsymbol{y}_t\}| < \infty$, for any $(\alpha,\boldsymbol{\beta},\boldsymbol{\mu}) \in \Theta_\alpha \times \Theta_{\boldsymbol{\beta}} \times \Theta_{\boldsymbol{m}}$.

Assumption A12: $\boldsymbol{\beta}_0 \in \Theta_{\boldsymbol{\beta}}, \alpha_0 + \overline{a}_0 \in \Theta_\alpha, \boldsymbol{\mu}_0 + \exp[(a_0-\alpha_0)\boldsymbol{C}](\boldsymbol{m}_0^* - \boldsymbol{m}_0) \in \Theta_{\boldsymbol{m}}$. Moreover, $\Theta_{\boldsymbol{\beta}}, \Theta_\alpha$ and $\Theta_{\boldsymbol{m}}$ are compact parameter sets.

A.6 For Proposition 4.4

Together with Assumptions **A4-A6, A11** and **A12**, the proposition requires the following assumptions:

Assumption A13: $(\alpha_0^*,\boldsymbol{\beta}_0,\boldsymbol{\mu}_0^*)$ belongs to the interior of $\Theta_\alpha \times \Theta_{\boldsymbol{\beta}} \times \Theta_{\boldsymbol{\mu}}$.

Assumption A14: For any $\boldsymbol{x}$, the functions $\boldsymbol{\beta} \to a(\boldsymbol{x};\boldsymbol{\beta})$ and $\boldsymbol{\beta} \to \boldsymbol{m}(\boldsymbol{x};\boldsymbol{\beta})$ have continuous third-order derivatives. The pseudo-density function g is three times continuously differentiable.

Assumption A15: the matrices $\boldsymbol{I}$ and $\boldsymbol{J}$ exist, and $\boldsymbol{J}$ is positive definite.

B Proofs

B.1 Proof of Proposition 3.1

By assumption (3.1), $\boldsymbol{\lambda}_0^*$ satisfies the first-order condition:

$$E_0\left\{\frac{\partial l}{\partial \boldsymbol{\lambda}}(\boldsymbol{\lambda}_0^*;\boldsymbol{u}_t)\right\} = 0. \tag{B.1}$$

We have

$$\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial l}{\partial \boldsymbol{\lambda}}\{\boldsymbol{\lambda}_t(\boldsymbol{\theta});\boldsymbol{u}_t\}. \tag{B.2}$$

It follows that, using Assumption **A0** and (B.1),

$$E_0\left\{\frac{\partial \ell_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}}\right\} = E_0\left\{\frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}}\right\} E_0\left\{\frac{\partial l}{\partial \boldsymbol{\lambda}}\{\boldsymbol{\lambda}_0^*; \boldsymbol{u}_t\}\right\} = 0, \tag{B.3}$$

which is the first-order condition for the enlarged model. We also have

$$\mathrm{Var}_0\left\{\frac{\partial \ell_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}}\right\} = E_0\left\{\frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}}\boldsymbol{I}^0\frac{\partial \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}'}\right\}.$$

Turning to the second-order derivatives, we have, for $j = 1, \dots, r$,

$$\frac{\partial^2 \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \theta_j} = \frac{\partial^2 \boldsymbol{\lambda}_t'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \theta_j}\frac{\partial l}{\partial \boldsymbol{\lambda}}\{\boldsymbol{\lambda}_t(\boldsymbol{\theta}); \boldsymbol{u}_t\} + \frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial^2 l}{\partial \boldsymbol{\lambda}\partial \boldsymbol{\lambda}'}\{\boldsymbol{\lambda}_t(\boldsymbol{\theta}); \boldsymbol{u}_t\}\frac{\partial \boldsymbol{\lambda}_t(\boldsymbol{\theta})}{\partial \theta_j}.$$

Thus, by taking the expectation of both sides at $\boldsymbol{\theta} = \boldsymbol{\theta}_0^*$, it follows that

$$E_0\left\{\frac{\partial^2 \ell_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right\} = E_0\left\{\frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}}\boldsymbol{J}^0\frac{\partial \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}'}\right\}.$$

Now suppose $\boldsymbol{I}^0$ is nonsingular and (3.5) holds. If $\boldsymbol{I}$ were nonsingular, there would exist $\boldsymbol{x} \in \mathbb{R}^r$, $\boldsymbol{x} \neq 0$ such that

$$\boldsymbol{x}' E_0\left\{\frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}}\boldsymbol{I}^0\frac{\partial \boldsymbol{\lambda}_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}'}\right\}\boldsymbol{x} = 0.$$

Because $\boldsymbol{I}^0$ is positive-definite by assumption, we would have $\boldsymbol{x}'\frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}} = 0$, in contradiction with (3.5). The same arguments apply to show the nonsingularity of $\boldsymbol{J}$.

B.2 Proof of Proposition 3.3

By Assumption **A02** and the ergodic theorem, the pseudo log-likelihood converges almost surely to the limit criterion. By the proof of Proposition 3.2 and Assumptions **A01** and **A04**, the asymptotic criterion is uniquely maximized at $\boldsymbol{\theta}_0^*$. The a.s. consistency of the PML estimator follows by a standard compactness argument.

Now we turn to the asymptotic normality. Since $\hat{\boldsymbol{\theta}}_T$ converges to $\boldsymbol{\theta}_0^*$, which stands in the interior of the parameter space by **A06**, the derivative of the criterion is equal to zero at $\hat{\boldsymbol{\theta}}_T$. We thus have:

$$\begin{aligned}
\mathbf{0} &= T^{-1/2}\sum_{t=1}^{T}\frac{\partial}{\partial \boldsymbol{\theta}}\ell_t(\hat{\boldsymbol{\theta}}_T) \\
&= T^{-1/2}\sum_{t=1}^{T}\frac{\partial}{\partial \boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0^*) + \left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2}{\partial \theta_i \partial \theta_j}\ell_t(\boldsymbol{\theta}_{ij}^*)\right)\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0^*\right) \\
&:= \boldsymbol{s}_T - \boldsymbol{J}_T\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0^*\right),
\end{aligned}$$

where the $\boldsymbol{\theta}_{ij}^*$'s are between $\hat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\theta}_0^*$. The proof is divided in several steps.

i) Asymptotic normality of s_T.
We have

$$\frac{\partial \ell_t(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}} = \frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}}\frac{\partial l}{\partial \boldsymbol{\lambda}}\{\boldsymbol{\lambda}_0^*; \boldsymbol{u}_t\}. \tag{B.4}$$

It follows from (B.1) and Assumption **A05** that $\frac{\partial}{\partial \boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0^*)$ is a martingale difference sequence with respect to the filtration $\mathcal{F}_t = \sigma\left(\left\{u_{t-i}, \frac{\partial \boldsymbol{\lambda}'_{t-i}(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}}, i \geq 0\right\}\right)$. Applying a central limit theorem for square integrable, ergodic and stationary martingale difference (see Billingsley, 1961), it follows that:

$$T^{-1/2}\sum_{t=1}^{T}\frac{\partial}{\partial \boldsymbol{\theta}}\ell_t(\boldsymbol{\theta}_0^*) \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{I}\right).$$

ii) Convergence of $\boldsymbol{J}_T$ to $\boldsymbol{J}$.

Denote by θ_i the i-th component of $\boldsymbol{\theta}$. By a Taylor expansion around $\boldsymbol{\theta}_0^*$ we have, for all i and j,

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\boldsymbol{\theta}_{ij}^*) \quad = \quad \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\boldsymbol{\theta}_0^*) + \frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial \boldsymbol{\theta}'}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\tilde{\boldsymbol{\theta}}_{ij})\right)(\boldsymbol{\theta}_{ij}^* - \boldsymbol{\theta}_0^*),$$

where $\tilde{\boldsymbol{\theta}}_{ij}$ is between $\tilde{\boldsymbol{\theta}}_{ij}^*$ and $\tilde{\boldsymbol{\theta}}_0^*$. The ergodic theorem shows that the first term in the right-hand side converges to $-\boldsymbol{J}$ a.s. The second term converges to 0 a.s. by Assumption **A09**. Indeed, by the ergodic theorem

$$\begin{aligned}\limsup_{T\to\infty}\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial \boldsymbol{\theta}'}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\tilde{\boldsymbol{\theta}}_{ij})\right)\right\| &\leq \limsup_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0^*)}\left\|\frac{\partial}{\partial \boldsymbol{\theta}'}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\boldsymbol{\theta})\right)\right\| \\ &= E_0\sup_{\boldsymbol{\theta}\in V(\boldsymbol{\theta}_0^*)}\left\|\frac{\partial}{\partial \boldsymbol{\theta}'}\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ell_t(\boldsymbol{\theta})\right)\right\| < \infty.\end{aligned}$$

iii) Invertibility of $\boldsymbol{J}$.

Follows straightforwardly from Assumption **A08** and Proposition 3.1.

iv) asymptotic distribution of $\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0^*\right)$.

Follows from $\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0^*\right) = \boldsymbol{J}_T^{-1}\boldsymbol{s}_T$.

B.3 Proof of Corollary 3.1

The asymptotic FOC corresponds to the maximization of:

$$\tilde{L}_\infty(\boldsymbol{\beta}, \alpha) = E_0 l[a(\boldsymbol{x};\boldsymbol{\beta}) + \alpha; \boldsymbol{y}],$$

where $l(\cdot)$ is defined in (3.6). By using the expression of the first-order derivative:

$$\frac{\partial l(a;\boldsymbol{y})}{\partial a} = \frac{\partial \log g}{\partial \boldsymbol{u}'}\{\exp(a\boldsymbol{C})\boldsymbol{y}\}\boldsymbol{C}\exp(a\boldsymbol{C})\boldsymbol{y} + \mathrm{Tr}(\boldsymbol{C}),$$

these FOC are:

$$E_0\left\{\gamma(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\beta}_0^*,\alpha_0^*)\begin{bmatrix}\frac{\partial a(\boldsymbol{x};\boldsymbol{\beta}_0^*)}{\partial \boldsymbol{\beta}} \\ 1\end{bmatrix}\right\} = 0,$$

where $\gamma(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\beta},\alpha)$ is given by (3.10) and $(\boldsymbol{\beta}_0^*,\alpha_0^*)$ denote the pseudo true values. These FOC are equivalently written as:

$$Cov_0[\gamma(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\beta}_0^*,\alpha_0^*), \frac{\partial a(\boldsymbol{x};\boldsymbol{\beta}_0^*)}{\partial \boldsymbol{\beta}}] = 0, \qquad E_0[\gamma(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\beta}_0^*,\alpha_0^*)] = 0.$$

In view of (3.11), the FOC (3.12) is automatically satisfied by the independence of $\boldsymbol{x}$ and u. Let us now look at the second equation for $\boldsymbol{\beta}_0^* = \boldsymbol{\beta}_0$:

$$E_0\gamma(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\beta}_0, \alpha_0^*) = 0. \tag{B.5}$$

This is the FOC in the generic pseudo-model:

$$\tilde{\boldsymbol{y}}_t = \exp[a(\boldsymbol{x}_t; \boldsymbol{\beta}_0)\boldsymbol{C}]\boldsymbol{y}_t = \exp(-a\boldsymbol{C})\boldsymbol{u}_t.$$

Therefore $\alpha_0^* = \alpha_0 + a_0 - a_0^*$. The equivalence in the second part of the Corollary follows from:

$$\gamma(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\beta}_0, \alpha_0) = \frac{\partial \log g}{\partial \boldsymbol{u}'}(u)\boldsymbol{C}u - \mathrm{Tr}(\boldsymbol{C}).$$

B.4 Proof of Proposition 3.5

The proposition is a direct consequence of Proposition 3.3, with $\boldsymbol{\lambda}_t(\boldsymbol{\theta})$ defined in (3.9) and

$$\frac{\partial \boldsymbol{\lambda}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \dfrac{\partial a(\boldsymbol{x}_t; \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \\ 1 \end{pmatrix}. \tag{B.6}$$

Thus, (3.5) is satisfied under **A10**. With $l(\lambda; \boldsymbol{u}_t) = \log g[\exp(\lambda\boldsymbol{C})u_t] + \lambda\mathrm{Tr}(\boldsymbol{C}), \lambda \in \mathbb{R}$ we have:

$$\begin{aligned}
\frac{\partial l}{\partial \lambda}(\lambda; \boldsymbol{u}_t) &= \frac{\partial \log g}{\partial \boldsymbol{u}'}\{\exp(\lambda\boldsymbol{C})\boldsymbol{u}_t\}\boldsymbol{C}\exp(\lambda\boldsymbol{C})\boldsymbol{u}_t + \mathrm{Tr}(\boldsymbol{C}), \\
\frac{\partial^2 l}{\partial \lambda^2}(\lambda; \boldsymbol{u}_t) &= \left(\frac{\partial \log g}{\partial \boldsymbol{u}'}[\exp\{\lambda\boldsymbol{C}\}\boldsymbol{u}_t]\right)\boldsymbol{C}^2\exp\{\lambda\boldsymbol{C}\}\boldsymbol{u}_t \\
&\quad + \boldsymbol{u}_t'\exp\{\lambda\boldsymbol{C}'\}\boldsymbol{C}'\left(\frac{\partial^2 \log g}{\partial \boldsymbol{u}\partial \boldsymbol{u}'}[\exp\{\lambda\boldsymbol{C}\}\boldsymbol{u}_t]\right)\boldsymbol{C}\exp\{\lambda\boldsymbol{C}\}\boldsymbol{u}_t,
\end{aligned}$$

from which the expressions for $\boldsymbol{I}^0 = i^*$ and $\boldsymbol{J}^0 = j^*$ follow.

B.5 Proof of Proposition 4.1

The proof is analogous to the proof of Proposition 3.4. The asymptotic pseudo log-likelihood is

$$E_0\left(\log g\left[\exp\left\{\sum_{j=1}^J [a_j(\boldsymbol{x}_t; \boldsymbol{\beta}) + \alpha_j]\boldsymbol{C}_j\right\}\boldsymbol{y}_t\right] + \sum_{j=1}^J [a_j(\boldsymbol{x}_t; \boldsymbol{\beta}) + \alpha_j]\mathrm{Tr}(\boldsymbol{C}_j)\right).$$

Replacing $\boldsymbol{y}_t$ by $\exp\left\{-\sum_{j=1}^J [a_j(\boldsymbol{x}_t; \boldsymbol{\beta}_0) + \alpha_{0j}]\boldsymbol{C}_j\right\}\boldsymbol{u}_t$, and using the assumption of commutative Lie group, we get

$$E_0\left(\log g\left[\exp\left\{\sum_{j=1}^J [a_j(\boldsymbol{x}_t; \boldsymbol{\beta}) - a_j(\boldsymbol{x}_t; \boldsymbol{\beta}_0) + \alpha_j - \alpha_{0j}]\boldsymbol{C}_j\right\}\boldsymbol{u}_t\right] + \sum_{j=1}^J [a_j(\boldsymbol{x}_t; \boldsymbol{\beta}) + \alpha_j]\mathrm{Tr}(\boldsymbol{C}_j)\right).$$

Therefore, with $l[\boldsymbol{a}; \boldsymbol{y}]$ defined in (4.2),

$$(\boldsymbol{\beta}_0^*, \boldsymbol{\alpha}_0^*) = \arg\max_{\boldsymbol{\beta}\in\Theta_{\boldsymbol{\beta}}, \boldsymbol{\alpha}\in\Theta_\alpha} E_0 l[\boldsymbol{a}(\boldsymbol{x}_t; \boldsymbol{\beta}) - \boldsymbol{a}(\boldsymbol{x}_t; \boldsymbol{\beta}_0) + \boldsymbol{\alpha} - \boldsymbol{\alpha}_0; \boldsymbol{u}_t].$$

As in the proof of Proposition 3.4, we get $E_0 l[\boldsymbol{a}(\boldsymbol{x}_t; \boldsymbol{\beta}) - \boldsymbol{a}(\boldsymbol{x}_t; \boldsymbol{\beta}_0) + \boldsymbol{\alpha} - \boldsymbol{\alpha}_0; \boldsymbol{u}_t] \leq E_0 l(\boldsymbol{a}_0^* - \boldsymbol{a}_0; \boldsymbol{u}_t)$, and the upper bound is reached for $\boldsymbol{\beta}_0^* = \boldsymbol{\beta}_0$ and $\boldsymbol{\alpha}_0^* = \boldsymbol{\alpha}_0 + \boldsymbol{a}_0^* - \boldsymbol{a}_0$.

B.6 Proof of Proposition 4.2

Because the proof is analogous to that of Proposition 3.5, we only develop the derivation of the asymptotic covariance matrix of the PML estimator.

For $\boldsymbol{\gamma} = (\text{Tr}(\boldsymbol{C}_1), \ldots, \text{Tr}(\boldsymbol{C}_J))'$, and $l_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) = l[a(\boldsymbol{x}_t; \boldsymbol{\beta}) + \boldsymbol{\alpha}, \boldsymbol{y}_t)$ we have:

$$\begin{aligned} l_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \log g \left[\exp\left\{ \sum_{j=1}^J [a_j(\boldsymbol{x}_t; \boldsymbol{\beta}) + \alpha_j] \boldsymbol{C}_j \right\} \boldsymbol{y}_t \right] + \sum_{j=1}^J [a_j(\boldsymbol{x}_t; \boldsymbol{\beta}) + \alpha_j] \text{Tr}(\boldsymbol{C}_j), \\ &:= \log g\{z_t(\boldsymbol{\beta})\} + \boldsymbol{\gamma}'\{\boldsymbol{a}(\boldsymbol{x}_t; \boldsymbol{\beta}) + \boldsymbol{\alpha}\}. \end{aligned}$$

Using the computations of Appendix C.3, it follows that:

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}'} l_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \frac{\partial \log g}{\partial \boldsymbol{u}'} \{z_t(\boldsymbol{\alpha}, \boldsymbol{\beta})\} \frac{\partial z_t(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + \boldsymbol{\gamma}'\{\boldsymbol{a}(\boldsymbol{x}_t; \boldsymbol{\beta}) + \boldsymbol{\alpha}\}, \\ &= \left(\frac{\partial \log g}{\partial \boldsymbol{u}'} \{z_t(\boldsymbol{\alpha}, \boldsymbol{\beta})\} \boldsymbol{C}[\boldsymbol{I}_J \otimes z_t(\boldsymbol{\alpha}, \boldsymbol{\beta})] + \boldsymbol{\gamma}' \right) \frac{\partial \boldsymbol{a}(\boldsymbol{x}_t; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \\ &:= h_t'(\boldsymbol{\alpha}, \boldsymbol{\beta}) \frac{\partial \boldsymbol{a}(\boldsymbol{x}_t; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}. \end{aligned}$$

Proceeding similarly with parameter $\boldsymbol{\alpha}$, we find that:

$$\begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\beta}} l_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) \\ \frac{\partial}{\partial \boldsymbol{\alpha}} l_t(\boldsymbol{\alpha}, \boldsymbol{\beta}) \end{pmatrix} = \begin{pmatrix} \frac{\partial \boldsymbol{a}'(\boldsymbol{x}_t; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \\ \boldsymbol{I}_J \end{pmatrix} h_t(\boldsymbol{\alpha}, \boldsymbol{\beta}).$$

We have $z_t(\boldsymbol{\alpha}_0^*, \boldsymbol{\beta}_0) = \exp\left\{ \sum_{j=1}^J (\alpha_{0j}^* - \alpha_{0j}) \boldsymbol{C}_j \right\} \boldsymbol{u}_t := \boldsymbol{\Gamma}_0^* \boldsymbol{u}_t$. Thus,

$$\begin{pmatrix} \frac{\partial}{\partial \boldsymbol{\beta}} l_t(\boldsymbol{\alpha}_0^*, \boldsymbol{\beta}_0) \\ \frac{\partial}{\partial \boldsymbol{\alpha}} l_t(\boldsymbol{\alpha}_0^*, \boldsymbol{\beta}_0) \end{pmatrix} = \begin{pmatrix} \frac{\partial a'(\boldsymbol{x}_t; \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \\ \boldsymbol{I}_J \end{pmatrix} \boldsymbol{k}(\boldsymbol{u}_t), \quad \boldsymbol{k}(\boldsymbol{u}_t) = [\boldsymbol{I}_J \otimes (\boldsymbol{\Gamma}_0^* \boldsymbol{u}_t)'] \boldsymbol{C}' \frac{\partial \log g}{\partial \boldsymbol{u}} \{\boldsymbol{\Gamma}_0^* \boldsymbol{u}_t\} + \boldsymbol{\gamma},$$

and, with $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$,

$$\boldsymbol{B} = V_0 \left[\frac{\partial}{\partial \boldsymbol{\theta}} l_t(\boldsymbol{\alpha}_0^*, \boldsymbol{\beta}_0) \right] = E_0 \left[\begin{pmatrix} \frac{\partial a'(\boldsymbol{x}_t; \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \\ \boldsymbol{I}_J \end{pmatrix} \boldsymbol{K} \begin{pmatrix} \frac{\partial a'(\boldsymbol{x}_t; \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \\ \boldsymbol{I}_J \end{pmatrix}' \right], \qquad \boldsymbol{K} = V_0 [\boldsymbol{k}(\boldsymbol{u}_t)].$$

Now,

$$\frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} l_t(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{\partial h_t'(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \frac{\partial \boldsymbol{a}(\boldsymbol{x}_t; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} + \{h_t'(\boldsymbol{\alpha}, \boldsymbol{\beta}) \otimes \boldsymbol{I}_p\} A(\boldsymbol{x}_t; \boldsymbol{\beta}),$$

where $A(\boldsymbol{x}_t; \boldsymbol{\beta})$ is the $Jp \times p$ matrix:

$$A(\boldsymbol{x}_t; \boldsymbol{\beta}) = \begin{pmatrix} \frac{\partial^2 a_1(\boldsymbol{x}_t; \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \\ \vdots \\ \frac{\partial^2 a_J(\boldsymbol{x}_t; \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \end{pmatrix}.$$

Noting that:

$$h_t'(\boldsymbol{\alpha},\boldsymbol{\beta}) = \left[\frac{\partial \log g}{\partial \boldsymbol{u}'}\left\{z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\right\}\boldsymbol{C}_1 z_t(\boldsymbol{\alpha},\boldsymbol{\beta}),\ldots,\frac{\partial \log g}{\partial \boldsymbol{u}'}\left\{z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\right\}\boldsymbol{C}_J z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\right] + \boldsymbol{\gamma}',$$

we compute:

$$\begin{aligned}
&\frac{\partial}{\partial\boldsymbol{\beta}}\left\{\frac{\partial \log g}{\partial \boldsymbol{u}'}\left\{z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\right\}\boldsymbol{C}_j z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\right\}\\
=\;&\frac{\partial}{\partial\boldsymbol{\beta}}\left\{\frac{\partial \log g}{\partial \boldsymbol{u}'}\left\{z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\right\}\right\}\boldsymbol{C}_j z_t(\boldsymbol{\alpha},\boldsymbol{\beta}) + \left[\frac{\partial}{\partial\boldsymbol{\beta}}\{\boldsymbol{C}_j z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\}'\right]\left\{\frac{\partial \log g}{\partial \boldsymbol{u}'}\left\{z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\right\}\right\}'\\
=\;&\frac{\partial z_t'(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\frac{\partial^2 \log g}{\partial \boldsymbol{u}\partial \boldsymbol{u}'}\left\{z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\right\}\boldsymbol{C}_j z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\\
&+\frac{\partial \boldsymbol{a}'(\boldsymbol{x}_t;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}[\boldsymbol{I}_J\otimes z_t(\boldsymbol{\alpha},\boldsymbol{\beta})]'\boldsymbol{C}'\boldsymbol{C}_j'\frac{\partial \log g}{\partial \boldsymbol{u}}\left\{z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\right\}\\
=\;&\frac{\partial \boldsymbol{a}'(\boldsymbol{x}_t;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}[\boldsymbol{I}_J\otimes z_t(\boldsymbol{\alpha},\boldsymbol{\beta})]'\boldsymbol{C}'\\
&\times\left\{\frac{\partial^2 \log g}{\partial \boldsymbol{u}\partial \boldsymbol{u}'}\left\{z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\right\}\boldsymbol{C}_j z_t(\boldsymbol{\alpha},\boldsymbol{\beta}) + \boldsymbol{C}_j'\frac{\partial \log g}{\partial \boldsymbol{u}}\left\{z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\right\}\right\}.
\end{aligned}$$

It follows that:

$$\begin{aligned}
&\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}l_t(\boldsymbol{\beta},\boldsymbol{\alpha})\\
=\;&\frac{\partial \boldsymbol{a}'(\boldsymbol{x}_t;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}[\boldsymbol{I}_J\otimes z_t(\boldsymbol{\alpha},\boldsymbol{\beta})]'\boldsymbol{C}'\\
&\times\sum_{j=1}^{J}\left\{\frac{\partial^2 \log g}{\partial \boldsymbol{u}\partial \boldsymbol{u}'}\left\{z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\right\}\boldsymbol{C}_j z_t(\boldsymbol{\alpha},\boldsymbol{\beta}) + \boldsymbol{C}_j'\frac{\partial \log g}{\partial \boldsymbol{u}}\left\{z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\right\}\right\}\frac{\partial a_j(\boldsymbol{x}_t;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}\\
&+\{h_t'(\boldsymbol{\alpha},\boldsymbol{\beta})\otimes\boldsymbol{I}_p\}A(\boldsymbol{x}_t;\boldsymbol{\beta})\\
=\;&\frac{\partial \boldsymbol{a}'(\boldsymbol{x}_t;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}[\boldsymbol{I}_J\otimes z_t(\boldsymbol{\alpha},\boldsymbol{\beta})]'\boldsymbol{C}'\frac{\partial^2 \log g}{\partial \boldsymbol{u}\partial \boldsymbol{u}'}\left\{z_t(\boldsymbol{\alpha},\boldsymbol{\beta})\right\}\boldsymbol{C}[\boldsymbol{I}_J\otimes z_t(\boldsymbol{\alpha},\boldsymbol{\beta})]\frac{\partial \boldsymbol{a}(\boldsymbol{x}_t;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'}\\
&+\frac{\partial \boldsymbol{a}'(\boldsymbol{x}_t;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}[\boldsymbol{I}_J\otimes z_t(\boldsymbol{\alpha},\boldsymbol{\beta})]'\boldsymbol{C}'\boldsymbol{G}[\boldsymbol{z}_t(\boldsymbol{\alpha},\boldsymbol{\beta})]\frac{\partial \boldsymbol{a}(\boldsymbol{x}_t;\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'} + \{h_t'(\boldsymbol{\alpha},\boldsymbol{\beta})\otimes\boldsymbol{I}_p\}A(\boldsymbol{x}_t;\boldsymbol{\beta}),
\end{aligned}$$

where $G(\boldsymbol{u})$ is the $n\times J$ matrix:

$$\boldsymbol{G}(\boldsymbol{u}) = \left[\boldsymbol{C}_1'\frac{\partial \log g}{\partial \boldsymbol{u}}(\boldsymbol{u})\;\;\boldsymbol{C}_2'\frac{\partial \log g}{\partial \boldsymbol{u}}(u)\;\;\ldots\;\;\boldsymbol{C}_J'\frac{\partial \log g}{\partial \boldsymbol{u}}(\boldsymbol{u})\right].$$

Note that the first-order conditions imply $Eh_t'(\boldsymbol{\alpha}_0^*,\boldsymbol{\beta}_0)=0$. Therefore

$$\boldsymbol{A} = E_0\left[-\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}l_t(\boldsymbol{\alpha}_0^*,\boldsymbol{\beta}_0)\right] \;=\; E_0\left[\begin{pmatrix}\frac{\partial \boldsymbol{a}'(\boldsymbol{x}_t;\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}\\ \boldsymbol{I}_J\end{pmatrix}\boldsymbol{L}\begin{pmatrix}\frac{\partial \boldsymbol{a}'(\boldsymbol{x}_t;\boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}}\\ \boldsymbol{I}_J\end{pmatrix}'\right]$$

where:

$$\boldsymbol{L} = -E_0\left\{[\boldsymbol{I}_J\otimes\boldsymbol{\Gamma}_0^*\boldsymbol{u}_t]'\boldsymbol{C}'\frac{\partial^2 \log g}{\partial \boldsymbol{u}\partial \boldsymbol{u}'}\{\boldsymbol{\Gamma}_0^*\boldsymbol{u}_t\}\boldsymbol{C}[\boldsymbol{I}_J\otimes\boldsymbol{\Gamma}_0^*\boldsymbol{u}_t] + [\boldsymbol{I}_J\otimes\boldsymbol{\Gamma}_0^*\boldsymbol{u}_t]'\boldsymbol{C}'\boldsymbol{G}\left(\boldsymbol{\Gamma}_0^*\boldsymbol{u}_t\right)\right\}.$$

Now we prove that $\boldsymbol{A}$ is nonsingular. Let $\boldsymbol{z} \in \mathbb{R}^{p+J}$ such that $\boldsymbol{z}'\boldsymbol{A}\boldsymbol{z} = 0$. In view of **A9***, it follows that

$$\boldsymbol{z}' \begin{pmatrix} \frac{\partial \boldsymbol{a}'(\boldsymbol{x}_t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \\ \boldsymbol{I}_J \end{pmatrix} = 0.$$

Writing $\boldsymbol{z} = (\boldsymbol{z}_1', \boldsymbol{z}_2')'$, where $\boldsymbol{z}_1 \in \mathbb{R}^p$, $\boldsymbol{z}_2 \in \mathbb{R}^J$, we thus have:

$$\boldsymbol{z}_1' \frac{\partial a_j(\boldsymbol{x}_t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} = \boldsymbol{z}_2'\boldsymbol{e}_j, \quad j = 1, \ldots, J.$$

Finally,

$$V_{as}\left[\sqrt{T}\begin{pmatrix} \hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0 \\ \hat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_0^* \end{pmatrix}\right] = \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}.$$

B.7 Proof of Proposition 4.3

The maximizer of the asymptotic PML of Model (4.11) is defined by:

$$\begin{aligned}
& (\boldsymbol{\beta}_0^*, \boldsymbol{\mu}_0^*, \alpha_0^*) \\
= \; & \arg\max_{\boldsymbol{\beta},\boldsymbol{\mu},\alpha} E_0 l[a(\boldsymbol{x}_t;\boldsymbol{\beta}) + \alpha, \boldsymbol{m}(\boldsymbol{x}_t;\boldsymbol{\beta}) + \exp\{-a(\boldsymbol{x}_t;\boldsymbol{\beta})\boldsymbol{C}\}\boldsymbol{\mu}; \boldsymbol{y}_t] \\
= \; & \arg\max_{\boldsymbol{\beta},\boldsymbol{\mu},\alpha} E_0 l[\Delta a(\boldsymbol{x}_t;\boldsymbol{\beta}) + \alpha - \alpha_0, \exp[\{a(\boldsymbol{x}_t;\boldsymbol{\beta}_0) + \alpha_0\}\boldsymbol{C}]\Delta\boldsymbol{m}(\boldsymbol{x}_t;\boldsymbol{\beta}) \\
& \qquad + \exp[\{-\Delta a(\boldsymbol{x}_t;\boldsymbol{\beta}) + \alpha_0\}\boldsymbol{C}]\boldsymbol{\mu} - \exp(\alpha_0\boldsymbol{C})\boldsymbol{\mu}_0; \boldsymbol{u}_t],
\end{aligned}$$

where $\Delta a(\boldsymbol{x}_t;\boldsymbol{\beta}) = a(\boldsymbol{x}_t;\boldsymbol{\beta}) - a(\boldsymbol{x}_t;\boldsymbol{\beta}_0)$, $\Delta\boldsymbol{m}(\boldsymbol{x}_t;\boldsymbol{\beta}) = \boldsymbol{m}(\boldsymbol{x}_t;\boldsymbol{\beta}) - \boldsymbol{m}(\boldsymbol{x}_t;\boldsymbol{\beta}_0)$, and the argmax are taken over $\Theta_{\boldsymbol{\beta}} \times \Theta_m \times \Theta_\alpha$ and $l(\cdot)$ is defined in (4.13).

Using the notations of Section 3.1, we have, with $\boldsymbol{\theta} = (\boldsymbol{\beta}', \alpha, \boldsymbol{\mu}')'$,

$$\boldsymbol{\lambda}_t(\boldsymbol{\theta}) = \begin{pmatrix} \Delta a(\boldsymbol{x}_t;\boldsymbol{\beta}) + \alpha - \alpha_0 \\ \exp[\{a(\boldsymbol{x}_t;\boldsymbol{\beta}_0) + \alpha_0\}\boldsymbol{C}]\Delta\boldsymbol{m}(\boldsymbol{x}_t;\boldsymbol{\beta}) + \exp[\{-\Delta a(\boldsymbol{x}_t;\boldsymbol{\beta}) + \alpha_0\}\boldsymbol{C}]\boldsymbol{\mu} - \exp(\alpha_0\boldsymbol{C})\boldsymbol{\mu}_0 \end{pmatrix}.$$

In view of (4.14), and using the same argument as in the proof of Proposition 3.4, we thus have:

$$\boldsymbol{\beta}_0^* = \boldsymbol{\beta}_0, \qquad \alpha_0^* = \alpha_0 + a_0^* - a_0, \qquad \boldsymbol{\mu}_0^* = \boldsymbol{\mu}_0 + \exp[(a_0 - \alpha_0)\boldsymbol{C}](\boldsymbol{m}_0^* - \boldsymbol{m}_0).$$

B.8 Proof of Proposition 4.4

It suffices to verify the conditions required for Proposition 3.3. First note that Assumption **A0** is satisfied, with

$$\boldsymbol{\lambda}_0^* = \begin{pmatrix} \alpha_0^* - \alpha_0 \\ \exp(\alpha_0\boldsymbol{C})(\boldsymbol{\mu}_0^* - \boldsymbol{\mu}_0) \end{pmatrix}.$$

We have:

$$\frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \frac{\partial a(\boldsymbol{x}_t;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} & \frac{\partial \boldsymbol{m}'(\boldsymbol{x}_t;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \exp[\{a(\boldsymbol{x}_t;\boldsymbol{\beta}_0) + \alpha_0\}\boldsymbol{C}'] - \frac{\partial a(\boldsymbol{x}_t;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\boldsymbol{\mu}' \exp[\{-\Delta a(\boldsymbol{x}_t;\boldsymbol{\beta}) + \alpha_0\}\boldsymbol{C}']\boldsymbol{C}' \\ 1 & \boldsymbol{0}_{1\times n} \\ \boldsymbol{0}_{n\times 1} & \exp[\{-\Delta a(\boldsymbol{x}_t;\boldsymbol{\beta}) + \alpha_0\}\boldsymbol{C}]. \end{pmatrix}.$$

Hence,

$$\frac{\partial \boldsymbol{\lambda}_t'(\boldsymbol{\theta}_0^*)}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \frac{\partial a(\boldsymbol{x}_t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} & \frac{\partial \boldsymbol{m}'(\boldsymbol{x}_t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \exp[\{a(\boldsymbol{x}_t;\boldsymbol{\beta}_0)+\alpha_0\}\boldsymbol{C}'] - \frac{\partial a(\boldsymbol{x}_t;\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \boldsymbol{\mu}' \exp[\alpha_0 \boldsymbol{C}']\boldsymbol{C}' \\ 1 & \mathbf{0}_{1\times n} \\ \mathbf{0}_{n\times 1} & \exp[\alpha_0 \boldsymbol{C}]. \end{pmatrix}.$$

It is easy to see that, under **A10**, this matrix has full row rank with probability 1. Thus (3.5) is satisfied.

We also have, with $\lambda \in \mathbb{R}$, $\boldsymbol{\lambda}_2 \in \mathbb{R}^n$ and $\boldsymbol{\lambda} = (\lambda_1, \boldsymbol{\lambda}_2')'$,

$$l(\boldsymbol{\lambda};\boldsymbol{u}_t) = \log g[\exp(\lambda_1 \boldsymbol{C})\{u_t + \exp(a_0 \boldsymbol{C})\boldsymbol{\lambda}_2\}] + \lambda_1 \mathrm{Tr}(\boldsymbol{C}).$$[21]

Thus

$$\frac{\partial l}{\partial \boldsymbol{\lambda}}(\boldsymbol{\lambda};\boldsymbol{u}_t) = \begin{pmatrix} \frac{\partial \log g}{\partial \boldsymbol{u}'}[\exp(\lambda_1 \boldsymbol{C})\{u_t + \exp(a_0 \boldsymbol{C})\boldsymbol{\lambda}_2\}]\boldsymbol{C}\exp(\lambda_1 \boldsymbol{C})\{u_t + \exp(a_0 \boldsymbol{C})\boldsymbol{\lambda}_2\} + \mathrm{Tr}(\boldsymbol{C}) \\ \exp[(a_0+\lambda_1)\boldsymbol{C}'] \frac{\partial \log g}{\partial \boldsymbol{u}}[\exp(\lambda_1 \boldsymbol{C})\{u_t + \exp(a_0 \boldsymbol{C})\boldsymbol{\lambda}_2\}] \end{pmatrix}.$$

Since $\boldsymbol{\lambda}_0^* = (a_0^* - a_0, \boldsymbol{m}_0' - \boldsymbol{m}_0^{*'})'$, we have

$$\frac{\partial l}{\partial \boldsymbol{\lambda}_0^*}(\boldsymbol{\lambda};\boldsymbol{u}_t) = \begin{pmatrix} \frac{\partial \log g}{\partial \boldsymbol{u}'}[\exp\{\overline{a}_0 \boldsymbol{C}\}\{u_t - \exp(a_0 \boldsymbol{C})\overline{\boldsymbol{m}}_0\}]\boldsymbol{C}\exp\{\overline{a}_0 \boldsymbol{C}\}\{u_t - \exp(a_0 \boldsymbol{C})\overline{\boldsymbol{m}}_0\} + \mathrm{Tr}(\boldsymbol{C}) \\ \exp[a_0^* \boldsymbol{C}'] \frac{\partial \log g}{\partial \boldsymbol{u}}[\exp\{\overline{a}_0 \boldsymbol{C}\}\{u_t - \exp(a_0 \boldsymbol{C})\overline{\boldsymbol{m}}_0\}] \end{pmatrix}$$

where $\overline{a}_0 = a_0^* - a_0$, $\overline{\boldsymbol{m}}_0 = \boldsymbol{m}_0^* - \boldsymbol{m}_0$.

The second-order derivative of $l(\boldsymbol{\lambda};\boldsymbol{u}_t)$ can be similarly evaluated at $\boldsymbol{\lambda}_0^*$.

C Derivatives for the index and multi-index models

C.1 Derivatives of $a \to \log g(e^{a\boldsymbol{C}}\boldsymbol{y})$

For $a \in \mathbb{R}$, $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{C}$ a $n \times n$ matrix, $g : \mathbb{R}^n \to \mathbb{R}^+$ a function,

$$\begin{aligned} \frac{\partial}{\partial a}(e^{a\boldsymbol{C}}\boldsymbol{y}) &= \boldsymbol{C}e^{a\boldsymbol{C}}\boldsymbol{y}, \qquad \frac{\partial^2}{\partial a^2}(e^{a\boldsymbol{C}}\boldsymbol{y}) = \boldsymbol{C}^2 e^{a\boldsymbol{C}}\boldsymbol{y} \\ \frac{\partial}{\partial a}\log g(e^{a\boldsymbol{C}}\boldsymbol{y}) &= \left[\frac{\partial \log g}{\partial \boldsymbol{u}'}(e^{a\boldsymbol{C}}\boldsymbol{y})\right]\boldsymbol{C}e^{a\boldsymbol{C}}\boldsymbol{y}, \\ \frac{\partial^2}{\partial a^2}\log g(e^{a\boldsymbol{C}}\boldsymbol{y}) &= \left(\boldsymbol{C}e^{a\boldsymbol{C}}\boldsymbol{y}\right)'\left[\frac{\partial^2 \log g}{\partial \boldsymbol{u}\partial \boldsymbol{u}'}(e^{a\boldsymbol{C}}\boldsymbol{y})\right]\boldsymbol{C}e^{a\boldsymbol{C}}\boldsymbol{y} + \left(\boldsymbol{C}^2 e^{a\boldsymbol{C}}\boldsymbol{y}\right)'\left[\frac{\partial \log g}{\partial \boldsymbol{u}}(e^{a\boldsymbol{C}}\boldsymbol{y})\right]. \end{aligned}$$

[21] This corresponds to a reparameterization of (4.13) but, to avoid new notations, we still denote by l the criterion function.

C.2 Derivatives of $\boldsymbol{\beta} \to e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y}$ and $\boldsymbol{\beta} \to \log g(e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y})$

For $a : \mathbb{R}^p \to \mathbb{R}$ and $\boldsymbol{\beta} \in \mathbb{R}^p$,

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\beta}'}\left\{e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y}\right\} &= \boldsymbol{C}e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y}\frac{\partial a(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'},\\
\frac{\partial}{\partial \boldsymbol{\beta}}\log g\left\{e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y}\right\} &= \frac{\partial}{\partial a}\log g\left\{e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y}\right\}\cdot\frac{\partial a(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\\
&= \left\{\left[\frac{\partial \log g}{\partial \boldsymbol{u}'}\left\{e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y}\right\}\right]\boldsymbol{C}e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y}\right\}\cdot\frac{\partial a(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}},\\
\frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\log g\left\{e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y}\right\} &= \frac{\partial}{\partial a}\log g\left\{e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y}\right\}\cdot\frac{\partial^2 a(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}+\frac{\partial^2}{\partial a^2}\log g(e^{a\boldsymbol{C}}\boldsymbol{y})\cdot\frac{\partial a(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\frac{\partial a(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\\
&= \left\{\left[\frac{\partial \log g}{\partial \boldsymbol{u}'}\left\{e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y}\right\}\right]\boldsymbol{C}e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y}\right\}\cdot\frac{\partial^2 a(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}+\\
&\quad \left\{\left(\boldsymbol{C}e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y}\right)'\left[\frac{\partial^2 \log g}{\partial \boldsymbol{u}\partial \boldsymbol{u}'}(e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y})\right]\boldsymbol{C}e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y}\right.\\
&\quad\quad \left.+\left(\boldsymbol{C}^2e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y}\right)'\left[\frac{\partial \log g}{\partial \boldsymbol{u}}(e^{a(\boldsymbol{\beta})\boldsymbol{C}}\boldsymbol{y})\right]\right\}\cdot\frac{\partial a(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\frac{\partial a(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}.
\end{aligned}
$$

In these equalities, "$\cdot$" indicates the multiplication of a matrix by a scalar.

C.3 Derivatives of $\boldsymbol{\beta} \to \exp\left\{\sum_{j=1}^J [a_j(\boldsymbol{x}_t;\boldsymbol{\beta})+\alpha_j]\boldsymbol{C}_j\right\}\boldsymbol{y}_t$

Let $\boldsymbol{z}_t(\alpha,\boldsymbol{\beta}) = \exp\left\{\sum_{j=1}^J [a_j(\boldsymbol{x}_t;\boldsymbol{\beta})+\alpha_j]\boldsymbol{C}_j\right\}\boldsymbol{y}_t$, where $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_J)' \in \mathbb{R}^J$, $a_j(\cdot)$ are real valued functions with $\boldsymbol{\beta} \in \mathbb{R}^p$, $\boldsymbol{y}_t \in \mathbb{R}^n$, $\boldsymbol{C}_j$ are $n\times n$ matrices. Let $a(\boldsymbol{x}_t;\boldsymbol{\beta}) = (a_1(\boldsymbol{x}_t;\boldsymbol{\beta}),\ldots,a_J(\boldsymbol{x}_t;\boldsymbol{\beta}))'$. For $i=1,\ldots,J$, let the $n\times 1$ vectors

$$
\begin{aligned}
\boldsymbol{z}_t^{(i)}(\boldsymbol{\alpha},\boldsymbol{\beta}) &= \exp\left\{\sum_{j=1}^{i-1}[a_j(\boldsymbol{x}_t;\boldsymbol{\beta})+\alpha_j]\boldsymbol{C}_j\right\}\boldsymbol{C}_i\exp\left\{[a_i(\boldsymbol{x}_t;\boldsymbol{\beta})+\alpha_i]\boldsymbol{C}_i\right\}\\
&\quad\times\exp\left\{\sum_{j=i+1}^{J}[a_j(\boldsymbol{x}_t;\boldsymbol{\beta})+\alpha_j]\boldsymbol{C}_j\right\}\boldsymbol{y}_t,
\end{aligned}
$$

where the first and last sums are replaced by 0 when $i=1$ and $i=J$, respectively. Let the $n\times J$ block-matrix

$$\boldsymbol{Z}_t(\boldsymbol{\alpha},\boldsymbol{\beta}) = [\boldsymbol{z}_t^{(1)}(\boldsymbol{\alpha},\boldsymbol{\beta})|\ldots|\boldsymbol{z}_t^{(J)}(\boldsymbol{\alpha},\boldsymbol{\beta})].$$

We have

$$\frac{\partial \boldsymbol{z}_t(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} = \boldsymbol{Z}_t(\boldsymbol{\alpha},\boldsymbol{\beta})\frac{\partial a(\boldsymbol{x}_t;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}.$$

When the matrices $\boldsymbol{C}_j$ commute, we have $\boldsymbol{z}_t^{(i)}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \boldsymbol{C}_i\boldsymbol{z}_t(\boldsymbol{\alpha},\boldsymbol{\beta})$ and

$$\boldsymbol{Z}_t(\boldsymbol{\alpha},\boldsymbol{\beta}) = [\boldsymbol{C}_1\boldsymbol{z}_t(\boldsymbol{\alpha},\boldsymbol{\beta})|\ldots|\boldsymbol{C}_J\boldsymbol{z}_t(\boldsymbol{\alpha},\boldsymbol{\beta})] = \boldsymbol{C}[\boldsymbol{I}_J\otimes\boldsymbol{z}_t(\boldsymbol{\alpha},\boldsymbol{\beta})],$$

where $\boldsymbol{C} = [\boldsymbol{C}_1|\dots|\boldsymbol{C}_J]$. Thus, when the $\boldsymbol{C}_j$ commute,

$$\frac{\partial \boldsymbol{z}_t(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} = \boldsymbol{C}[\boldsymbol{I}_J \otimes \boldsymbol{z}_t(\boldsymbol{\alpha},\boldsymbol{\beta})]\frac{\partial a(\boldsymbol{x}_t;\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}.$$

References

[1] Alexander, G. (2000): "On Back Testing Zero Investment Strategies", Journal of Business,73,255-278.

[2] Arsigny, V., Fillard, P., Pennec, X., and N., Ayache (2007): "Geometric Means in a Novel Vector Space Structure on Symmeteric Positive Definite Matrices", SIAM J. Matrix Annal. Appl., 29, 328–347.

[3] Basawa, I., Feignin, P., and C., Heyde (1976): "Asymptotic Properties of Maximum Likelihood Estimators for Stochastic Processes", Sankhya, The Indian Journal of Statistics, 38, 259-270.

[4] Begelfor, E., and M., Werman (2006): "Affine Invariance Revisited", in *IEE International Conference on Computer Vision and Pattern Recognition*, 2087-2094.

[5] Berkes, I. and L. Horváth (2004): "The Efficiency of the Estimators of the Parameters in GARCH Processes", *The Annals of Statistics*, 32, 633–655.

[6] Bhattacharya, A., and R., Bhattacharya (2015): "Nonparametric Inference on Manifolds", Cambridge Univ. Press.

[7] Billingsley, P.(1961): "The Lindeberg-Lévy Theorem for Martingales", *Proceedings of the American Mathematical Society* 12, 788–792.

[8] Bollerslev, T. (1990): "Modelling the Coherence in Short Run Nominal Exchange Rates : A Multivariate Generalized ARCH Approach", Review of Economics and Statistics, 72, 498-505.

[9] Bollerslev, T., and J., Woolridge (1992): "Quasi-Maximum Likelihood Estimators and Inference in Dynamic Models with Time-Varying Covariances", Econometric Reviews, 11, 143-172.

[10] Boothby, W. (1975): "An Introduction to Differentiable Manifolds and Riemannian Geometry", Academic Press.

[11] Calzolari, G., Fiorentini, G., and E., Sentana (2000): "The Score of Conditionally Heteroscedastic Dynamic Regression Models with Student-t Innovations and an LM Test for Multivariate Normality", DP CEMFI.

[12] Cheng, M., and H., Wu (2013): "Local Linear Regression on Manifolds and its Geometric Interpretations", Journal of the American Statistical Association, 108, 1421-1434.

[13] Chiappori, P.A., Komunjer, I. and D., Kristensen (2015): "Nonparametric Identification and Estimation of Transformation Models", Journal of Econometrics, 188, 22-39.

[14] Chirikjian, G. (2011): "Stochastic Models, Information Theory and Lie Groups", Vol 2, Springer.

[15] Dellaportas, P., and M., Pourahmadi (2012) : "Cholesky-GARCH Models with Application to Finance", Stat. Comput., 22, 849-855.

[16] Fan, J., Qi, L., and D., Xiu (2014): "Quasi-Maximum Likelihood Estimation of GARCH Models with Heavy-Tailed Likelihoods", Journal of Business and Economic Statistics, 32, 178-191.

[17] Francq, C., Lepage, G., and J.M., Zakoian (2011): "Two-Stage NonGaussian QML Estimation of GARCH Models and Testing the Efficiency of the Gaussian QMLE", Journal of Econometrics, 165, 246-257.

[18] Gelfand, I., Minlos, R., and J., Shapiro (1962): "Representations of the Rotation and Lorentz Groups and Their Applications", New-York, Pergamon Press.

[19] Godement, R. (2004): "Introduction à la théorie des groupes de Lie", Springer.

[20] Gouriéroux, C., and J., Jasiak (2016): "Semi-Parametric Estimation of Noncausal Vector Autoregressions", CREST Discussion Paper.

[21] Gouriéroux, C., and A., Monfort (1995) : "Statistics and Econometrics Models", Cambridge University Press.

[22] Gouriéroux, C., Monfort, A., and E., Renault (2016) : "Consistent Pseudo Maximum Likelihood Estimators", CREST DP.

[23] Gouriéroux, C., Monfort, A., and A., Trognon (1984): "Pseudo Maximum Likelihood Methods: Theory", Econometrica, 52, 681-700.

[24] Hansen, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators", Econometrica, 50, 1029-1054.

[25] Harvey, C., and A., Siddique (1999): "Autoregressive Conditional Skewness", Journal of Financial and Quantitative Analysis, 34, 465-487.

[26] Henderson, H.V., and S. R., Searle (1979): "Vec and Vech Operators for Matrices, with Some Uses in Jacobians and Multivariate Statistics", The Canadian Journal of Statistics, 7, 65-81.

[27] Huang, Z., Wang, R., Shao, S., Li, X. and X., Chen (2015): "Log-Euclidian Metric Learning on Symmetric Positive Definite Manifolds with Applications to Image Set Classification", in Proceedings of the 32nd International Conference on Machine Learning, JMLR, vol 37.

[28] Hull, B. (2015): "Lie Groups, Lie Algebras and Representation : An Elementary Introduction", Springer, 2nd edition.

[29] Karcher, H. (1977): "Riemannian Center of Mass and Mollifier Smoothing", Communications on Pure and Applied Mathematics, 30, 509-541.

[30] Karlin, S., and H., Taylor (1998): "An Introduction to Stochastic Modelling", 3^{rd} edition, Academic Press.

[31] Liu, Z., and S., Sarkar (2006): "Improved Gait Recognition by Gait Dynamics Normalization", IEEE Transactions on Pattern Analysis and Machine Intelligence, 28, 863-876.

[32] McCullagh, P. (1983): "Quasi-likelihood Functions", Annals of Statistics, 11, 59-67.

[33] Newey, W., and D., Steigerwald (1997): "Asymptotic Bias for Quasi Maximum Likelihood Estimators in Conditional Heteroscedastic Models", Econometrica, 65, 587-599.

[34] Norris, J. (1997): "Markov Chains", Cambridge University Press.

[35] Tse, Y. (2000): "A Test for Constant Correlations in Multivariate GARCH Model", Journal of Econometrics, 98, 107-127.

[36] Turaga,P., Veeraraghavan, A., Srivastava, A., and R., Chellappa (2010): "Statistical Analysis on Manifolds and Its Applications to Video Analysis", in *Video Search and Mining*, D. Schonfeld, C. Shan, D. Tao (Eds), Springer, 115-144.

[37] Wedderburn, R. (1974): "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method", Biometrika, 61, 439-447.

[38] White, H. (1994): "Estimation, Inference and Specification Analysis", Cambridge Univ. Press.

[39] Yuan, Y., Zhu, H., Lin, W., and J., Marron (2012): "Local Polynomial Regression for Symmetric Positive Definite Matrices", Journal of the Royal Statistical Society: Series B, 74, 697-719.