

The R packages `VGAM` and `VGAMextra` handling systems of cointegrated time series (Bivariate case)

by Victor Miranda
PhD Candidate, University of Auckland
Supervisor: Thomas Yee.

February 9, 2018

VGLMs/VGAMs possess infrastructure to handle systems of **cointegrated** time series (SCTSs), but we are still restricted to the bivariate (Normal) case due to absence of R software to estimate the multivariate normal, such as `trinormal()`.

Here, using VGLMs, I show how to handle a very popular choice for SCTSs: The Error Correction Model (ECM), as introduced by Engle and C. (1987). Pfaff (2011) presents a compendium of techniques and methodologies to handle cointegrated time series, with examples in R (including the aforementioned), where the ECM explored in this section is a special case. The R code choices, however, does not handle the general case, when the off-diagonal elements of the covariance matrix are non-zero. As we will see later in this document and compared to Pfaff (2011), VGLMs not only become the natural choice accommodating ECMs operating the general case, but also opens further areas for development.

Let's start by simulating $\mathbf{y}_t = (y_{1,t}, y_{2,t})^T$, two **random walks**, as follows ($n = 280$):

$$\begin{aligned} y_{2,t} &= \beta_0 + \beta_1 y_{1,t} + \beta_2 y_{2,t-1} + \varepsilon_{2,t}, \\ y_{1,t} &= y_{1,t-1} + \varepsilon_{1,t}, \end{aligned} \tag{1}$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \varepsilon_{2,t})^T \sim N_2(\mathbf{0}, \mathbf{V})$, $\mathbf{V} = \begin{pmatrix} \sigma_{\varepsilon_{1,t}}^2 & \sigma_{\varepsilon_{1,t}}\sigma_{\varepsilon_{2,t}}\rho \\ \sigma_{\varepsilon_{1,t}}\sigma_{\varepsilon_{2,t}}\rho & \sigma_{\varepsilon_{2,t}}^2 \end{pmatrix}$, with

$$\sigma_{\varepsilon_{1,t}} = \exp(\log(1.5)), \quad \sigma_{\varepsilon_{2,t}} = \exp(0), \quad \rho = 0.75, \quad (\beta_0, \beta_1, \beta_2)^T = (0.0, 2.5, -0.32)^T.$$

The R code is:

```
> library(VGAM, lib.loc = "~/phdvgam/myRlibs")
> library(VGAMextra, lib.loc = "~/phdvgam/myRlibs")
>
> set.seed(2017081901)
```

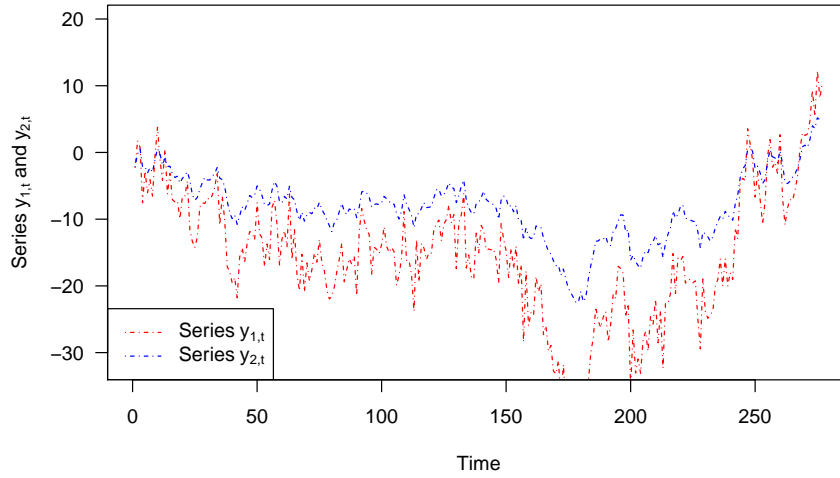


Figure 1. Non-stationary (simulated) series $y_{1,t}$ and $y_{2,t}$ from Model (1).

```

> nn <- 280
> rho <- 0.75 * 1
> # Gaussian noise1
> s2u <- exp(log(1.5))
> ut <- rnorm(nn, 0, s2u)
> # Gaussian noise2
> s2w <- exp(0)
> wt <- rnorm(nn, 0, s2w)
> my.errors <- rbinorm(nn, mean1 = 0, mean2 = 0,
                      var1 = s2u, var2 = s2w, cov12 = rho)
> ut <- my.errors[, 1]
> wt <- my.errors[, 2]
> yt <- xt <- numeric(0)
> xt[1] <- ut[1]
> yt[1] <- wt[1]
>
> beta <- c(0.0, 2.5, -0.32)
>
> for (ii in 2:nn) {
  xt[ii] <- xt[ii - 1] + ut[ii]
  yt[ii] <- beta[1] + beta[2] * xt[ii] + beta[3] * yt[ii - 1] + wt[ii]
}

```

The plot of both, $y_{1,t}$ and $y_{2,t}$ can be seen from Figure (1).

To model the dynamic behaviour of (1), I will use an Order(u, v)-ECM [ECM(u, v)], whose general form is given by (equations 4.5a and 4.5b, Ch. 4 in Pfaff (2011)):

$$\Delta y_{2,t} = \phi_0 + \gamma_1 \hat{z}_{t-1} + \sum_{i=1}^u \phi_{1,i} \Delta y_{1,t-i} + \sum_{j=1}^v \phi_{2,j} \Delta y_{2,t-j} + \varepsilon_{2,t}, \quad (2)$$

$$\Delta y_{1,t} = \psi_0 + \gamma_2 \hat{z}_{t-1} + \sum_{i=1}^u \psi_{1,i} \Delta y_{2,t-i} + \sum_{j=1}^v \psi_{2,j} \Delta y_{1,t-j} + \varepsilon_{1,t},$$

$t = 1, \dots, T$, where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \varepsilon_{2,t})^T \stackrel{\text{iid}}{\sim} N_2(\mathbf{0}, \boldsymbol{\Sigma})$, \widehat{z}_t are the residuals of the static regression $y_{2,t} \sim y_{1,t}$, and $\Delta(\cdot) = (1 - L)(\cdot)$. The adjustment rate of the error from the long-run equilibrium is determined by γ_1 , expected to be negative, if the system converges from its long-run equilibrium path. Note, while the vector-error terms $\boldsymbol{\varepsilon}_t$ are iid, its components may be correlated in the same time period (contemporaneous correlation).

However, we firstly must verify that $y_{1,t}$ and $y_{2,t}$ are cointegrated. For this, the residuals, \widehat{z}_t , of the static regression $y_{2,t} \sim y_{1,t}$ must conform with autoregressive stationarity conditions (Engle and C., 1987). This may be verified via the unit root test implemented (by default) in VGLM-ARIMA family functions, from the package [VGAMextra](#). In this case, `ARff()` is called, along with `noChecks = FALSE`, requesting the unit root test. The results, shown below, confirm stationarity of the residuals (no roots lying inside the unit circle), that is, $y_{1,t}$ and $y_{2,t}$ are cointegrated.

```
> errors.coint <- residuals(lm(yt ~ xt))
> data.errors <- data.frame(est.errors = errors.coint)
> fit.root.test <- vglm(est.errors ~ 1, family = ARff(nodrift = TRUE,
                                                    noChecks = FALSE),
                        data = data.errors, trace = FALSE)

Checks on stationarity / invertibility successfully performed.
No roots lying inside the unit circle.
Further details within the 'summary' output.
```

Interestingly, however, the aforementioned ECM(u, v) (cf. (2)) can be seen as a VGLM with two-responses, $\Delta y_{1,t} | \phi_{t-1} = \Delta y_{1,t}$, and $\Delta y_{2,t} | \phi_{t-1} = \Delta y_{2,t}$, with bivariate normal errors $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \varepsilon_{2,t})^T$, hence resulting in the following VGLM associated to (2):

$$\begin{aligned}
 (\Delta y_{1,t}, \Delta y_{2,t})^T &\sim N_2((\mu_{\Delta y_{1,t}}, \mu_{\Delta y_{2,t}})^T, \boldsymbol{\Sigma}) \tag{3} \\
 \mu_{\Delta y_{2,t}} &= \phi_0 + \gamma_1 \widehat{z}_{t-1} + \sum_{i=1}^u \phi_{1,i} \Delta y_{1,t-i} + \sum_{j=1}^v \phi_{2,j} \Delta y_{2,t-j}, \\
 \mu_{\Delta y_{1,t}} &= \psi_0 + \gamma_2 \widehat{z}_{t-1} + \sum_{i=1}^u \psi_{1,i} \Delta y_{2,t-i} + \sum_{j=1}^v \psi_{2,j} \Delta y_{1,t-j},
 \end{aligned}$$

with $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{\varepsilon_{1,t}}^2 & \sigma_{\varepsilon_{1,t}, \varepsilon_{2,t}} \\ \sigma_{\varepsilon_{1,t}, \varepsilon_{2,t}} & \sigma_{\varepsilon_{2,t}}^2 \end{pmatrix}$, $\sigma_{\varepsilon_{1,t}, \varepsilon_{2,t}} = \sigma_{\varepsilon_{1,t}} \cdot \sigma_{\varepsilon_{2,t}} \cdot \rho$, and five linear predictors¹:

$$\boldsymbol{\eta}_{coint} = (\mu_{\Delta y_{2,t}}, \mu_{\Delta y_{1,t}}, \log \sigma_{\varepsilon_{1,t}}, \log \sigma_{\varepsilon_{2,t}}, \text{rhobit}(\rho))^T.$$

Thus, for illustrative purposes, an ECM(2, 2) as in (3) will be fitted to (1). Note the use of the family function `binormal()` from [VGAM](#). See the the R code in Table 1.

¹`rhobit`(θ) = $\log \frac{1+\theta}{1-\theta}$, for $-1 < \theta < 1$.

Table 1. R code to fit an ECM(2, 2) to (1) using VGLMs and RR-VGLMs.

```

> ### Using VGLMs
> fit.coint1 <- vglm(cbind(dify1, difx1) ~ errors.cointLag1 + difxLag1 + difyLag1 +
                    difxLag2 + difyLag2,
                    binormal(zero = c("sd", "rho")), # 'sigma', 'rho' intercept-only.
                    trace = FALSE, data = coint.data)

```

The estimated coefficients obtained from `vglm()` are retrieved from the object `fit.coint1`, producing the next output

```

> coef(fit.coint1, matrix = TRUE)

```

	mean1	mean2	loge(sd1)	loge(sd2)	rhobit(rho)
(Intercept)	0.049363	0.0244895	1.3259	0.20276	4.4712
errors.cointLag1	-1.073665	-0.0019501	0.0000	0.00000	0.0000
difxLag1	-0.276705	-0.1462313	0.0000	0.00000	0.0000
difyLag1	0.078858	0.0781579	0.0000	0.00000	0.0000
difxLag2	-0.230352	-0.0691028	0.0000	0.00000	0.0000
difyLag2	0.204452	0.0800435	0.0000	0.00000	0.0000

As expected, $\hat{\gamma}_1 = -1.07$ is negative in sign (and close to unity), assuring the system convergence to its long-run equilibrium path. Overall, results show that $y_{1,t}$ and $y_{2,t}$ are two cointegrated $I(1)$ -variables guaranteeing *Granger causality* in one direction. More precisely, one series may be predicted with help the other.

Further improvements to be incorporated over time are: Employ the Reduced-Rank class of VGLMs (Yee, 2015, RR-VGLMs)) to aid the number of coefficients as u and v increase, to write, e.g., `trinormal()`, and then, write suitable family functions for cointegrated time series (beyond the bivariate case).

References

- R. Engle and G. C. Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55(2):98–1007, 1987.
- V. Miranda and T. Yee. Vector generalized linear time series models. Manuscript in preparation, 2018.
- B. Pfaff. *Analysis of integrated and cointegrated time series with R*. Springer, Seattle, Washington, USA, 2011.
- R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL <https://www.R-project.org/>.
- R. Rigby and D. Stasinopouls. Generalized additive models for location, scale and shape, (with discussion). *Applied Statistics*, 54(3):507–554, 2005.
- T. Yee. *Vector generalized linear and additive models with an implementation in R*. Springer, New York, USA, 2015.