Explanatory Analysis for Semi-Parametric Estimation of Noncausal Vector Autoregression

Christian Gourieroux [*]
Joann Jasiak [†]

0.1 Second-Order Identification of causal and noncausal directions

Let us now explain how to use the autocovariances in order to identify the causal/noncausal directions for given n_{10}, or $n - n_{10}$. More precisely, in a mixed process with causal characteristics (n_1, A^1) and noncausal characteristics (n_2, A^2), we have (see Proposition 2) :

$$
\begin{aligned}
\Gamma^*_{1,2}(h) &= Cov(Y^*_{1,t}, Y^*_{2,t-h}) \\
&= Cov(A^1 Y_t, A^2 Y_{t-h}) \\
&= A^1 \Gamma(h)(A^2)' \\
&= 0 \text{ , for } h \leq 0.
\end{aligned}
\tag{1}
$$

The above covariance conditions are used to consistently estimate the causal and noncausal directions A^1, A^2 for given dimensions $n_1, n_2 = n - n_1$.

step one) **Search for the causal and noncausal directions**

Let us denote the causal/noncausal dimensions by n_1 and $n_2 = n - n_1$, respectively. Then the causal/noncausal directions can be estimated as the solutions of the following constrained minimization:

$$
(\hat{A}^1, \hat{A}^2) = \arg \min_{A^1, A^2} \sum_{h=0}^{-H} ||A^1 \hat{\Gamma}(h)(A^2)'||^2,
\tag{2}
$$

s.t. $A^1 \hat{\Gamma}(0)(A^1)' = Id_{n_1}, A^2 \hat{\Gamma}(0)(A^2)' = Id_{n-n_1},$

where $||C||^2 = Tr(CC'), \hat{\Gamma}(h)$ is the sample counterpart of $\Gamma(h)$, the lag, $H, H > 0$, is sufficiently large, and the standardization concerns matrix A^{-1} directly (see (3.2) in section 3.1). The value of the objective function at the optimum is denoted by $\hat{L}(n_1, n - n_1)$.

In model (2.1) and under Assumptions A1-A2, the sample autocovariances $\hat{\Gamma}(h), h = 0, \ldots, -H$ converge a.s. to their theoretical counterpart $\Gamma(h)$, when the number of observations T tends to infinity. Thus the solutions $\hat{A}^1, \hat{A}^2$ that minimize objective function (3.8) converge a.s. to the solutions of the associated asymptotic minimization, in which the sample-based $\hat{\Gamma}(h)$ are replaced by their theoretical counterparts $\Gamma(h)$. If n_1 is equal to the true causal dimension $n_{1,0}$, then, the asymptotic objective function is

minimized for the true causal and noncausal directions. Thus $\hat{A}^j$ is a consistent estimator of A_0^j, $j = 1, 2$. If n_1 is equal to $n - n_{1,0}$, the minimum of the asymptotic objective function is equal to zero. If n_1 is different from $n_{1,0}$ and $n - n_{1,0}$, the minimum of the asymptotic objective function is strictly positive.

step two) **Identification of causal and noncausal dimensions**

We identify the causal dimension, which is either n_{10}, or $n - n_{10}$ from the residual analysis as follows. For this purpose, we apply the estimation method of A^1, A^2 introduced above, for a given causal dimension n_{10}, say, and compute the estimated causal and noncausal components : $\hat{Y}_{1,t}^* = \hat{A}^1 Y_t$, $\hat{Y}_{2,t}^* = \hat{A}^2 Y_t$. Next, we proceed in the following order:

a) regress $\hat{Y}_{1,t}^*$ on $\hat{Y}_{1,t-1}^*$ to find $\hat{J}_1$ and the associated residuals $\hat{\varepsilon}_{1,t}^*$ [resp. $\hat{Y}_{2,t}^*$ on $\hat{Y}_{2,t+1}^*$ to find $-\hat{J}_2^{-1}$].

b) estimate matrix Φ as $\hat{\Phi} = \hat{A} \begin{pmatrix} \hat{J}_1 & 0 \\ 0 & \hat{J}_2 \end{pmatrix} \hat{A}^{-1}$, and compute the mixed residuals $\hat{\varepsilon} = Y_t - \hat{\Phi} Y_{t-1}$.

c) plot the nonlinear ACF of $(\hat{\varepsilon}_t)$ [1]. If the nonlinear ACF are not significant, the process is mixed $(n_{10}, n - n_{10})$. Otherwise, we have the mixed process $(n - n_{10}, n_{10})$.

The above identification procedure relies mainly on moments up to order two, except for the analysis of nonlinear autocorrelograms of the residuals, which requires nonlinear methods and relies on the serial independence of the error terms.

We have mentioned in the introduction the importance of noncausal components that capture speculative bubbles, in processes with fat-tailed errors $\varepsilon_{2,t}^*$. The presence of fat tails is not compatible with the existence of second-order moments of the error term. Nevertheless, the procedure described above relies on the sample autocovariances $\hat{\Gamma}(h)$ and not on the theoretical autocovariances $\Gamma(h)$. It is known that $\hat{\Gamma}(h)$ can preserve the consistency and asymptotic distributional properties, even in the presence of fat tail errors, such as errors with stable distributions [see e.g. Davis, Resnick (1986)]. Therefore, the method proposed above will provide consistent estimators of

---

[1] See Gourieroux, Jasiak (2001) for the definition and implementation of nonlinear autocorrelograms.

A^1, A^2, provided that the standardization $A^{j'}\hat{\Gamma}(0)A^j = Id$ is used in order to control the possibly different speeds of convergence of the elements of $\hat{\Gamma}$.

1 Illustration

In this section, we first illustrate the application of the Generalized Covariance estimator introduced in Section 4 to simulated data and discuss its finite sample properties. Next, the GCov estimator is used to analyze the dynamics of commodity futures.

1.1 The simulated data

Let us consider a bivariate process $n = 2$ of causal and noncausal dimensions equal to 1: $n_1 = n - n_1 = 1$. The following parameter values are fixed: $J_1 = 0.7, J_2 = 2$,

$$A = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, A^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

The errors $\epsilon_t = (\epsilon_{1,t}, \epsilon_{2,t})'$ are such that $\epsilon_{1,t}, \epsilon_{2,t}$ are drawn independently in the same t-Student distribution with the degree of freedom $\nu = 4$, zero mean and variance equal to $\nu/(\nu - 2)$.

The autoregressive matrix is equal to:

$$\Phi = A \begin{pmatrix} J_1 & 0 \\ 0 & J_2 \end{pmatrix} A^{-1} = \begin{pmatrix} 0.7 & -1.3 \\ 0.0 & 2.0 \end{pmatrix}.$$

We create a sample of length $T = 1000$ by simulating 2000 errors $\epsilon_t^s, t = 1, ..., T$, and by computing the simulated transformed errors $\epsilon_t^{*s} = A^{-1}\epsilon_t^s$, from which the values of the causal component $Y_{1,t}^{*s} = J_1 Y_{1,t-1}^{*s} + \epsilon_{1,t}^{*s}, t = 1, ..., T$, with initial value $Y_{1,0}^{*s} = 0$, and the values of the noncausal component: $Y_{2,t}^{*s} = 1/J_2 Y_{2,t+1}^{*s} - 1/J_2 \epsilon_{2,t+1}^{*s}, t = 1, ..., T$, with terminal value $Y_{2,T+1}^{*s} = 0$ are obtained. Next, we compute the values of the series $Y_t^s = AY_t^{*s}, t = 1, ..., T$ and discard the first and last 500 realizations. They are related to the observed process as follows: $Y_{1,t} = Y_{1,t}^*, Y_{2,t} = Y_{1,t}^* + Y_{2,t}^* \iff Y_{1,t}^* = Y_{1,t}, Y_{2,t}^* = Y_{2,t} - Y_{1,t}$. Hence, the first component of Y_t is purely causal and its second component is a mixture of a causal and a noncausal process.

4

Figure 1 shows the path of the two components of the observed process Y^s and Figure 2 shows its autocorrelation function.

[Figure 1: Simulated Y_t]

The first observed component has mean -0.147 and variance 8.296 and the second component has mean 0.028 and variance 0.633. Their contemporaneous correlation is -0.221.

[Figure 2: Autocorrelation Function of Y_t]

The simulated data display multiple peaks in the trajectory due to the fat tails of the errors. Indeed, for the selected value of parameter $\nu = 4$, the kurtosis of the errors does not exist. The marginal and cross autocorrelations are significant up to lag 10 with exponential decay rates determined by the values of $J_1 = 0.7$ and $1/J_2 = 0.5$.

Let us now consider the auto- and cross-correlations of the causal and non-causal components of Y_t^*.

[Figure 3: Autocorrelation Function of Y_t^*]

The cross-correlations are almost all not significant in the South-West panel of Figure 3. This illustrates the condition: $\Gamma_{1,2}^*(h) = 0$, for $h \leq 0$, derived in Proposition 2, and used in the exploratory analysis below.

1.2   The exploratory analysis

Let us first apply the exploratory analysis along the lines of Section 3.4 [see also Appendix 5 in the complementary material] and consider various possible combinations that are:

$(n_1, n - n_1) = (2,0)$ for a pure causal process,
$(n_1, n - n_1) = (0,2)$ for a pure noncausal process,
$(n_1, n - n_1) = (1,1)$ for a mixed causal/noncausal process.

Pure causal model

In the pure causal model, matrix Φ is estimated by the Ordinary Least Squares (OLS) from the Seemingly Unrelated Regression of Y_t on Y_{t-1}. We get the estimated matrix $\hat{\Phi} = \begin{pmatrix} 0.721 & -1.272 \\ 0.008 & 0.473 \end{pmatrix}$ with eigenvalues $\lambda_1 = 0.67$ and $\lambda_2 = 0.52$. As expected, the eigenvalues are close to $J_1 = 0.7$ and $1/J_2 = 0.5$. The explosive root is captured by the estimate of its stationary counterpart.

5

It is clear that the pure causal model is **misspecified** as the second row of matrix $\hat{\Phi}$ is very different from the second row of the true matrix Φ. In practice, the true matrix Φ is unknown and such a misspecification will be detected from the analysis of the causal residuals.

Figure 4 below displays the ACF of the SUR-based causal residuals: $\hat{\epsilon}_t = Y_t - \hat{\Phi}Y_{t-1}$.

[Figure 4: Autocorrelation Function of Causal Residuals]

The correlations are not significant, which implies that the causal residuals can be considered as weak white noises. Thus, the misspecification cannot be detected from the second-order properties of the residuals alone. Let us now consider the ACF computed from the squared causal residuals.

[Figure 5: Autocorrelation Function of Squared Causal Residuals]

We observe significant autocorrelations in the South-East panel. This implies that causal errors ϵ_t are not serially independent. Thus, the pure causal dynamics is rejected.

Pure noncausal process

A similar approach is used for the pure noncausal process. More precisely, the initial model $Y_t = \Phi Y_{t-1} + \epsilon_t$ is transformed into its forward-looking representation $Y_t = \Phi^{-1}Y_{t+1} - \Phi^{-1}\epsilon_{t+1}$. Therefore, matrix Φ^{-1} can be estimated by the OLS in the SUR regression of Y_t on Y_{t+1}.

The estimated autoregressive coefficient is $\hat{\Phi}^{-1} = \begin{pmatrix} 0.823 & 0.392 \\ -0.117 & 0.368 \end{pmatrix}$. Its inverse provides the estimate of Φ: $\hat{\Phi} = \begin{pmatrix} 1.054 & -1.122 \\ 0.336 & 2.356 \end{pmatrix}$. The eigenvalues of $\hat{\Phi}$ are $\lambda_1 = 1.492$ and $\lambda_2 = 1.917$ and are close to $1/J_1 = 1.428$ and $J_2 = 2.0$. The autocorrelation functions of the noncausal residuals and their squares are provided in Figures 6 and 7.

[Figure 6: Autocorrelation Function of Noncausal Residuals]

[Figure 7: Autocorrelation Function of Squared Noncausal Residuals]

The noncausal residuals satisfy the weak white noise condition. However, the pure noncausal specification is rejected due to significant squared autocorrelations of the squared residuals in the top panels.

Mixed process

In the mixed case, we estimate the rows A^1, A^2 of matrix A^{-1} by minimizing the objective function (3.8). This constrained minimization involves the autocovariances up to lag $H = 4$ and yields the estimated matrix $\hat{A}^{-1} = \begin{pmatrix} 0.356 & 0.293 \\ -0.008 & 1.250 \end{pmatrix}$. Let us now compare the matrices $\hat{A}^{-1}$ and A^{-1}. We know that A^1, A^2 are defined up to linear invertible transforms (as well as the associated causal and noncausal components) (see the discussion in Section 3.1). Therefore in our framework, we verify if the first rows of $\hat{A}^{-1}$ and A^{-1} (resp. the second rows) are close to being proportional. The cosine between the row vectors are:

$cos_1 = 0.995$, $cos_2 = 0.999$, showing a quasi-proportionality.
Given these estimates, we compute the fitted components:
$\hat{Y}_{1,t}^* = \hat{A}^1 Y_t$, $\hat{Y}_{2,t}^* = \hat{A}^2 Y_t$, by Corollary 2. Figure 8 displays the scatterplots of $(\hat{Y}_{j,t}^*, Y_{j,t}^*), j = 1, 2$.

[Figure 8: Scatterplots of Fitted and True Components]

The true and fitted components satisfy a quasi-linear relationship, which is compatible with the definition of these components up to a multiplicative scalar. The R^2 of the associated regressions are $R_1^2 = 0.997$ and $R_2^2 = 0.999$, respectively. Let us now consider the auto- and cross-correlations of $\hat{Y}_t^*$.

[Figure 9: Autocorrelation Function of $\hat{Y}_t^*$]

As expected, the autocorrelations of $\hat{Y}_t^*$ in the South-West panel are almost nonsignificant.

The regression coefficient obtained by regressing $\hat{Y}_{1,t}^*$ on $\hat{Y}_{1,t-1}^*$ [resp. $\hat{Y}_{2,t}^*$ on $\hat{Y}_{2,t+1}^*$] provide the estimated values $\hat{J}_1 = 0.725$ and $1/\hat{J}_2 = 0.472$. Given these and the previously estimated matrix $\hat{A}^{-1}$ the estimated Φ matrix is:

$\hat{\Phi} = \begin{pmatrix} 0.732 & -1.141 \\ -0.008 & 2.111 \end{pmatrix}$.

Next, we compute the mixed causal-noncausal residuals as:
$\hat{\epsilon}_t = Y_t - \hat{\Phi} Y_{t-1}$,
and display the ACF of the mixed residuals and of the squared mixed residuals in Figures 10 and 11.

[Figure 10: Autocorrelation Function of Mixed Residuals]

[Figure 11: Autocorrelation Function of Squared Mixed Residuals]

All autocorrelations are non-significant and the mixed causal/noncausal model is not rejected.

The exploratory analysis outlined may be considered as a preliminary step prior to applying more sophisticated estimation methods. It provides values of J_1, J_2, A, Φ, which can be used to initiate the algorithms for computing more efficient semi-parametric estimators, such as the Generalized Covariance estimator discussed in the next section.

1.3   Application to commodity futures

[Insert Figure D: VAR(1) Residuals]

[Insert Figure E: VAR(3) residuals]

Appendix 5

**Complementary Material: Direct Search of the Causal and
Noncausal Directions**

A.5.1 Discussion of Optimization (3.8)

The constrained minimization (3.8) is similar to the canonical correlation analysis. We are looking for the linear transformations A^1, A^2, which are the least correlated at any nonpositive lag [2]. Therefore, one could replace the global optimization of objective function (3.8) by a recursive optimization in the spirit of the recursive optimization in Independent Component Analysis (ICA) [see e.g. Ilmonen et al. (2012)].

The optimization of objective function (3.8) involves n^2 arguments, that are the elements of matrix A^{-1}, subject to $\dfrac{n_1(n_1 + 1)}{2} + \dfrac{(n - n_1)(n - n_1 + 1)}{2}$ constraints; thus the number of functionally independent arguments is equal to :

$$2n_1(n - n_1) + \frac{n_1(n_1 - 1)}{2} + \frac{(n - n_1)(n - n_1 - 1)}{2}. \qquad (a.3)$$

In standard economic applications the dimension n is rather small, and the global optimization of objective function (3.8) is easily performed. Table 1 below provides the number of independent arguments.

Table 1 : Number of independent arguments

	Causal dimension n_1	
Size n	$n_1 = 1$	$n_1 = 2$
2	2	/
3	5	/
4	9	10
5	14	16

---

[2]The objective function in (3.8) has an analogue in Second-Order Blind Identification (SOBI) of jointly uncorrelated, but serially correlated sources [see Belouchrani et al. (1997)]. The criterion is of the type $\sum\limits_{h=1}^{h} ||A\hat{\Gamma}(h)A'||^2$, s.t. $A\hat{\Gamma}(0)A' = Id$.

There are no numerical outcomes for $n_1 = 0$, which is a degenerate case. For $n_1 > n/2$ the outcomes are symmetric.

For each lag h, we get $n_1(n - n_1)$ quadratic functions of A^1, A^2 to be minimized by means of the norm $|| \quad ||^2$, and in the global optimization $(H + 1)n_1(n - n_1)$ such quadratic elements. Thus there is a minimum value of H to select for given size n and causal dimension n_1 in order to have a unique minimizer of objective function (3.8). The order condition, that is the minimum H, is given in Table 2.

Table 2 : Order Condition, i.e. Minimum Value of H

	Causal dimension n_1		
Size n	$n_1 = 1$	$n_1 = 2$	uniform in n_1
2	1	/	1
3	2	/	2
4	2	2	2
5	3	2	3

A.5.2 First-order conditions

Let us now derive the first-order conditions (FOC) to the constrained minimization of objective function (3.8). Below, it is shown how to eliminate the estimated Lagrange multipliers from the FOC in order to obtain a system defining $\hat{A}^1, \hat{A}^2$ only. The row vectors of matrix A^1 (resp. A^2) are denoted by $a_i^1, i = 1, \ldots, n_1$ (resp. $a_k^2, k = 1, \ldots, n_2$) .
The objective function to be minimized is :

$$\min_{a_i^1, a_k^2} \sum_{h=0}^{-H} \sum_i \sum_k (a_i^{1'} \hat{\Gamma}(h) a_k^2)^2, \qquad (a.4)$$

$$\text{s.t. } a_i^{1'} \hat{\Gamma}(0) a_i^1 = 1, \forall i, a_i^{1'} \hat{\Gamma}(0) a_j^1 = 0, \forall i < j, \qquad (a.5)$$
$$a_k^{2'} \hat{\Gamma}(0) a_k^2 = 1, \forall k, a_k^{2'} \hat{\Gamma}(0) a_l^2 = 0, \forall k < l. \qquad (a.6)$$

Let us introduce the Lagrange multipliers : $\lambda_{i,i}/2, \lambda_{i,j}, i < j, \mu_{k,k}/2, \mu_{k,l}, k < l$, associated with the orthonormality restrictions. The FOC for the optimization of the Lagrangian are :

10

$$\frac{\partial \mathcal{L}}{\partial a_i^1} = 0 \Leftrightarrow \Sigma_h \Sigma_k (\hat{a}_i^{1'} \hat{\Gamma}(h) \hat{a}_k^2) \hat{\Gamma}(h) \hat{a}_k^2 - \hat{\lambda}_{i,i} \hat{\Gamma}(0) \hat{a}_i^1 - \Sigma_{j,j>i} \hat{\lambda}_{i,j} \hat{\Gamma}(0) \hat{a}_j^1 = 0, \forall i,$$

$$\frac{\partial \mathcal{L}}{\partial a_k^2} = 0 \Leftrightarrow \Sigma_h \Sigma_i (\hat{a}_i^{1'} \hat{\Gamma}(h) \hat{a}_k^2) \hat{\Gamma}(h)' \hat{a}_i^1 - \hat{\mu}_{k,k} \hat{\Gamma}(0) \hat{a}_k^2 - \Sigma_{l:l>k} \hat{\mu}_{k,l} \Gamma(0) \hat{a}_l^2 = 0, \forall k.$$

The Lagrange multipliers can be eliminated from the FOC by premultiplying the FOC by $\hat{a}_j^1, j < i$, and by $\hat{a}_l^2, l < k$, respectively. We get the following set of FOC for the estimates of A^1, A^2, only :

$$
\begin{cases}
\Sigma_h \Sigma_k [(\hat{a}_i^{1'} \hat{\Gamma}(h) \hat{a}_k^2)(\hat{a}_j^{1'} \hat{\Gamma}(h) \hat{a}_k^2)] = 0, \forall j < i, \\[2ex]
\Sigma_h \Sigma_i [(\hat{a}_i^{1'} \hat{\Gamma}(h) \hat{a}_k^2)(\hat{a}_i^{1'} \hat{\Gamma}(h) \hat{a}_l^2)] = 0, \forall l < k, \\[2ex]
\hat{a}_i^{1'} \hat{\Gamma}(0) \hat{a}_i^1 = 1, \forall i, \hat{a}_i^{1'} \hat{\Gamma}(0) \hat{a}_j^1 = 0, \forall i < j, \\[2ex]
\hat{a}_k^{2'} \hat{\Gamma}(0) \hat{a}_k^2 = 1, \forall k, \hat{a}_k^{2'} \hat{\Gamma}(0) \hat{a}_l^2 = 0, \forall k < l.
\end{cases}
\qquad (a.7)
$$
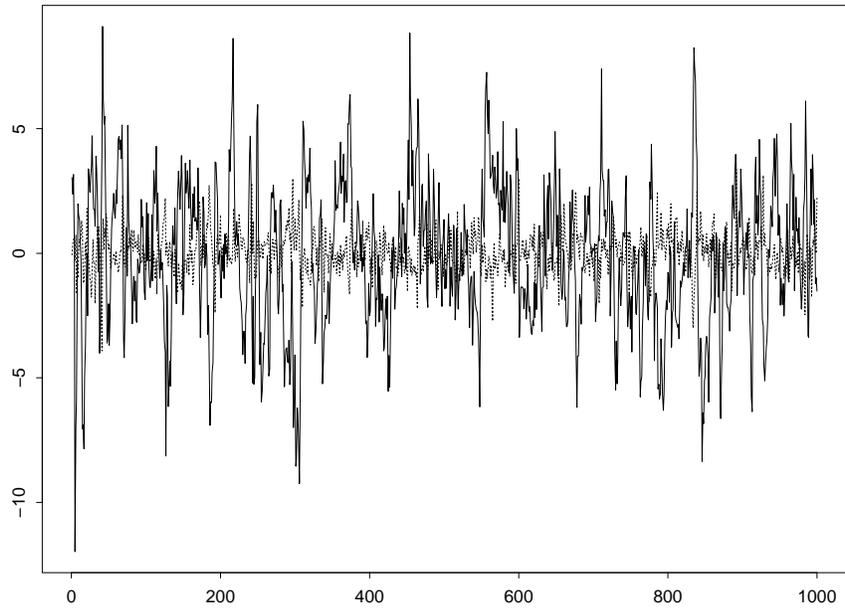
Figure 1: Simulated Y_t: Y_{1t}-solid line, Y_{2t}-dashed line

Multivariate Series : y
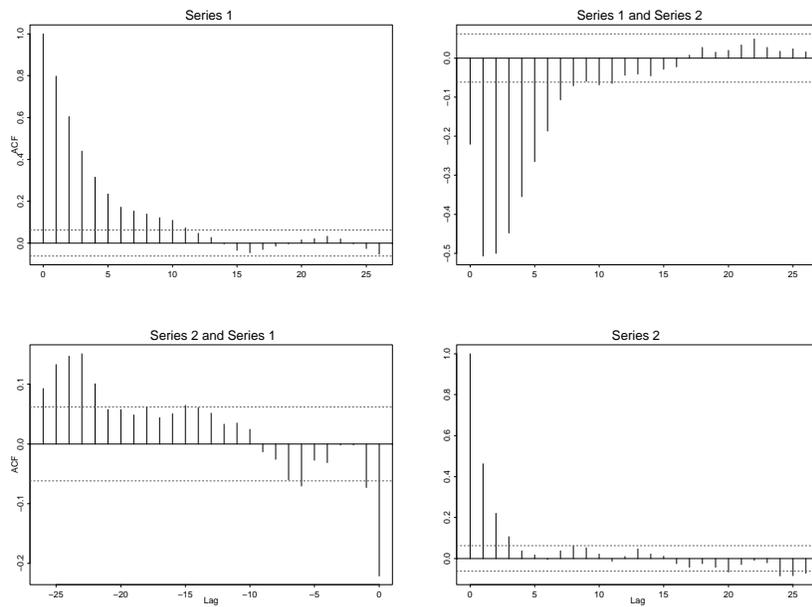
Figure 2: Autocorrelation Function of Y_t
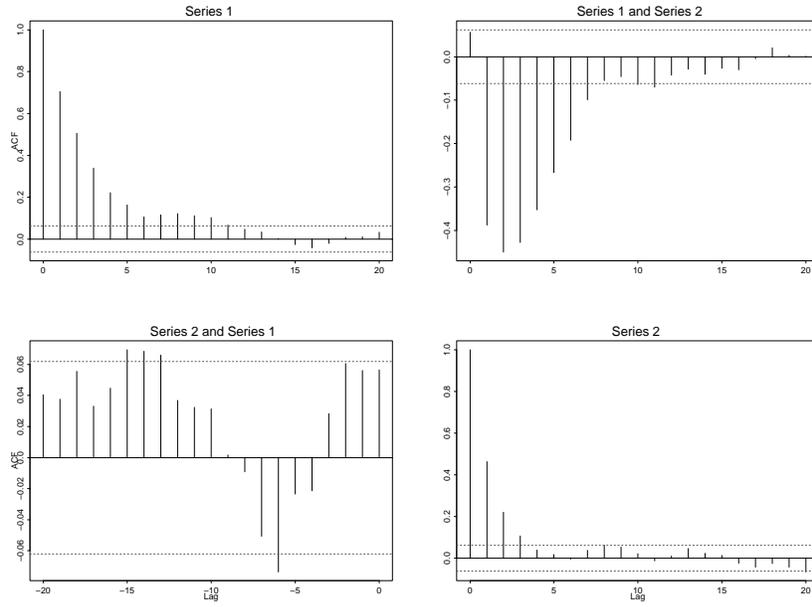
Figure 3: Autocorrelation Function of Y_t^*
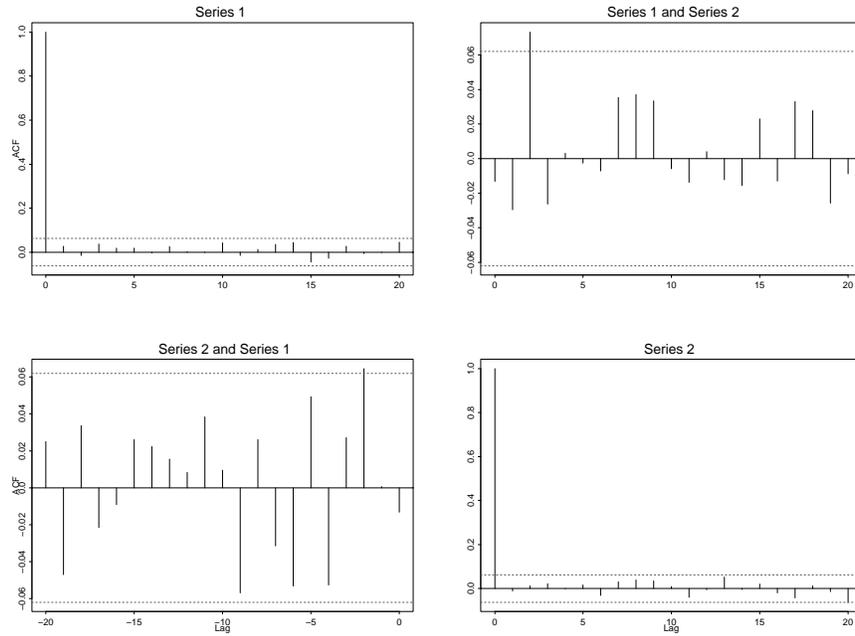
Figure 4: Autocorrelation Function of Causal Residuals
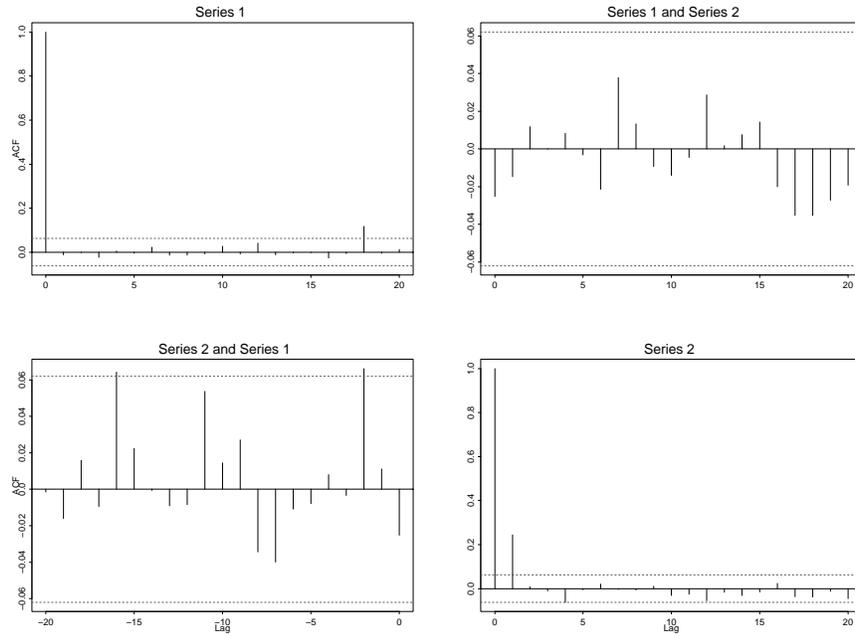
Figure 5: Autocorrelation Function of Squared Causal Residuals
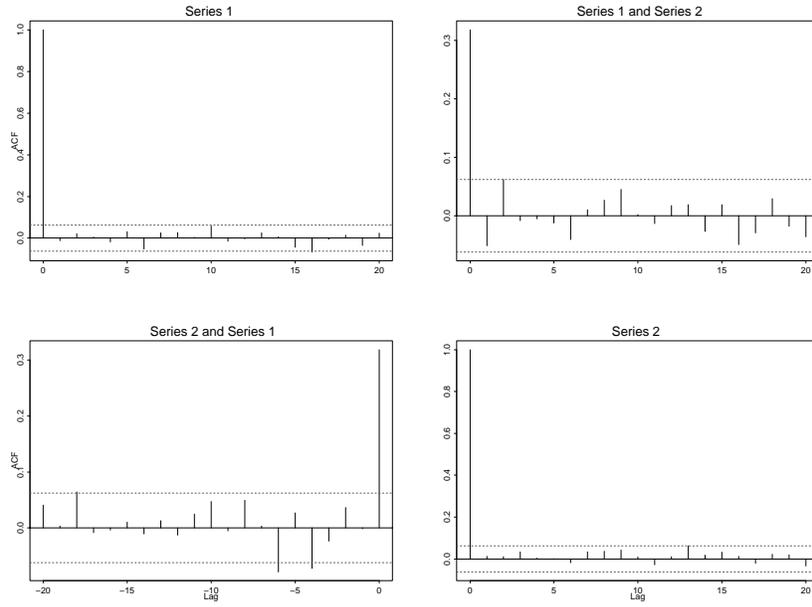
Figure 6: Autocorrelation Function of Noncausal Residuals
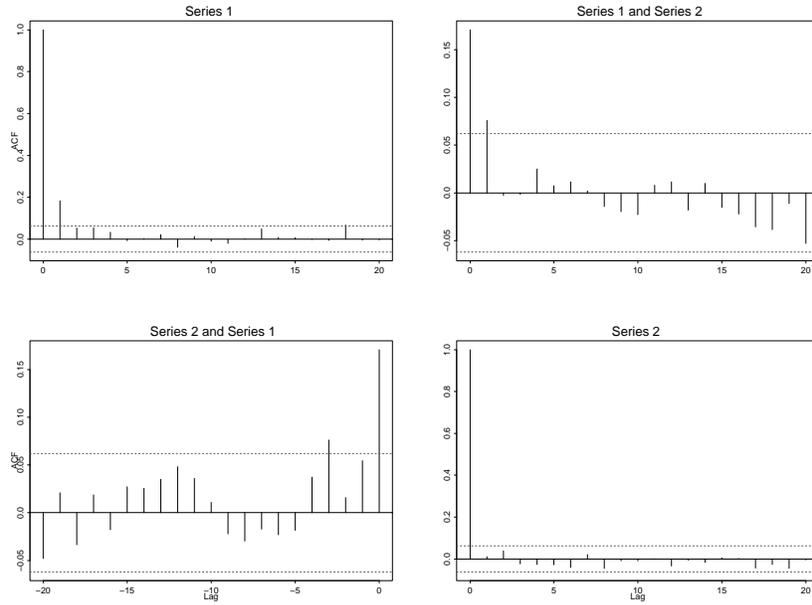
Figure 7: Autocorrelation Function of Squared Noncausal Residuals
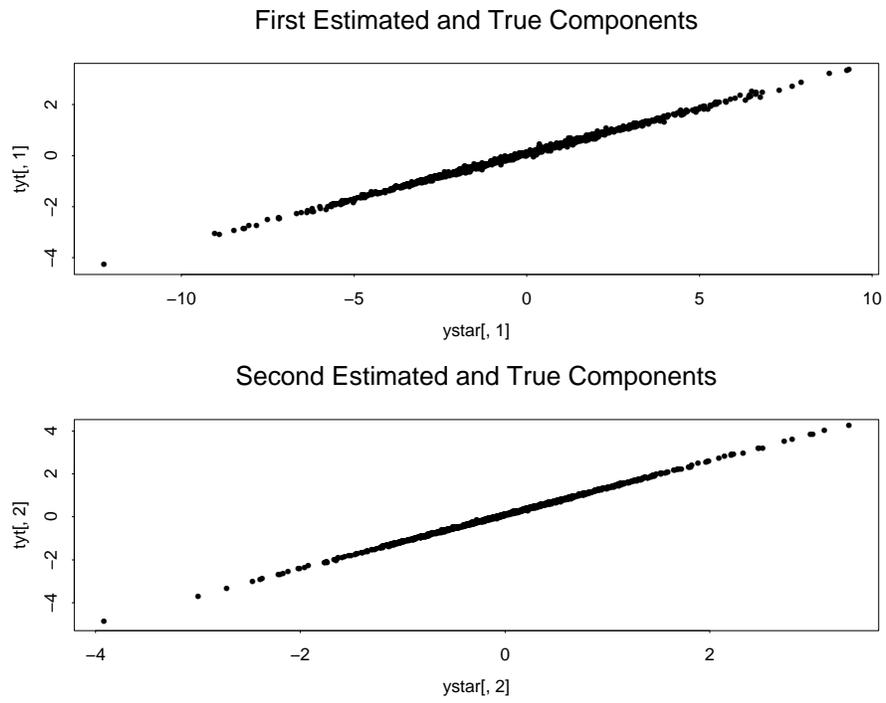
Figure 8: Scatterplots of Fitted and True Components
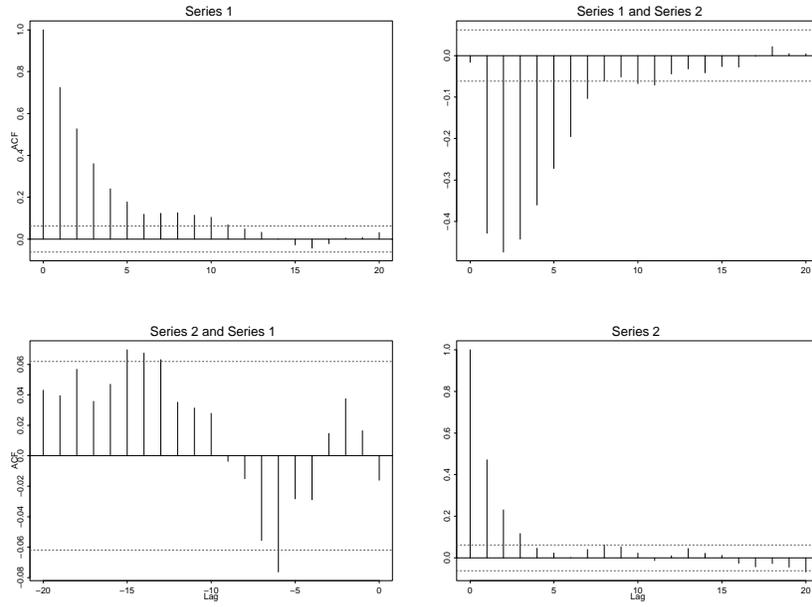
Figure 9: Autocorrelation Function of $\hat{Y}_t^*$
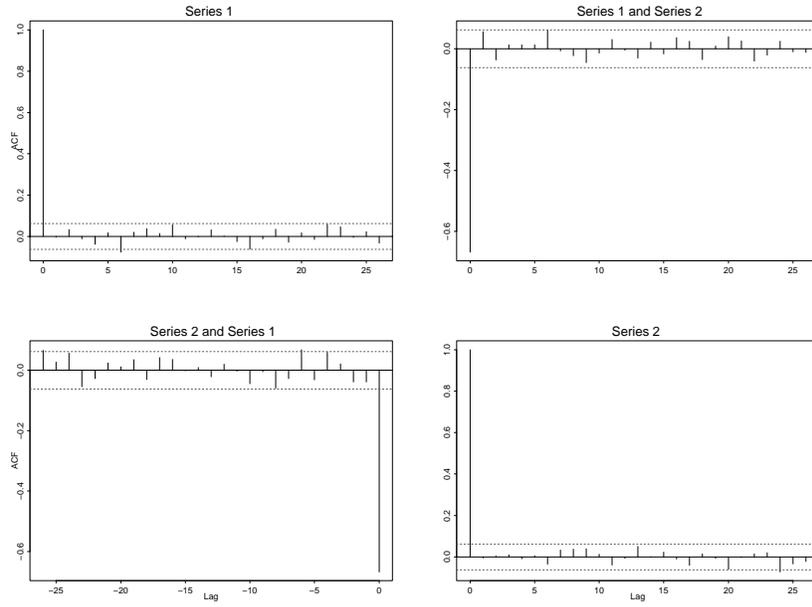
Figure 10: Autocorrelation Function of Mixed Residuals
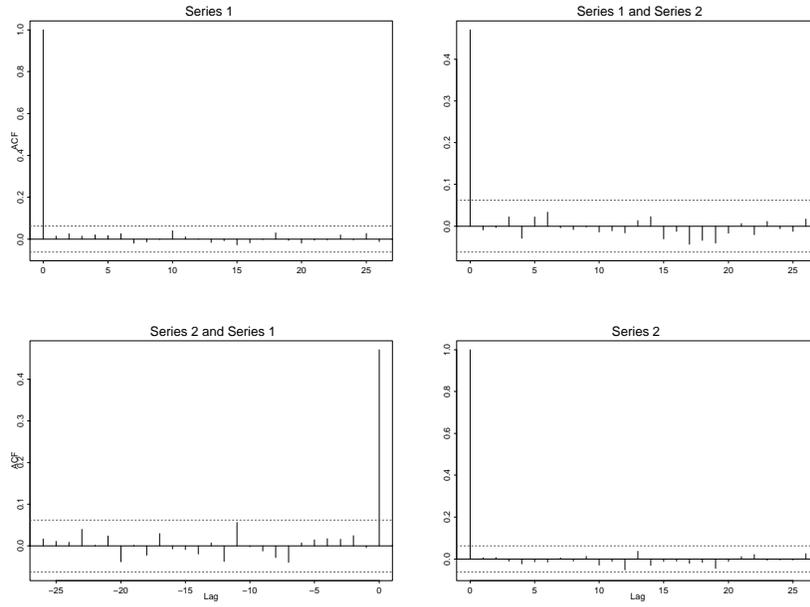
Figure 11: Autocorrelation Function of Squared Mixed Residuals
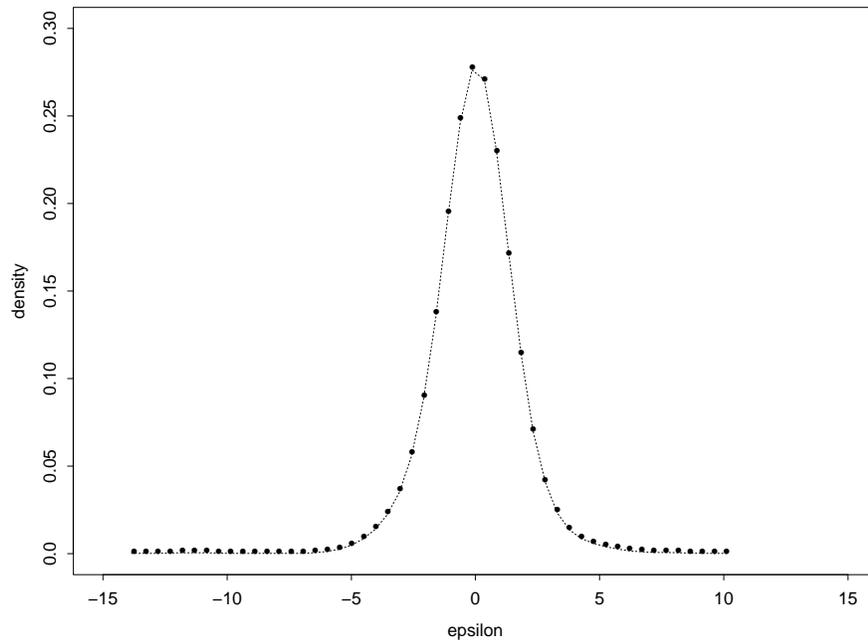
Figure 12: Empirical Density of GCov Residuals *dotted line* and True Errors *dot symbols* (causal component)
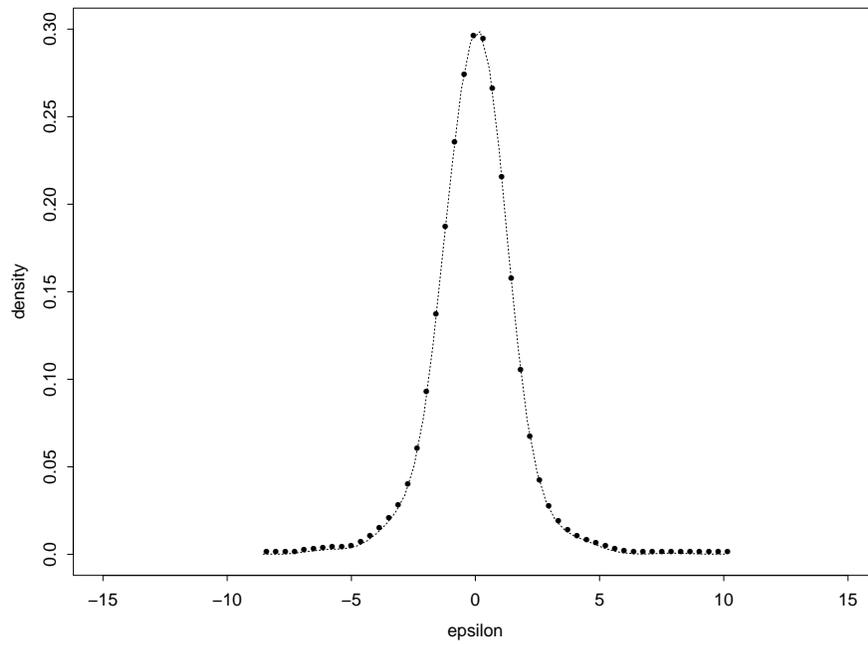
Figure 13: Empirical Density of GCov Residuals *dotted line* and True Errors *dot symbols* (noncausal component)

US Corn Futures

US Soybean Futures
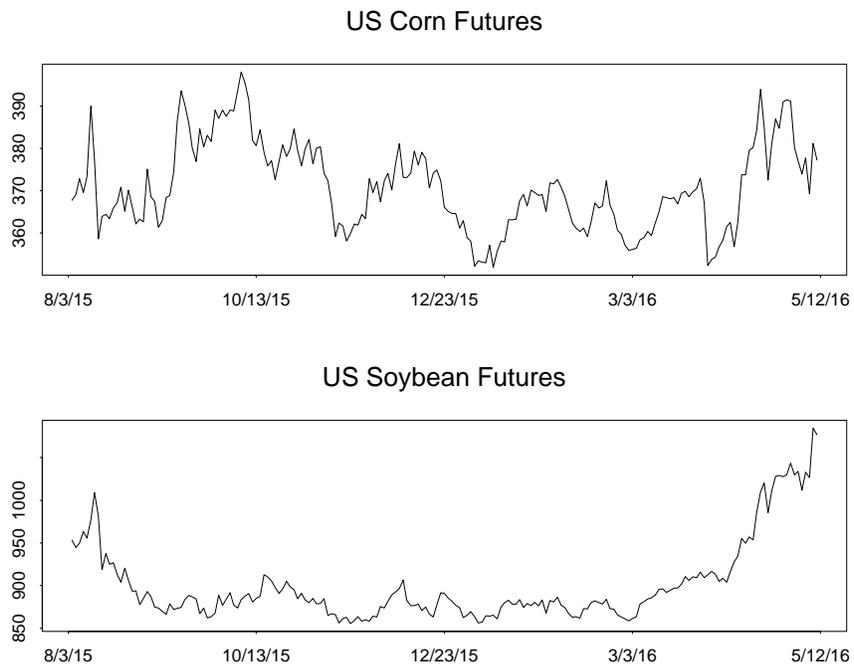
Figure 14: Evolution of Commodity Futures
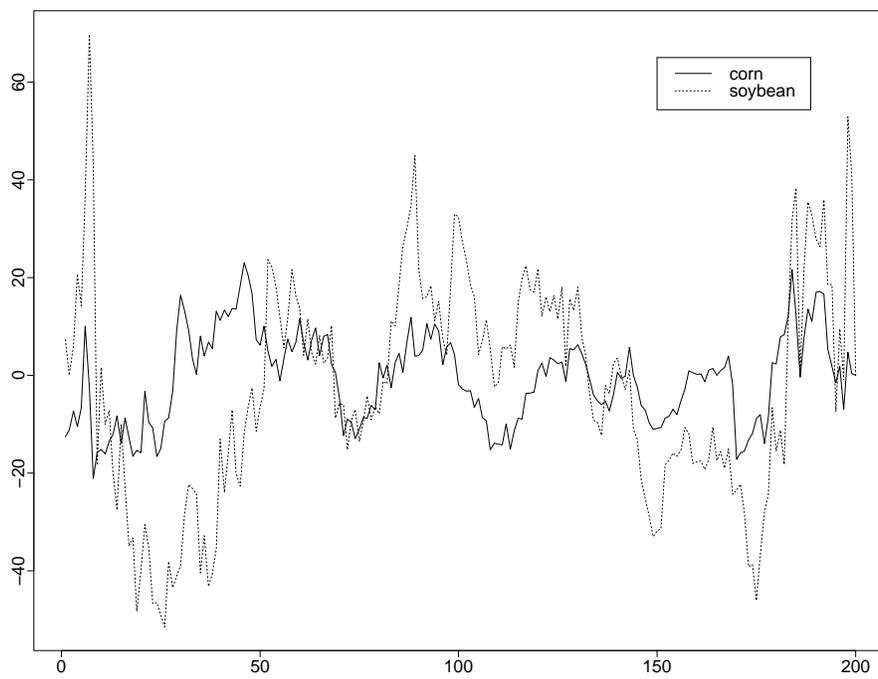
Figure 15: Adjusted data
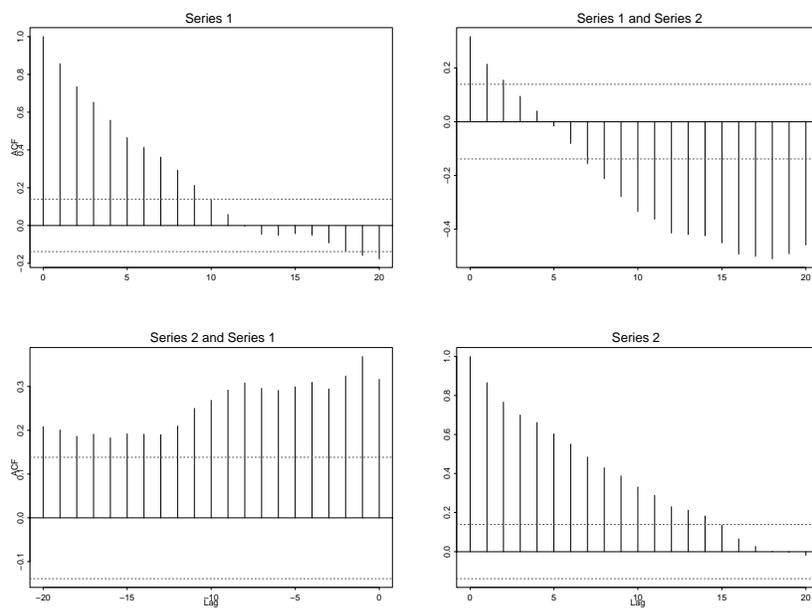
Figure 16: ACF, Adjusted data
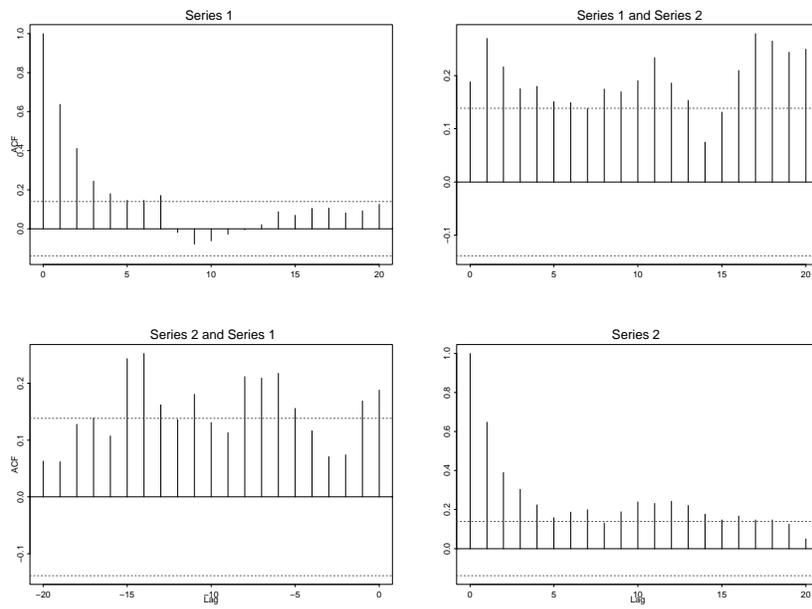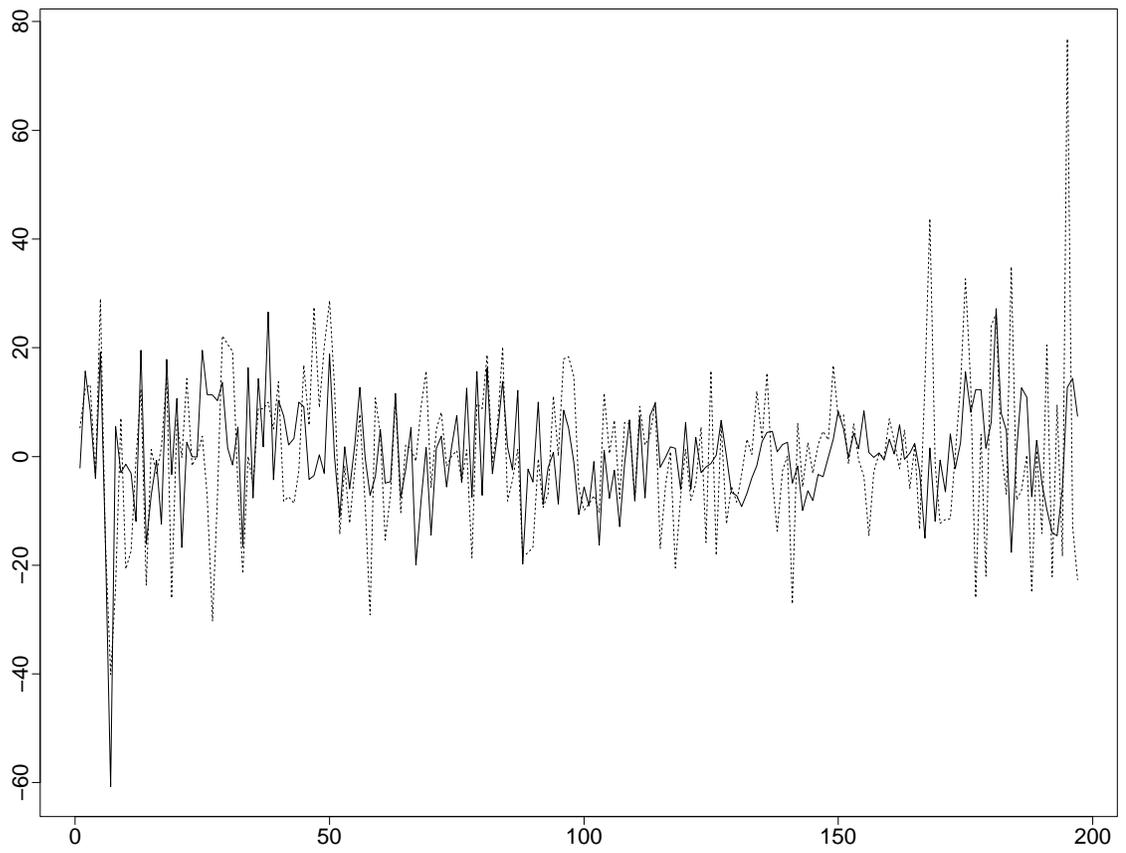
Figure 17: ACF, Squared Adjusted data
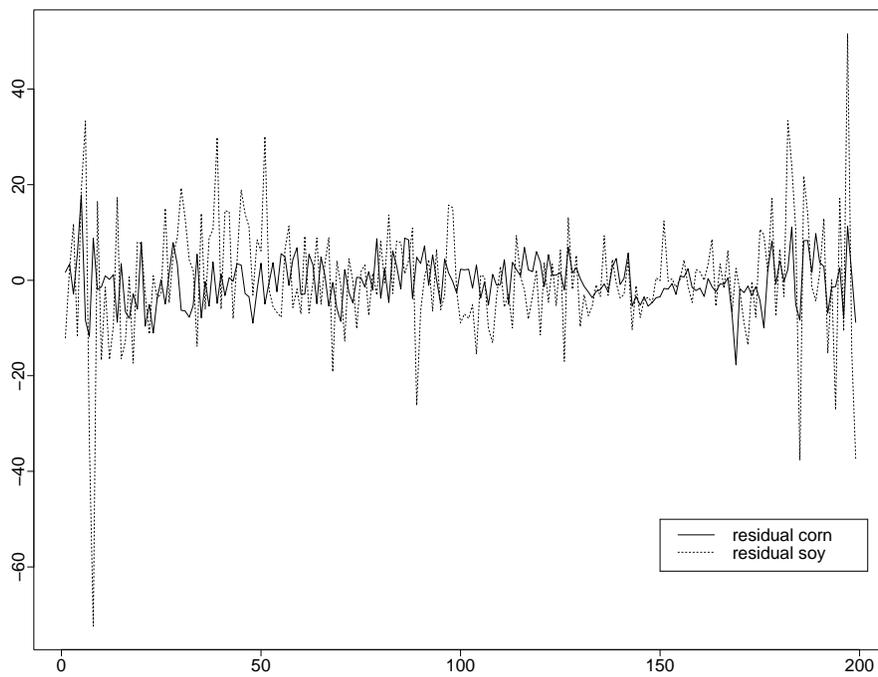
Figure 18: VAR(1) residuals
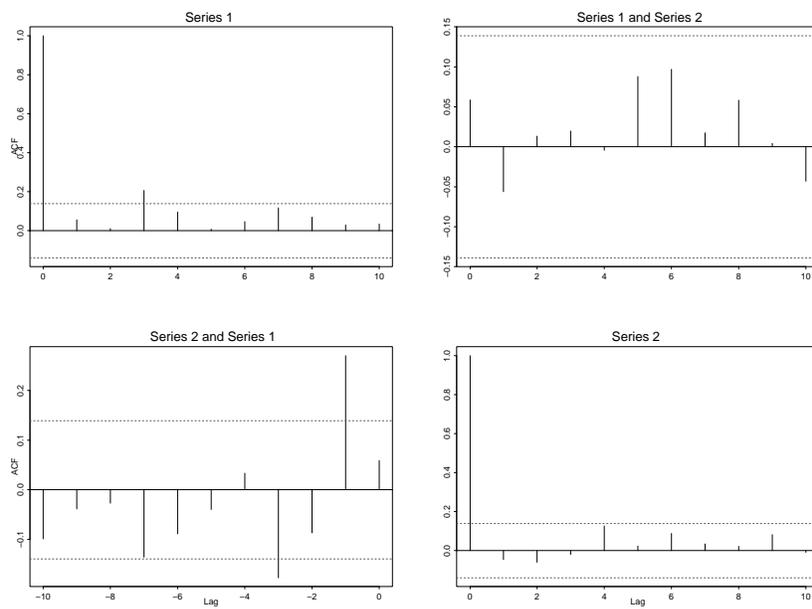
Figure 19: ACF, VAR(1) residuals
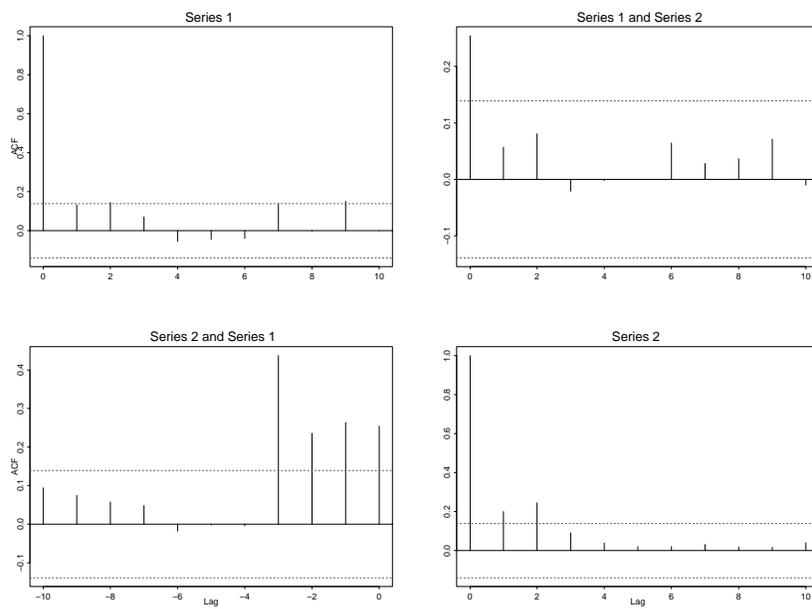
Figure 20: ACF, Squared VAR(1) residuals
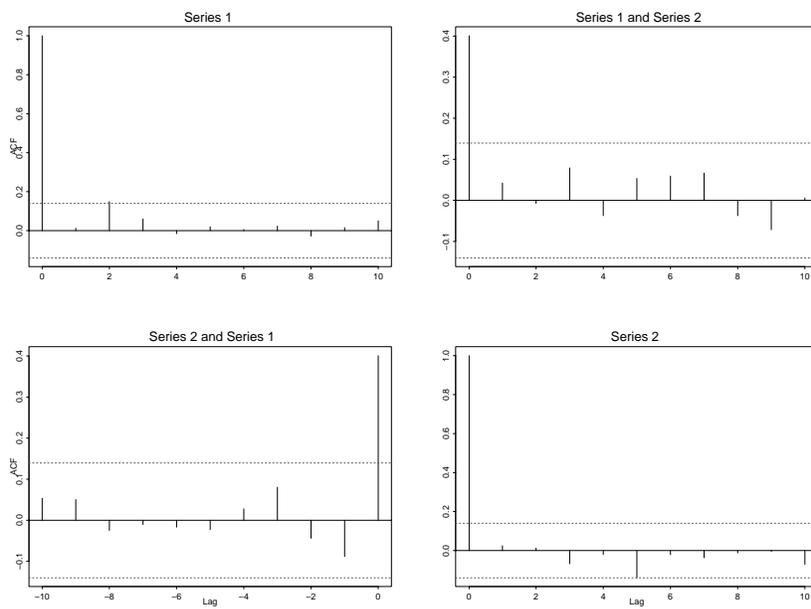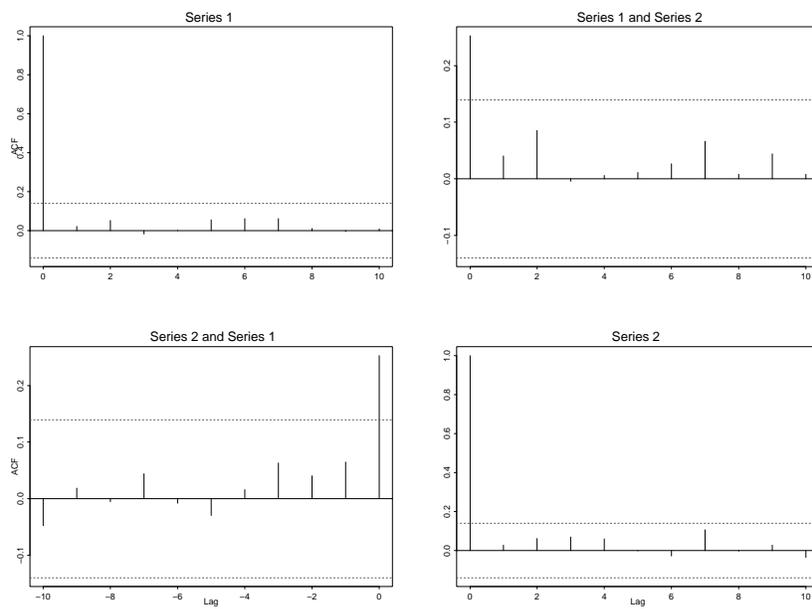
Figure 21: VAR(3) residuals

Figure 22: ACF, VAR(3) residuals

Figure 23: ACF, Squared VAR(3) residuals