Supplementary Material to Structural Modelling of Dynamic Networks and Identifying Maximum Likelihood

Gourieroux*, C. and J. Jasiak†

*University of Toronto, Toulouse School of Economics and CREST, e-mail: *Christian.Gourieroux@ENSAE.fr.*

†York University, e-mail: *jasiakj@yorku.ca.*

On-line Appendix 1

Comparison of the Decompositions

1.1 The NMF

After the transformation the new NMF is :

$$\begin{aligned} A &= \tilde{\beta}_1\tilde{\gamma}_1' + \tilde{\beta}_2\tilde{\gamma}_2', \\ \text{with } \tilde{\beta}_1 &= \beta_1 + q_{21}\beta_2 = q_{12}\beta_1 + \beta_2, \\ \tilde{\gamma}_1 &= \frac{1}{1-q_{12}q_{21}}(\gamma_1 - q_{12}\gamma_2), \tilde{\gamma}_2 = \frac{1}{1-q_{12}q_{21}}(-q_{21}\gamma_1 + \gamma_2). \end{aligned}$$

It is easily checked that $A = \beta_1\gamma_1' + \beta_2\gamma_2'$.

1.2 The decomposition (2.13).

To get the new decomposition (2.13) :

$$A = a(\tilde{\pi}\tilde{\beta}_1^*\tilde{\gamma}_1^{*'} + (1-\tilde{\pi})\tilde{\beta}_2^*\tilde{\gamma}_2^{*'}],$$

the new factorial directions have to be normalized with the components that sum up to one. We get :

$$\begin{aligned} \tilde{\beta}_1^* &= (\beta_1 + q_{21}\beta_2)/(\beta_1'e + q_{21}\beta_2'e), \\ \tilde{\beta}_2^* &= (q_{12}\beta_1 + \beta_2)/(q_{12}\beta_1'e + \beta_2'e), \\ \tilde{\gamma}_1^* &= (\gamma_1 - q_{12}\gamma_2)/(\gamma_1'e - q_{12}\gamma_2'e), \\ \tilde{\gamma}_2^* &= (-q_{21}\gamma_1 + \gamma_2)/(-q_{21}\gamma_1'e + \gamma_2'e), \\ \tilde{\pi}/(1-\tilde{\pi}) &= (\beta_1'e + q_{21}\beta_2'e)(\gamma_1'e - q_{12}\gamma_2'e)/[(q_{12}\beta_1'e + \beta_2'e)[-q_{21}\gamma_1'e + \gamma_2'e)]. \end{aligned}$$

On-line Appendix 2

Additional Assumptions for Asymptotic Results

2.1 Consistency

We provide below a set of additional assumptions a.1 to get the consistency. They require some uniform convergence of the objective function on the set $\mathcal{A}^*$ of all possible α on which the optimization is performed.

Assumption a.1 :

i) The set $\mathcal{A}^*$ is compact.

ii) $\log l(y_t|y_{t-1};\alpha)$ is integrable for all $\alpha \in \mathcal{A}^*$.

iii) $\mathcal{A}_0 \subset \mathcal{A}^*$.

iv) Uniform convergence of the objective function :

$$\sup_{\alpha\in\mathcal{A}^*} |\frac{1}{T}\sum_{t=1}^{T} \log l(y_t|y_{t-1};\alpha) - E_0 \log l(y_t|y_{t-1};\alpha)|$$

$$= 0_P(1/\sqrt{T}),$$

where E_0 is the expectation with respect to the true stationary distribution of (Y_{t-1}, Y_t).

v) $\lim_{p\to\infty} \frac{1}{T}\sum_{t=1}^{T} \log l(y_t|y_{t-1};\hat{\alpha}_T(\alpha_0,p)) = \max_{\alpha\in\mathcal{A}^*} \frac{1}{T}\sum_{t=1}^{T} \log l(y_t|y_{t-1};\alpha)$.

The three first conditions are standard and used to prove the convergence to $\mathcal{A}_0$ of the set of solutions of the finite sample optimisations. They imply condition C_1 on Chernozhukov et al. (2007) for instance (See also this reference for a proof, in which, in our framework, the objective function is the log-likelihood function instead of a moment criterion function). This is the convergence result in Proposition 3. Proposition 4 follows since this convergence of sets is uniform.

The last condition iv) is usually not introduced. It concerns the algorithm used to approximate the solutions of the finite sample optimisations. This condition explains why in Assumption A.4 we have introduced stronger conditions on the concavity of the log-likelihood function with respect to A.

By construction the domain for $\pi, \beta_k^*, \gamma_k^*, k = 1, \ldots, K$ is compact and its bounds

as $\pi_k = 0$, for some k, or $\beta_k = 0$, for some k cannot be reached due to the rank condition, i.e. Assumptions A.2 and A.3. Therefore assumption a.1 i) concerns mainly scalar parameter a.

2.2 Asymptotic Normality

When T tends to infinity, the estimator $(\hat{\alpha}_T, \hat{q}_T = q(\hat{\alpha}_T))$ will tend to $(\alpha_0^*, 0)$. Let us assume :

Assumption a.2 :

i) The true set $\mathcal{A}_0$ has a non-empty interior and α_0^* is in the interior of the true set $\mathcal{A}_0$.

ii) 0 is in the interior of the set $Q(\alpha_0^*)$ of admissible values of q constructed from α_0^*.

iii) The log-likelihood function is twice continuously differentiable with respect to α.

iv) The additional objective function $\tilde{g}$ is continuously differentiable with respect to q and continuously cross differentiable with respect to q and α.

v) The function $q(.)$ exists and is continuously differentiable.

vi) The score $\dfrac{\partial \log l}{\partial \alpha}(y_t|y_{t-1};\alpha)$ has finite second-order moments.

vii) The Hessian $\dfrac{\partial^2 \log l}{\partial \alpha \partial \alpha'}(y_t|y_{t-1};\alpha)$ has finite first-order moments.

viii) The matrix $(Id - P)J_0$ has rank $\dim \alpha - \dim q - 1 - 2K$.

Let us discuss these additional assumptions. Condition $a.2$ i) eliminates the case $K = 1$, when the NMF is point-identified and the standard asymptotic theory applies. This is assumption (4*) in Shi, Shum (2015), Theorem 2.1. Then, the rank condition in their assumption (4) is automatically satisfied in our framework by assumptions A-2-A.3 and a.2 viii). Their condition (***) in Theorem 3.1 is automatically satisfied in our framework of maximum likelihood estimation.

On-line Appendix 3

Additional Estimation Results

3.1 The Limiting Identifying Criterion

i) The concentration measure

The concentration measure can be applied to the true parameter values A_0. This leads to the objective function:

$$g_0(q_{12}, q_{21}) = -\left\{\pi(q_{12}, q_{21}, A_0) \log \pi(q_{12}, q_{21}, A_0) + [(1 - \pi(q_{12}, q_{21}, A_0) \log(1 - \pi(q_{12}, q_{21}, A_0)]\right\},$$

to be maximized to get the least concentrated heterogeneity distributiion. It is easily checked that $\pi(q_{12}, q_{21}, A_0) = \frac{1}{12}\frac{(2+3q_{21})(3-q_{12})}{1-q_{12}q_{21}}$ is symmetric in q_{12}, q_{21}. We show in Figure 3 this function for the values of $-0.5 \leq q_{12} \leq 0.5, \ -0.5 \leq q_{21} \leq 0.5$.

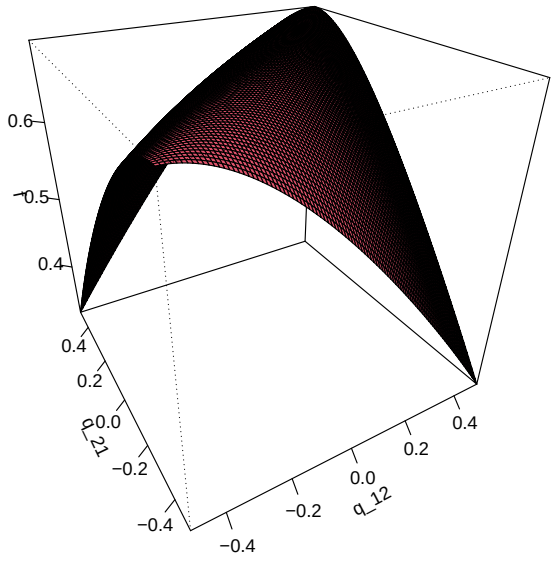

Figure 1: The Concentration Criterion

Since the true heterogeneity distribution $\pi = \{0.5, 0.5\}$ is the maximizer of the concentration function, the maximum is reached if and only if $\pi(q_{12}, q_{21}, A_0) = 0.5$, or equivalently if:

$$(2 + 3q_{21})(3 - q_{12}) = 6(1 - q_{12}q_{21}) \iff q_{12} = q_{21}.$$

As shown in Figure 3, the optimum is reached for the 45 degree line with respect to q_{12}, q_{21}. This means that the concentration measure as an additional optimization criterion allows us to reduce the dimension of the identified set to one, but it does not suffice to select a single point in the identified set. This is due to the uniform distribution of heterogeneity.

ii) The additional collinearity measure

Let us now consider the collinearity measure computed for $B^*(q)$. The pattern of this measure is illustrated in Figure 4. Then, for optimization, we add to the concentration criterion, a collinearity measure of the normalized B^*. This new maximization criterion becomes:

$$g(q_{12}, q_{21}) = g_0(q_{12}, q_{21}) + \lambda |det(B^{*'} B^*|,$$

where λ is a positive tuning parameter. Figure 5 shows the extended objective function for $\lambda = 5$. We see that the maximum will be reached on the boundary for $q_{1,2}, q_{2,1}$.

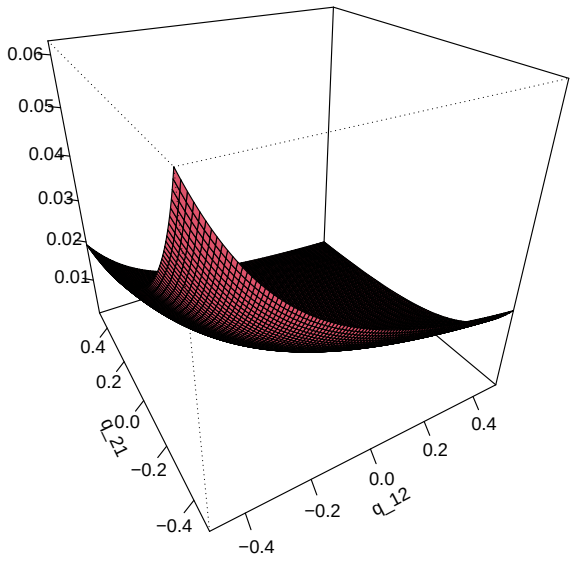

Figure 2: The Determinant Criterion

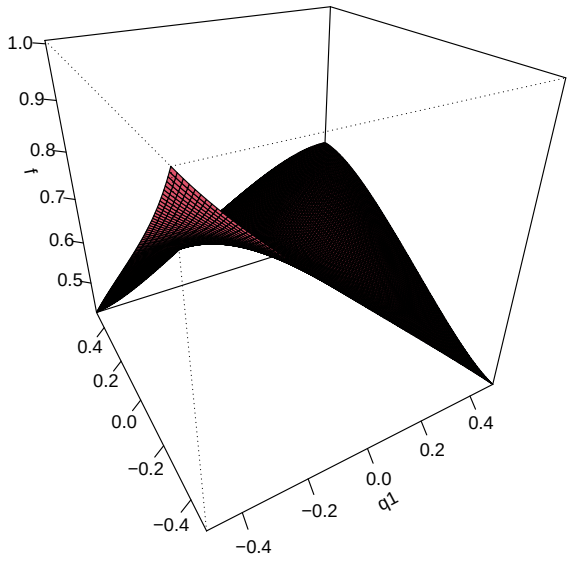

Figure 3: The Identifying Criterion with Determinant

iii) The repulsion criterion

In order to get the maximum in the interior of the identified set, we add the repulsion term $\sum_i \sum_j (ln(\beta^*_{i,j}) + ln(\gamma^*_{i,j}))$ with coefficient 1.7 in the combined identifying criterion in Figure 6:

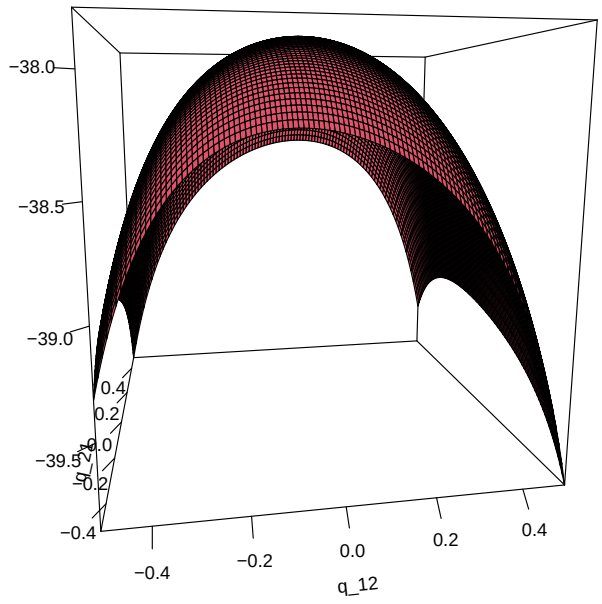

Figure 4: The Combined Identifying Criterion

3.2 Seemingly Unrelated Regressions

The multivariate Poisson model (6.l1) implies:

$$E(Y_t|Y_{t-1}) = AY_t + \mu$$

Therefore in the case of low dimension $n = 4$, the matrix A can be estimated by the (unconstrained) OLS. This approach is not efficient since it does not account for i) the reduced rank assumption, ii) the Poisson assumption and the conditional heteroscedasticity, in particular. However, it provides us information on the impact of the log-difference of the first two singular values on the accuracy of the estimation procedure. In this simulation, the relative accuracy will be measured by the weighted distance between the estimate $\hat{A}_T$ and the true value A_0, defined as:

$$dist = \left(\frac{1}{16} \sum_{i=1}^{4} \sum_{j=1}^{4} (\frac{\hat{a}_{i,j,T} - a_{0,i,j}}{a_{0,i,j}})^2 \right)^{1/2}.$$

We use the same simulated series as in Section 6.1 and show the estimated matrix $\hat{A}_T$ and distance measure for different numbers of observations $T = 200, 500, 1000$.

T=200:

$\widehat{dist} = 2.0748$ and

$$\hat{A}_{100} = \begin{pmatrix} 0.0597 & -0.0256 & 0.1077 & -0.0320 \\ 0.0917 & 0.0786 & 0.1589 & 0.0291 \\ 0.0715 & 0.0612 & 0.0111 & 0.1314 \\ 0.0850 & 0.1113 & -0.0643 & 0.0408 \end{pmatrix}$$

T=500:

$\widehat{dist} = 1.6912$ and

$$\hat{A}_{500} = \begin{pmatrix} 0.0076 & -0.0200 & 0.0252 & -0.0303 \\ 0.0224 & 0.0417 & 0.0413 & 0.0515 \\ 0.0964 & 0.0069 & -0.0130 & 0.1543 \\ 0.0634 & 0.0488 & -0.0380 & 0.0872 \end{pmatrix}$$

T=1000:

$\widehat{dist} = 1.2691$ and

$$\hat{A}_{1000} = \begin{pmatrix} -0.0076 & 0.0217 & 0.0172 & 0.0127 \\ 0.0478 & 0.0522 & 0.0225 & 0.0700 \\ 0.0753 & 0.0509 & 0.0806 & 0.1116 \\ 0.0326 & -0.0017 & -0.0258 & 0.0605 \end{pmatrix}$$

Although the unconstrained OLS approach is consistent and the estimated relative accuracy increases with the number of observations, the unconstrained OLS is not sufficiently accurate and should be avoided. In particular, it can produce negative values that are at odds with the assumption of a Poisson autoregressive model.

3.3 AML Estimates

To focus on the NMF estimation of A, we consider 1) the log-likelihood function (6.6) concentrated with respect to the intercept. This concentration is done by using 1) the OLS estimated intercept, and 2) the log-likelihood function in which the intercept is replaced by its true value. The AML algorithm provides the values of matrix $\hat{A}_T$ for T=200, 500 and T=1000 estimated from the two likelihood functions described above.

1) **Likelihood with estimated intercept**

The results are as follows:

T=200

$$\hat{B}_{200} = \begin{pmatrix} 0.1209 & 0.0203 \\ 0.2273 & 0.1571 \\ 0.0906 & 0.2927 \\ 0.0375 & 0.2648 \end{pmatrix}$$

$$\hat{C}_{200} = \begin{pmatrix} 0.2536 & 0.2117 \\ 0.1645 & 0.2459 \\ 0.2954 & 0.0465 \\ 0.1229 & 0.2362 \end{pmatrix}$$

estimated A

$$\hat{A}_{200} = \begin{pmatrix} 0.0350 & 0.0249 & 0.0367 & 0.0197 \\ 0.0909 & 0.0760 & 0.0745 & 0.0650 \\ 0.0850 & 0.0869 & 0.0404 & 0.0803 \\ 0.0656 & 0.0713 & 0.0234 & 0.0671 \end{pmatrix}$$

$\widehat{dist} = 1.4315$, $\hat{\pi} = 0.4227$.

T=500

$$\hat{B}_{500} = \begin{pmatrix} 0.0242 & 0.0001 \\ 0.0997 & 0.0981 \\ 0.0988 & 0.2886 \\ 0.0456 & 0.2372 \end{pmatrix}$$

$$\hat{C}_{500} = \begin{pmatrix} 0.2325 & 0.1914 \\ 0.1600 & 0.1579 \\ 0.1864 & 0.0014 \\ 0.2204 & 0.3326 \end{pmatrix}$$

estimated A

$$\hat{A}_{500} = \begin{pmatrix} 0.0057 & 0.0039 & 0.0045 & 0.0054 \\ 0.0420 & 0.0314 & 0.0187 & 0.0546 \\ 0.0782 & 0.0614 & 0.0188 & 0.1178 \\ 0.0560 & 0.0448 & 0.0088 & 0.0890 \end{pmatrix}$$

$\widehat{dist} = 1.2513$, $\hat{\pi} = 0.3346$.

T=1000

$$\hat{B}_{1000} = \begin{pmatrix} 0.0363 & 0.0475 \\ 0.1493 & 0.1974 \\ 0.3161 & 0.2351 \\ 0.1179 & 0.0398 \end{pmatrix}$$

$$\hat{C}_{1000} = \begin{pmatrix} 0.0353 & 0.2682 \\ 0.0506 & 0.1763 \\ 0.1468 & 0.0662 \\ 0.3510 & 0.0395 \end{pmatrix}$$

estimated A

$$\hat{A}_{1000} = \begin{pmatrix} 0.0140 & 0.0102 & 0.0085 & 0.0146 \\ 0.0582 & 0.0423 & 0.0350 & 0.0602 \\ 0.0742 & 0.0574 & 0.0620 & 0.1202 \\ 0.0148 & 0.0130 & 0.0199 & 0.0430 \end{pmatrix}$$

$\widehat{dist} = 1.1696$, $\hat{\pi} = 0.5584$.

2) **Likelihood with known intercept**

T=200

$$\hat{B}_{200} = \begin{pmatrix} 0.0480 & 0.1375 \\ 0.1332 & 0.2438 \\ 0.0022 & 0.3079 \\ 0.0735 & 0.3779 \end{pmatrix}$$

$$\hat{C}_{200} = \begin{pmatrix} 0.1032 & 0.1112 \\ 0.0751 & 0.1639 \\ 0.1268 & 0.0060 \\ 0.0289 & 0.1107 \end{pmatrix}$$

estimated A

$$\hat{A}_{200} = \begin{pmatrix} 0.0202 & 0.0261 & 0.0069 & 0.0166 \\ 0.0408 & 0.0500 & 0.0183 & 0.0308 \\ 0.0345 & 0.0506 & 0.0021 & 0.0341 \\ 0.0496 & 0.0675 & 0.0116 & 0.0439 \end{pmatrix}$$

$\widehat{dist} = 0.6750$, $\hat{\pi} = 0.1704$.

T=500

$$\hat{B}_{500} = \begin{pmatrix} 0.0192 & 0.1146 \\ 0.0319 & 0.0881 \\ 0.1328 & 0.0036 \\ 0.1156 & 0.0300 \end{pmatrix}$$

$$\hat{C}_{500} = \begin{pmatrix} 0.1645 & 0.4003 \\ 0.1783 & 0.3013 \\ 0.2152 & 0.2266 \\ 0.4697 & 0.1237 \end{pmatrix}$$

estimated A

$$\hat{A}_{500} = \begin{pmatrix} 0.0490 & 0.0379 & 0.0301 & 0.0232 \\ 0.0405 & 0.0322 & 0.0268 & 0.0259 \\ 0.0233 & 0.0248 & 0.0294 & 0.0628 \\ 0.0310 & 0.0297 & 0.0317 & 0.0580 \end{pmatrix}$$

$\widehat{dist} = 0.5136$, $\hat{\pi} = 0.5532$.

T=1000

$$\hat{B}_{1000} = \begin{pmatrix} 0.0504 & 0.0609 \\ 0.0654 & 0.0649 \\ 0.1380 & 0.0006 \\ 0.1234 & 0.0217 \end{pmatrix}$$

$$\hat{C}_{1000} = \begin{pmatrix} 0.1720 & 0.3674 \\ 0.1312 & 0.3290 \\ 0.1835 & 0.1713 \\ 0.4967 & 0.1718 \end{pmatrix}$$

estimated A

$$\hat{A}_{1000} = \begin{pmatrix} 0.0310 & 0.0266 & 0.0197 & 0.0355 \\ 0.0351 & 0.0299 & 0.0231 & 0.0436 \\ 0.0240 & 0.0183 & 0.0254 & 0.0687 \\ 0.0292 & 0.0233 & 0.0264 & 0.0650 \end{pmatrix}$$

$\widehat{dist} = 0.4119$, $\hat{\pi} = 0.7068$.

We observe:

a) A significant decrease of relative accuracy of AML estimators in comparison with the relative accuracy of OLS estimators. The accuracy also significantly decreases when the intercept is given.

b) The strict positivity of the estimates of B and C.

c) Sometimes a large fluctuation of the AML estimates $\hat{B}_T$ (e.g. compare the second column of $\hat{B}_T$ for T=200 and T=1000 and known intercept), whereas these fluctuations are smaller in $\hat{A}_T$. This is due to the identification issue of B and C, while A is identifiable and consistently estimated.

On-line Appendix 4

Identifying Maximum Likelihood and Singular Value Decomposition

Singular Value Decomposition (SVD) is a factorization technique for decomposing matrix A into a product of matrices $A = BC'$ with $Rk(A) = Rk(B) = Rk(C)$. It can be applied to any matrix A, not necessarily with positive elements. Matrices B and C do not necessarily have positive elements either. The SVD is obtained from a joint spectral decomposition of matrices AA' and $A'A$, or equivalently from a sequence of optimizations. For example, the first eigenvalue λ_1 and the first eigenvector v_1 are obtained from the optimization: $Max_v v'AA'v$, subject to $v'v = 1$. Then, the second eigenvalue λ_2 and the second eigenvector v_2 are obtained from the optimization $Max_v v'AA'v$, subject to $v'v_1 = 0$ and $v'v = 1$, and so on. The IML method can be implemented by applying the above sequence of SVD optimizations at each step of the AML algorithm. This allows us to derive the pointwise convergence and asymptotic normality of the IML/SVD estimator under the assumption of a given rank $Rk(A) = K$. The asymptotic variance-covariance matrix of this estimator can be derived along the lines of Sections 4.3.3-4.3.4, by taking into account the K consecutive SVD optimisations instead of a single optimization for the NMF. It is out of the scope of this paper to derive the complicated closed-form expression of this variance-covariance matrix. The results in Sections 4.3.3-4.3.4 show that this complicated expression is greatly simplified if matrix A_0 has non-negative elements and is such that $Rk(A) = Rk_+(A) = K$, since the sequence of optimizations leading to the

SVD that identifies a specific factorization can be replaced by a single optimization providing an alternative factorization.