Unified asymptotic theory for all partial directed coherence forms

L. A. Baccalá¹, C. S. N. de Brito², D. Y. Takahashi²,³,⁴,† and K. Sameshima⁴

¹Telecommunications and Control Department, Escola Politécnica,  
²Bioinformatics Graduate Program,  
³Institute of Mathematics and Statistics, and  
⁴Department of Radiology and Oncology, Faculdade de Medicina,  
University of São Paulo, São Paulo, Brazil

This paper presents a unified mathematical derivation of the asymptotic behaviour of the three main forms of partial directed coherence (PDC). Numerical examples are used to contrast PDC, gPDC (generalized PDC) and iPDC (information PDC) as to meaning and applicability and, more importantly, to show their essential statistical equivalence insofar as connectivity inference is concerned.

1. Introduction

Partial directed coherence (PDC), a multivariate time-series technique obtained from the factorization of partial coherence [1], has become popular for evaluating the connectivity between neural structures. In the frequency domain, it closely reflects the idea of Granger causality [2], i.e. a time series \(x(k)\) is Granger-caused by \(y(k)\) only if knowledge of \(y(k)\)’s past proves helpful in predicting \(x(k)\), and it is, as such, an important measure given that many neuroscience research scenarios, such as sleep staging [3], have long been linked to typical neuroelectric oscillatory behaviour.

PDC has been finding increasing applications [4–6], most of which have been carried out by comparing sample connectivities between groups classified as presenting some known disorder against normal controls. Only recently have objective trial-by-trial criteria appeared that allow the evaluation of PDC’s estimator asymptotic properties. This is the case with Takahashi et al. [7], who confirmed a previously available connectivity hypothesis test [8] by the addition of hitherto unavailable PDC confidence intervals.
Table 1. Defining variables according to PDC type in (2.3).

<table>
<thead>
<tr>
<th>variable</th>
<th>PDC</th>
<th>gPDC</th>
<th>iPDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>1</td>
<td>1/2</td>
<td>1/2</td>
</tr>
<tr>
<td>S</td>
<td>$I_K$</td>
<td>$(I_K \otimes \Sigma_w)^{-1}$</td>
<td>$\Sigma_w^{-1}$</td>
</tr>
</tbody>
</table>

This scenario has recently been compounded by the introduction of other more general expressions for PDC, which can be more suitable under certain special circumstances. Generalized PDC (gPDC) [9], for example, is better suited to cases of large prediction errors or power spectral disparity between the multivariate channels. Its asymptotics have been presented without proof in reference [10].

Another, perhaps more relevant, form of PDC, termed information PDC (iPDC) has been introduced in references [11,12] in connection with a precise interpretation of PDC in terms of the mutual information between adequately partialized processes, thereby establishing it as a measure of direct connectivity strength.

This paper extends reference [7] to explicitly obtain the asymptotic behaviour of all PDC forms in a unified way. It starts by briefly reviewing PDC’s various formulations (§2) and a statement of the asymptotic results (§3), which are numerically illustrated in §4. A final brief discussion appears in §5, leaving mathematical details to the appendix.

2. Background

Defining PDC and its variants calls for an adequately fitted multivariate autoregressive time-series (i.e. vector time-series) model of simultaneously acquired time series $x(n)$ made up by $x_k(n)$, $k = 1, \ldots, K$:

$$x(n) = \sum_{l=1}^{p} A(l)x(n-l) + w(n), \quad (2.1)$$

where $w(n)$ is a zero-mean white innovations process with covariance matrix $\Sigma_w = [\sigma_{ij}]$ and model order $p$. The $a_{ij}(l)$ coefficients from each $A(l)$ matrix describe the lagged effect of the $j$th onto the $i$th series, wherefrom one can also define a frequency domain representation of (2.1) via the $\tilde{A}(\lambda)$ matrix whose entries are given by

$$\tilde{A}_{ij}(\lambda) = \begin{cases} 1 - \sum_{l=1}^{p} a_{ij}(l) e^{-j2\pi \lambda l}, & \text{if } i = j, \\ -\sum_{l=1}^{p} a_{ij}(l) e^{-j2\pi \lambda l}, & \text{otherwise}, \end{cases} \quad (2.2)$$

where $j = \sqrt{-1}$.

Denote the columns of $\tilde{A}(\lambda)$ by $\tilde{a}_j$. Then, the general expression summarizing all the forms of PDC from $j$ to $i$ is

$$\pi_{ij}(\lambda) = \frac{1}{s} \frac{\tilde{A}_{ij}(\lambda)}{\sqrt{\tilde{a}_j^H(\lambda)\tilde{a}_j(\lambda)}}, \quad (2.3)$$

where $H$ denotes Hermitian transposition. Each PDC version uses its own $s$ and $S$ in (2.3) as summed up in Table 1.
(a) Problem formulation

As in reference [7] for the original PDC, one must reparametrize (2.3) before applying the delta method [13], which consists of an appropriate Taylor expansion of the distribution of (2.3) with respect to $a_{ij}(r)$ and $\Sigma_w$, which are estimated when fitting (2.1) and whose dispersion depends on $n_s$, the number of available observations.

To do so, use the vec matrix-column-stacking operator on the model coefficients

$$\alpha = \text{vec}[A(1)A(2)\cdots A(p)],$$

and on the residual noise covariance matrix $\sigma = \text{vec} \Sigma_w$ to produce the parameter vector

$$\theta^T = \begin{bmatrix} \alpha^T & \sigma^T \end{bmatrix}^T.$$

The delta method [13] rests on the continuous mapping nature of the parameters onto the statistics of interest (PDC) by using the asymptotic distribution parameter behaviour in terms of the number of data points $n_s$ and the Taylor expansion of the mapping.

Lemma 2.1. If vector time series $x(n)$ is stationary zero-mean Gaussian, then the full model parameter vector $\theta$ obeys

$$\sqrt{n_s}(\hat{\theta} - \theta) \to \mathcal{N}(0, \Omega_\theta),$$

where

$$\Omega_\theta = \begin{bmatrix} \Omega_\alpha & 0 \\ 0 & \Omega_\sigma \end{bmatrix},$$

with $\Omega_\alpha = \Gamma_x^{-1} \otimes \Sigma_w$, where $\Gamma_x = E[\bar{x}(n)\bar{x}^T(n)]$ for

$$\bar{x}(n) = [x_1(n) \cdots x_k(n) \cdots x_1(n-p+1) \cdots x_k(n-p+1)]^T$$

and

$$\Omega_\sigma = 2D_KD_K^+(\Sigma_w \otimes \Sigma_w)D_K^+D_K^T,$$

where $D_K^+$ is the Moore–Penrose pseudo-inverse of the standard duplication matrix.

To simplify notation, except where essential for comprehension, $i, j$ and $\lambda$ are omitted as they are fixed with respect to (2.1) estimation.

Lemma 2.1 holds even if $x(n)$ Gaussianity is relaxed [14]. In fact, finite auto- and cross-fourth-order moments of $w_i(n)$ suffice with the proviso that model (2.1) adequately describes the problem.

3. Result overview

Under lemma 2.1, confidence intervals and null-hypothesis thresholds for PDC (2.3) can be computed by splitting its parameter dependence on the parameter vector $\theta$ by considering its decomposition into numerator and denominator:

$$|\pi_{ij}(\lambda)|^2 = \nu(\theta) = \frac{\pi_n(\theta)}{\pi_d(\theta)},$$

The following results hold, thanks to the quadratic form dependence of $\pi_n(\theta)$ and $\pi_d(\theta)$ on $\theta$:

1. **Confidence intervals.** For large $n_s$, (3.1) is asymptotically normal, i.e.

$$\sqrt{n_s}(|\hat{\nu}_{ij}(\lambda)|^2 - |\nu_{ij}(\lambda)|^2) \to \mathcal{N}(0, \gamma^2(\lambda)),$$

where $\gamma^2(\lambda)$ is a frequency-dependent variance whose expression requires the notation introduced in the appendix.

2. **Null hypothesis threshold.** Under the null hypothesis,

$$H_0 : |\nu_{ij}(\lambda)|^2 = 0,$$

$\gamma^2(\lambda)$ vanishes identically, and (3.2) no longer holds, requiring the next Taylor expansion term [15] in the asymptotic expression of (3.1), which has an $O(n_s^{-1})$ dependence. The
Figure 1. (a) Original $|\text{PDC}|^2$ compared with (b) $|\iota\text{PDC}|^2$ for example 4.1 correctly detect connection direction. Black solid lines portray significant estimates ($\alpha = 1\%$), and grey lines non-significant estimates. Original PDC Melanoma $\rightarrow$ Sunspot is below threshold even though it is close to 1. Threshold values appear as black dashed lines. Error bars Sunspot $\rightarrow$ Melanoma $|\text{PDC}|^2$ depict estimated 99% confidence level interval. Note the different squared PDC magnitudes in each panel. The grey background calls attention to squared PDC magnitudes below 0.01. Conventions also apply to other figures.

The resulting distribution is that of a linear combination of at most two properly weighted $\chi^2$ with frequency-dependent weights:

$$n_s \tilde{a}_j^H(\lambda) S \tilde{a}_j(\lambda) (|\tilde{\pi}_{ij}(\lambda)|^2 - |\pi_{ij}(\lambda)|^2) \quad \overset{d}{\rightarrow} \sum_{k=1}^{q} l_k(\lambda) \chi^2_1,$$

where $l_k(\lambda)$ (details left to the appendix) depend only on PDC’s numerator $\pi_{ij}(\theta(\lambda))$. PDC denominator dependence is implicit on the left-hand side of (3.4).

The $\chi^2$ dependence comes from the quadratic behaviour of the $\theta$ parameters, which are themselves asymptotically normal according to lemma 2.1.

4. Numerical illustrations

(a) Example 4.1: Sunspot $\times$ melanoma incidence data

Computing $|\text{PDC}|^2$ for the relationship between sunspot activity and Connecticut melanoma incidence [16] over only 37 years leads to the upper panels of figure 1, whose striking feature is that the Sunspot $\rightarrow$ Melanoma PDC relationship counterintuitively turns out to be small as opposed to the Melanoma $\rightarrow$ Sunspot one. Despite its small size, closer examination reveals that Sunspot $\rightarrow$ Melanoma $|\text{PDC}|^2$ is above threshold around the characteristic sunspot periodicity and that Melanoma $\rightarrow$ Sunspot $|\text{PDC}|^2$, though numerically high, is below the test threshold and is, hence, not significant.

Apparent interaction size discrepancies such as this were the primary reason for introducing $g\text{PDC}$ [9], and later $\iota\text{PDC}$ [11,12], shown in figure 1b, where the relative estimated sizes are more in accord with intuition and retain the correct directed inferential significance relationship. The $g\text{PDC}$ results are similar to those for $\iota\text{PDC}$ because $\Sigma_w$ is diagonal, and may be appreciated in reference [10].
The difference between PDC and gPDC/iPDC stems from the different units/magnitudes of the observed quantities (sunspot number versus melanoma cases per 100,000 inhabitants). The scale invariance of gPDC/iPDC solves this and dispenses with the common often inexplicit practice of data standardization, i.e. division of each original time series by its standard deviation (see also Baccalà et al. [9]).

Despite possible misleading impressions caused by the computed magnitude of the original PDC, it is important to keep in mind that connectivity inference using it is, nonetheless, correct under the appropriate statistical threshold.

Now consider the model
\[
\begin{align*}
x_1(n) &= 0.95\sqrt{2}x_1(n-1) - 0.9025x_1(n-2) + w_1(n), \\
x_2(n) &= 0.5x_1(n-2) + w_2(n), \\
x_3(n) &= -0.4x_1(n-3) + w_3(n), \\
x_4(n) &= -0.5x_1(n-2) + 0.25\sqrt{2}x_4(n-1) + 0.25\sqrt{2}x_5(n-1) + w_4(n) \\
x_5(n) &= -0.25\sqrt{2}x_4(n-1) + 0.25\sqrt{2}x_5(n-1) + w_5(n)
\end{align*}
\]
from reference [1] for the next two examples. In all cases, \(n_s = 2000\) observations were used after a burn-in period of 1000 to ensure stationarity. The model order \(p = 3\) was obtained via Akaike’s information criterion.

5. Discussion

By establishing a common PDC asymptotics framework, this paper epitomizes the effort begun in reference [7] for the original PDC and leads to rigorous gPDC derivation details for illustrative simulations first shown in reference [10]. The present results also apply to the iPDC case [11,12].

Outside this framework is Schelter et al.’s PDC-related work [18], the statistical behaviour of which lies outside of the present scope.
Figure 2. Illustration of $|\mu\text{PDC}|^2$ results for $a_1 = 0.59, a_2 = 0.52, a_3 = 0.72, a_4 = 0.98$ and $a_5 = 0.66$ for example 4.2 leading to correctly detected connections from $j$ (columns) to $i$ (rows). Time domain Granger causality Wald tests $p$-values from reference [14] are also shown.

Figure 3. $|\mu\text{PDC}|^2$ at $\lambda = 0.2$, example 4.2, quantile Monte Carlo performance results (grey points, 2000 realizations using $a_i$ from figure 2) for (a) the non-existing connection $x_3(n) \rightarrow x_1(n)$ and (b) the normally distributed existing link $x_4(n) \rightarrow x_5(n)$. In panel (a) deviations from the solid theoretical line are consistent with $\alpha = 0.01$.

The derivation’s unified nature implies that the PDC forms addressed herein have similar statistical behaviour and that, despite PDC magnitude issues as for the original PDC, they all correctly detect directional connections provided proper null hypothesis thresholds are used in each PDC case, and, most importantly, as long as the fitted models are good, i.e. the observed time-series durations are compatible with model orders, and the fitted residuals are white. Naturally,
Figure 4. Original $|\text{PDC}|^2$ results for example 4.3 portraying successful connectivity inference are further confirmed by $p$-values for the time domain Granger causality Wald tests from reference [14]. $|\text{PDC}|^2$ 99% confidence intervals for the significant connections are also shown.

Figure 5. $|g\text{PDC}|^2$ results for example 4.3 portraying successful connectivity inference. Note a confidence interval pattern similar to that of the squared PDC results (figure 4).
Figure 6. \(|\iota PDC|^2\) results for example 4.3 portraying successful connectivity inference and confidence intervals for significant \(|\iota PDC|^2\) estimates. Grey backgrounds call attention to scales below 0.01.

because of their scale invariance, \(g PDC\) and \(i PDC\) should be preferred over PDC, specially now that they, too, have objective connectivity detection criteria.

The examples in this paper were chosen to illustrate some specific points. The sunspot/melanoma example had only very few observations (merely \(n_s = 37\) data points), and results were consistent even though sunspot behaviour hardly conforms to signal Gaussianity as rigorously required or to the desirable \(n_s\) large limit. This is in line with the more general validity of lemma 2.1 in a more general setting.

Example 4.2 was the most canonical one in that all theoretical premises were imposed by construction, thus permitting full appreciation of distribution quantiles. \(i PDC\) was stressed for its objective ability to address interaction size effects in the presence of instantaneous causality.

Example 4.3 involved imposing coloured inputs whose presence can introduce biases into the estimated models. Yet, even in this case, using model parameters borrowed from other published results led to correct connectivity inference.

Because interval/threshold computations are intricate, their derivations were left to the appendix even though they constitute an important part of the present paper. Further illustrations of the examples are provided as electronic supplementary material that furthermore contains reference to the routines from the open-source AsympPDC package (available from http://www.lcs.poli.usp.br/~baccala/philosophical) used in producing the examples.

Acknowledgements. Support from FAPESP (grant no. 2005/56464-9; CInAPCe Programme), CAPES Bioinformatics Graduate Programme (D.Y.T. and C.S.N.B.), FAPESP (grant no. 2008/08171-0; D.Y.T), CNPq (grant nos 306964/2006-6 and 304404/2009-8; L.A.B.) are gratefully acknowledged.

Appendix A

In the following, function arguments are shown explicitly only when essential for comprehension and must otherwise be deduced from context.
Stating PDC in a general form requires preliminary definitions. To include $\lambda$ dependence, and avoid using complex numbers for PDC definition, consider the change of variables from $\alpha$ to $a$,

$$a(\lambda) = \begin{bmatrix} \text{vec}(\text{Re}(\bar{A}(\lambda))) \\ \text{vec}(\text{Im}(\bar{A}(\lambda))) \end{bmatrix} = \begin{bmatrix} \text{vec}(I_{pK^2}) \\ 0 \end{bmatrix} - C(\lambda)\alpha,$$

(A 1)

where $I_N$ is an $N \times N$ identity matrix and

$$C(\lambda) = \begin{bmatrix} C(\lambda) \\ -S(\lambda) \end{bmatrix},$$

whose blocks are $K^2 \times pK^2$ dimensional of the form

$$C(\lambda) = [C_1(\lambda) \cdots C_p(\lambda)], \quad S(\lambda) = [S_1(\lambda) \cdots S_p(\lambda)],$$

for

$$C_r(\lambda) = \text{diag}([\cos(2\pi r\lambda) \cdots \cos(2\pi r\lambda)]), \quad S_r(\lambda) = \text{diag}([\sin(2\pi r\lambda) \cdots \sin(2\pi r\lambda)]).$$

This reparametrizes the problem using $\bar{\theta}(\lambda)^T = [a^T(\lambda) \sigma^T]^T$, i.e.

$$\pi(\bar{\theta}) = \frac{\pi_n(\bar{\theta})}{\pi_d(\bar{\theta})}$$

(A 2)

with implicit $i$, $j$ and $\lambda$ dependence.

**Lemma A.1 ($\alpha$-asymptotics).** The asymptotic properties of the $\bar{\theta}$ are given by

$$\sqrt{n}\frac{\hat{\bar{\theta}} - \bar{\theta}}{\Omega_{\bar{\theta}}} \to \mathcal{N}(0, \Omega_{\bar{\theta}}),$$

(A 3)

where

$$\Omega_{\bar{\theta}} = \begin{bmatrix} \Omega_{\alpha} & 0 \\ 0 & \Omega_{\sigma} \end{bmatrix}$$

(A 4)

with

$$\Omega_{\alpha}(\lambda) = C(\lambda)\Omega C^T(\lambda)$$

(A 5)

for

$$\Omega = \begin{bmatrix} \Omega_{\alpha} & \Omega_{\alpha} \\ \Omega_{\alpha} & \Omega_{\sigma} \end{bmatrix}.$$

The proof is straightforward by direct computation using (A 1).

To usefully write (2.3) as the (A 2) ratio, consider the $K^2 \times K^2$ matrices:

- $I_{ij}$, whose entries are zero except for indices of the form $(l, m) = ((j-1)K + i, (j-1)K + i)$, which equal 1; and
- $I_j$, which is non-zero only for entries whose indices are of the form $(l, m)$ with $(j-1)K + 1 \leq l = m \leq jK$.

These matrices can be used to construct

$$\Gamma_{ij} = \begin{bmatrix} I_{ij} & 0 \\ 0 & I_{ij} \end{bmatrix}, \quad \Gamma_j = \begin{bmatrix} I_j & 0 \\ 0 & I_j \end{bmatrix},$$

which are parameter choice operators for the $j$ to $i$ pair of interest.

**Lemma A.2.** The absolute value of squared (2.3) can be written as a ratio of quadratic forms:

$$|\pi_{ij}(\lambda)|^2 = \frac{a(\lambda)^T \Gamma_{ij} \mathcal{S}_n(\sigma) \Gamma_{ij} a(\lambda)}{a(\lambda)^T \Gamma_j \mathcal{S}_d(\sigma) \Gamma_j a(\lambda)},$$

(A 6)

where $\mathcal{S}_n(\sigma)$ and $\mathcal{S}_d(\sigma)$ do not depend on $\lambda$ and take on different values according to the desired PDC form being listed in table 2.
Table 2. Defining variables according to PDC type in (A.6).

<table>
<thead>
<tr>
<th>variable</th>
<th>PDC</th>
<th>gPDC</th>
<th>jPDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_n$</td>
<td>$I_2 \otimes I_K$</td>
<td>$I_2 \otimes (I_K \otimes \Sigma_w)^{-1}$</td>
<td>$I_2 \otimes (I_K \otimes \Sigma_w)^{-1}$</td>
</tr>
<tr>
<td>$S_d$</td>
<td>$I_2 \otimes I_K$</td>
<td>$I_2 \otimes (I_K \otimes \Sigma_w)^{-1}$</td>
<td>$I_2 \otimes \Sigma_w^{-1}$</td>
</tr>
</tbody>
</table>

The proof follows by direct substitution.
Thus,
$$
\pi_n(\theta) = a(\lambda)^T I_j^a S_n(\sigma) I_j^a a(\lambda), \quad \pi_d(\theta) = a(\lambda)^T I_j^a S_d(\sigma) I_j^a a(\lambda)
$$
(A.7)

have naturally split factors in their dependence on $a$ and $\sigma$ (table 1).

**Theorem A.3 (delta method).** Let the distribution of $v_n = (v_1, \ldots, v_k)^T$ estimated from $n$ observations converge in distribution as
$$
\sqrt{n} (v_n - \mu) \xrightarrow{d} N(0, \Sigma_v).
$$
(A.8)

Let $g(v)$ be a real-valued function with continuous partials of order $m > 1$ in the neighbourhood of $v = \mu$, with all the partials of order $j$ with $1 \leq j \leq m - 1$ vanishing at $v = \mu$ and non-vanishing $m$th-order partials at $v = \mu$. Then
$$
(\sqrt{n})^m (\hat{g}(v_n) - g(\mu)) \xrightarrow{d} \frac{1}{m!} \sum_{i_1=1}^k \cdots \sum_{i_m=1}^k \frac{\partial^m g}{\partial x_{i_1} \cdots \partial x_{i_m}} \bigg|_{x=\mu} \prod_{j=1}^m Z_{ij},
$$
(A.9)

with
$$
Z = (Z_1, \ldots, Z_k)^T \sim N(0, \Sigma_v).
$$
(A.10)

From this one can readily deduce the following for large $n$ and non-null first derivatives in (A.9).

**Corollary A.4.** For a real differentiable function $g(v)$ asymptotically distributed as in (A.8), then
$$
\sqrt{n} (\hat{g}(v_n) - g(\mu)) \xrightarrow{d} N(0, g^T \Sigma_v g)
$$
(A.11)
is the first delta method approximation, where $g = \nabla_v g$ is the gradient of $g(v)$ computed at $\mu$ [15].

### A.1. Confidence interval theorem

**Proposition A.5.** Omitting $\lambda$ and $\sigma$ for the sake of simplicity, one has
$$
\sqrt{n_o}(|\hat{\pi}_{ij}|^2 - |\pi_{ij}|^2) \sim N(0, \gamma^2),
$$
(A.12)

where $n_o$ is the number of observations and
$$
\gamma^2 = g_a \Omega_a g_a^T + g_\sigma \Omega_\sigma g_\sigma^T,
$$
(A.13)

for
$$
g_a = 2 \frac{a^T I_j^a S_n I_j^a}{a^T I_j^a S_d I_j^a} - 2 \frac{a^T I_j^a S_n I_j^a}{(a^T I_j^a S_d I_j^a)^2} a^T I_j^a S_d I_j^a
$$
(A.14)

and
$$
g_\sigma = \frac{1}{a^T I_j^a S_d I_j^a} \{ (I_j^a)^T \otimes (a^T I_j^a) \} \Theta K \xi_d - \frac{a^T I_j^a S_n I_j^a}{(a^T I_j^a S_d I_j^a)^2} \{ (I_j^a)^T \otimes (a^T I_j^a) \} \Theta K \xi_d,
$$
(A.15)

where the values of $\xi_n$ and $\xi_d$ are listed in table 3 and where $\Theta_K = (T_{2K,K} \otimes I_{2K})(I_K \otimes \text{vec}(I_{2K} \otimes I_K))$ and $T_{LM}$ stands for the commutation matrix [14].

When $\Sigma_w$ is known a priori or does not need to be estimated, $g_\sigma \Omega_\sigma g_\sigma^T$ in (A.15) is zero.
Table 3. Defining variables according to PDC type in (A15).

<table>
<thead>
<tr>
<th>variable</th>
<th>PDC</th>
<th>gPDC</th>
<th>iPDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_n$</td>
<td>0</td>
<td>$-\text{diag} (\text{vec}(S_n^{-1})))$</td>
<td>$-\text{diag} (\text{vec}(S_n^{-1})))$</td>
</tr>
<tr>
<td>$\xi_d$</td>
<td>0</td>
<td>$-\text{diag} (\text{vec}(S_n^{-1})))$</td>
<td>$-\Sigma_w^{-1} \otimes \Sigma_w^{-1}$</td>
</tr>
</tbody>
</table>

Proof. The proof follows from theorem A.4 under lemma A.1, by computing the gradient of $\pi(\tilde{\theta})$. This is simplified as $a$ and $\sigma$ are asymptotically independent (hence $\Omega_{\tilde{\theta}}$ is block diagonal), allowing their separate consideration.

The transpose of the required gradients can be written as

$$\frac{\partial \pi}{\partial \psi^T} = \frac{1}{\pi_d} \left[ \pi_d \frac{\partial \pi_n}{\partial \psi^T} - \pi_n \frac{\partial \pi_d}{\partial \psi^T} \right],$$

(A16)

where $\psi$ represents either $a$ or $\sigma$ and respectively leads to $g_a$ and $g_\sigma$.

These gradients operate on the quadratic forms whose differentiation with respect to $a$ yields

$$\frac{\partial (a^T \Gamma^c S^c a)}{\partial a^T} = 2a^T \Gamma^c S^c$$

because all appropriately valued $\Gamma^c S^c$ matrices are symmetric. Inserting the terms in (A16) leads to (A14).

The nested dependence of $\pi(\tilde{\theta})$ on $\sigma$ calls for the use of the chain rule (see reference [19], p. 183 and equation (3)):

$$\frac{\partial (a^T \Gamma^c S^c a)}{\partial \sigma^T} = \left( \frac{\partial (a^T \Gamma^c S^c a)}{\partial S} \right) \frac{\partial S}{\partial \sigma^T},$$

(A18)

where

$$\frac{\partial (a^T \Gamma^c S^c a)}{\partial S} = (\Gamma^c a)^T \otimes (\Gamma^c a^c).$$

(A19)

In its general form, $S = I_{2K} \otimes \tilde{S}$. Thus, by the chain rule (see reference [19], p. 184 and equation (11)(c)).

$$\frac{\partial S}{\partial \sigma^T} = \frac{\partial \tilde{S}}{\partial \sigma^T},$$

(A20)

where

$$\frac{\partial \tilde{S}}{\partial S} = \Theta_K = (T_{2K,K} \otimes I_{2K^2})(I_K \otimes \text{vec}(I_{2K} \otimes I_K)).$$

(A21)

All that is left to compute are the derivatives of $\tilde{S}$ in each case (table 1). For each possible $\tilde{S}$ value, one has

$$\frac{\partial \tilde{S}_K}{\partial \sigma^T} = 0,$$

(A22)

which follows from lack of $\sigma$ dependence.

When $\tilde{S} = \Sigma_w^{-1}$ (see reference [20], p. 183 and equation (19)) then

$$\frac{\partial \Sigma_w^{-1}}{\partial \sigma^T} = -\Sigma_w^{-1} \otimes \Sigma_w^{-1},$$

(A23)

and finally let $\tilde{S} = I_K \otimes \Sigma_w$ so that $\tilde{S} = \Sigma_w^{-1}$ implies

$$\frac{\partial (\Sigma^{-1})}{\partial \sigma^T} = \frac{\partial (\Sigma^{-1})}{\partial S} \frac{\partial S}{\partial \Sigma_w} \frac{\partial \Sigma_w}{\partial \sigma^T},$$

(A24)
which requires the chain rule twice where (see reference [19], p. 185 and equation (16))

\[
\frac{\partial (S^{-1})}{\partial S} = -S^{-T} \otimes S^{-1} = -(I_K \otimes \Sigma_w)^{-T} \otimes (I_K \otimes \Sigma_w)^{-1}
\]

(A 25)

and

\[
\frac{\partial \tilde{S}}{\partial \Sigma_w} = \text{diag}(\text{vec}(I_K)),
\]

(A 26)

and trivially

\[
\frac{\partial \Sigma_w}{\partial \sigma} = I_K^2,
\]

(A 27)

which together in (A 24) reduce to

\[
\frac{\partial \tilde{S}}{\partial \sigma} = -\text{diag}(\text{vec}( (I_K \otimes \Sigma_w)^{-2} )) = -\text{diag}( \Sigma_n^{-2} )
\]

(A 28)

after some additional algebra. The latter results are summarized as \( \xi_n \) and \( \xi_d \) quantities in table 3 and comprise (A 15).

Use of Slutsky’s lemma concludes the proof by allowing the use of estimated quantities. ■

### A.2. Null hypothesis test

Under the null hypothesis

\[
H_0 : |\pi_{ij}|^2 = 0 \iff \Gamma_{ij}^c a = 0,
\]

(A 29)

both (A 14) and (A 15) equal zero, and (3.2) no longer applies, so that the next Taylor term becomes necessary [15] weighted by one-half of PDC’s Hessian at the point of interest with an \( O(n_\pi^{-1}) \) dependence. Via a device similar to that used in reference [7], one can show the following.

**Proposition A.6.** Under (A 29)

\[
n_s (a_i^c \Sigma_d a_i^c) (|\tilde{\pi}_{ij}|^2 - |\pi_{ij}|^2) \overset{d}{\to} \sum_{k=1}^{q} l_k \chi^2_1,
\]

(A 30)

(see (3.2)) where \( l_k \) are the eigenvalues of \( D = L^T \Gamma_{ij}^c S_n \Gamma_{ij}^c L \), where \( L \) is the Cholesky factor of \( \tilde{\Sigma}_a \) and \( q = \text{rank}(D) \leq 2 \) and reduces to 1 whenever \( \lambda \in \{0, \pm 0.5\} \) or \( p = 1 \) for all \( \lambda \in [-0.5, 0.5] \).

Equation (A 30) is a linear combination of \( \chi^2_1 \) variables whose weights depend on estimated parameter and covariance values appearing in \( \tilde{\Sigma}_a \), which depend on \( \lambda \) (A 5).

**Proof.** Using theorem A.3 with \( m = 2 \) as both (A 14) and (A 15) vanish requires (A 14) and (A 15) derivation with respect to \( \sigma \) a second time. This does not alter the \( \Gamma_{ij}^c a \) dependence and so also produces null results. The same holds when deriving (A 15) with respect to \( a \) because it is quadratic in \( \Gamma_{ij}^c a \). The only non-zero surviving term comes from deriving (A 14) with respect to \( a \) and reduces to

\[
\frac{2}{a_i^c S_d a_i^c} \Gamma_{ij}^c S_n \Gamma_{ij}^c.
\]

(A 31)

Therefore, the Hessian in (A 9) only has an upper non-zero block corresponding to (A 31) implying that only the distribution of \( a \) needs to be considered in writing

\[
n_s (\tilde{a}^T \Gamma_{ij}^c S_d \tilde{a}) (|\tilde{\pi}_{ij}|^2 - |\pi_{ij}|^2) \overset{d}{\to} x^T \Gamma_{ij}^c S_n \Gamma_{ij}^c x,
\]

(A 32)

under theorem A.3 for \( x \overset{d}{\to} \mathcal{N}(0, \Sigma_a) \). Slutsky’s lemma concludes the first part of the proof by allowing the use of estimated quantities.

Diagonalizing \( x^T \Gamma_{ij}^c S_n \Gamma_{ij}^c x \) is done via the Cholesky factor \( L_a \) from \( \Sigma_a = L_a L_a^T \). Making \( x = Ly \), where \( L = S_n^{1/2} \Gamma_{ij}^c L_a \), yields \( x^T \Gamma_{ij}^c S_n \Gamma_{ij}^c x = y^T L_a^T \Gamma_{ij}^c S_n \Gamma_{ij}^c L_a y = y^T D y \), which means the vector elements
\( y = (L^T L)^{-1} L^T x \) become mutually independent zero-mean and unit-variance so that diagonalizing \( D = U A U^T \) with \( U U^T = I_{q \times q} \) produces

\[
y^T D y = \sum_{k=1}^{q} l_k y^T u_k u_k^T y = \sum_{k=1}^{q} l_k \xi_k^2,
\]

where \( u_k \) is the \( k \)th column of \( U \). It is easy to show that the variables \( \xi_k = u_k^T y \) are mutually independent, unit-variance zero-mean normal, implying \( \xi_k^2 \) are \( \chi^2 \) random variables.

As \( \text{rank}(X) = \text{rank}(X^T) \) and \( \text{rank}(XY) \leq \min(\text{rank}(X), \text{rank}(Y)) \), after recalling explicit \( \lambda \) dependence, this implies that

\[
\text{rank}(D) = \text{rank}(L_a^T I_{ij}^c S_n I_{ij}^c) = \text{rank}(L_a L_a^T I_{ij}^c S_n I_{ij}^c)
\]

\[
= \text{rank}(\Omega(\lambda) I_{ij}^c S_n I_{ij}^c) = \text{rank}(C(\lambda) \Omega(\lambda) I_{ij}^c S_n I_{ij}^c)
\]

is upper bounded by \( \text{rank}(I_{ij}^c) = 2 \) and is given by \( \text{rank}(C(\lambda)) = 1 \) when \( \lambda \in (0, \pm 0.5) \) or by \( \text{rank}(C(\lambda) I_{ij}^c) = 1 \) when \( p = 1 \) for all \( \lambda \), thereby concluding the proof.

Details on obtaining the threshold values from the linear combination followed those adopted in Takahashi et al. [7] and references therein.

**References**


