Spurious Spatial Regression, Spatial Cointegration and Heteroscedasticity

by

Jørgen Lauridsen

and

Reinhold Kosfeld

Discussion Papers on Business and Economics
No. 6/2006

FURTHER INFORMATION
Department of Business and Economics
Faculty of Social Sciences
University of Southern Denmark
Campusvej 55
DK-5230 Odense M
Denmark

Tel.: +45 6550 3271
Fax: +45 6615 8790
E-mail: lho@sam.sdu.dk
http://www.sam.sdu.dk/depts/virkl/about.shtml

ISBN 87-91657-05-9

Spurious Spatial Regression, Spatial Cointegration and Heteroscedasticity

Jørgen Lauridsen[1]

Reinhold Kosfeld[2]

[1] Corresponding author: Associate Professor, Department of Business and Economics, Campusvej 55, DK-5230 Odense M, Denmark

E-mail jtl@sam.sdu.dk

Homepage http://www.sam.sdu.dk/ansat/jtl

[2] Professor, Department of Economics, University of Kassel, D-34109 Kassel, Germany

E-mail kosfeld@wirtschaft.uni-kassel.de

Homepage http://www.wirtschaft.uni-kassel.de/Kosfeld/

Spurious Spatial Regression, Spatial Cointegration and Heteroscedasticity

ABSTRACT

A test strategy consisting of a two-step application of a Lagrange Multiplier test was recently suggested as a device to reveal spatial nonstationarity, spurious spatial regression and spatial cointegration. The present paper generalises the test procedure by incorporating control for biased test values emerging from unobserved heteroscedasticity. Using Monte Carlo simulation, the behaviour of several relevant tests for nonstationarity and/or heteroscedasticity are investigated. The two-step test for spatial nonstationarity turns out to be robust towards heteroscedasticity. While several tests for heteroscedasticity prove to be inconclusive under certain circumstances, a Lagrange Multiplier test for heteroscedasticity based on spatially differenced variables serves well as an indication of heteroscedasticity irrespective of stationarity status.

JEL Classifications: C21; C40; C51; J60.

Keywords: Spatial autocorrelation; spatial autoregression; spatial nonstationarity, spatial cointegration, spurious regression, unobserved heteroscedasticity.

1. INTRODUCTION

Spatial regression has been discussed widely in books dedicated to developments in spatial econometrics, notably by Anselin (1988a), and Anselin and Florax (1995). The consequenses for estimation and inference in the presence of stable spatial processes have been extensively investigated (Haining 1990; Anselin 1988a; Bivand 1980; Richardson 1990; Richardson and Hèmon 1981; Clifford and Richardson 1985; Clifford, Richardson and Hèmon 1989). The study of Fingleton (1999) took the first steps into analyses of implications of spatial unit roots, spatial cointegration and spatial error correction models. A follow-up to this study is found in Mur and Trívez (2003), where the concept of spurious spatial regression was established in a framework of spatial trend (non)stationarity. In Lauridsen (2006) estimation of spatial error-correction models using an IV approach was investigated, while Lauridsen and Kosfeld (2006) suggested an LM test procedure to test for spatial nonstationarity, spurious spatial regression and spatial cointegration. A study of Kosfeld and Lauridsen (2004) applied this methodology to a model for regional convergence across German labour market regions. Lauridsen and Kosfeld (2004) established a Wald post-test for spatial nonstationarity.

Fingleton (1999) suggested that “very high” values of Moran’s I test for spatial residual autocorrelation indicate spatial nonstationarity and spurious regression. It was, however, left as an open question how to distinguish between stationary positive autocorrelation and nonstationarity. Lauridsen and Kosfeld (2006) showed that a two-step LM error test can provide a better founded basis to separate these two cases. It was further shown that the same procedure works as a diagnostic for spurious regression. Next, it was suggested that the test procedure works well as a test for spatial cointegration, using a specific two-variable data generating process. In all cases, the small-sample properties of the suggested procedures were derived using Monte Carlo simulation. It was concluded that the procedure works well in all cases, especially for medium and large sample sizes.

The purpose of the present paper is to extend the two-step LM test procedure established by Lauridsen and Kosfeld (2006) to account for unobserved heteroscedasticity. Specifically, this will be obtained by incorporating the suggested modifications in Anselin (1988b) into the two-step LM test procedure suggested by Lauridsen and Kosfeld (2006). It is concluded that the unadjusted tests for spatial nonstationarity are not biased by presence of heteroscedasticity. While several tests for heteroscedasticity are partly flawed by nonstationarity, it is found that an LM test for heteroscedasticity based on spatially differenced variables is robust toward spatial nonstationarity or spurious regression. Thus, a modified procedure applies well, so that inference can be based on three LM tests.

2. MODELS WITH SPATIAL DYNAMICS

2.1. The regressive, spatially autoregressive model.

The first order spatially autoregressive model (SAR(1) model) was initially studied by Whittle (1954) and has been used extensively in works by Ord (1975), Cliff and Ord (1981), Ripley (1981), Upton and Fingleton (1985), Anselin (1988a), Haining (1990), Griffith (1992), Anselin et al. (1996), Florax et al. (2003), Lauridsen (2006). For applied research the SAR(1) model is extended by explanatory variables (see Upton and Fingleton, 1985; Anselin, 1988a; Haining, 1990; Anselin et al. (1996); Florax et al. (2003); Lauridsen, 2006). The regressive, spatially autoregressive model (SARX(1) model) is established as

(1) $\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{v}$,

in which $\mathbf{y}$ is an $n\times 1$ vector, $\mathbf{X}$ an $n\times K$ matrix of explanatory variables, ρ the autoregressive parameter, $\mathbf{I}$ the $n\times n$ identity matrix and $\mathbf{v}$ an $n\times 1$ vector of independently normally distributed errors with zero expectation and variances σ^2, i.e. $\mathbf{v} \sim N(\mathbf{0},\sigma^2\mathbf{I})$, $\mathbf{W}$ denotes an $n\times n$ spatial weight matrix. It is obtained by row-standardisation of the $n\times n$ contiguity matrix $\mathbf{W^*}$ which is defined by $W^*_{ij} = 1$ if observation j is assumed to impact observation i, and $W^*_{ij} = 0$ otherwise, i.e. $W_{ij} = W^*_{ij} / \Sigma_{j=1..n} W^*_{ij}$. For alternative specifications of the spatial weight matrix, see e.g. Cliff and Ord (1981) and Anselin (1988a). $\mathbf{W}$ may be multidirectional, which is not the case for the time-series case where $W_{ij} = 1$ if $j = i$-1, for $i = 2,3,..,n$. For the general spatial case, $\mathbf{W}$ is generally multidirectional. As proved by Anselin (1988a), multidirectionality of $\mathbf{W}$ renders OLS estimation of the parameters inconsistent. Finally, for the general case, ρ is restricted to the interval between -1 and +1 and thus may assume positive as well as negative values. Although meriting interest in itself, the negative case is conceptually different from the usual positive case. We thus narrow our focus in the present investigation to the common case where ρ is positive.

2.2. Spurious regression and spatial nonstationarity.

If $\mathbf{y}$ and one or more of the $\mathbf{x}$ variables are generated according to SAR schemes with positive autoregressive parameters and $\mathbf{y}$ is regressed on $\mathbf{X}$, i.e.

(2) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$,

with $\boldsymbol{\varepsilon}$ as the error term, a risk of spurious regression occurs. Especially, in the case of spatial nonstationarity, where $\mathbf{y}$ and one or more of the $\mathbf{x}$ variables have autoregressive parameters close to 1, the risk of spurious regression is alarmingly high. It manifests in the OLS residuals $\mathbf{e}$ of the regression tending to be highly spatially autocorrelated. This is demonstrated in Fingleton (1999) where extremely high values of the test statistics of the Moran test for spatial autocorrelation (Whittle, 1954; Anselin, 1988a) have been found. In this setting high values of Moran's I can be viewed as the counterpart of low values of the Durbin-Watson statistic having been established in spurious time-series regression. In both cases the behaviour of the test statistics is used as an indication of nonstationarity.

The stochastic process the OLS residuals **e** of the regression (2) are generated from usually has to be inferred by inspecting their behaviour. Fingleton (1999) leaves it as an open question how to separate the case of stationary positive autocorrelation ($0<\rho<1$) from the nonstationarity case ($\rho=1$). This means that the implicitly assumed error process

(3) $\boldsymbol{\varepsilon} = \rho_\varepsilon \mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\mu}, \qquad \boldsymbol{\mu} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}),$

is considered with $\rho_\varepsilon = 0$ under the null hypothesis of independently and identically distributed (i.i.d.) disturbances and with $\rho_\varepsilon > 0$ under the alternative hypothesis of spatially autocorrelated errors. The error process (3) is exactly the spatial analogue of the Markov process underlying the Durbin-Watson test taking only first order error correlation into account as a possible alternative. Note that spatial autocorrelation can be caused by both a SAR(1) and SMA(1), see e.g. Kelejian and Robinson (1995), Hepple (1995a, 1995b). However, manifestation of spatial nonstationarity can only be attributed to a SAR process. Moreover, Fingleton (1999) does not address the well-known power of the Moran I test towards misspecifications in the form of e.g. spatial heterogeneity (Anselin, 1988a). Being an advantage in some circumstances, this feature of the Moran I is not necessarily an advantage when investigating specific features of the data generating processes underlying the model in consideration.

Lauridsen and Kosfeld (2006) suggested a two-step application of a Lagrange Multiplier test for spatially autocorrelated errors. The LM error statistic (LME) developed in Anselin (1988a, 1988b),

(4) $$\text{LME} = (\mathbf{e'We} / \sigma^2)^2 / \text{tr}(\mathbf{W}^2 + \mathbf{W'W}) .$$

is asymptotical χ^2 distributed with 1 degree of freedom under H_0: $\rho_e = 0$. Therefore, a large LME value indicates either spatial nonstationarity or stationary, spatial error autocorrelation. This result corresponds to the suggestions of Fingleton (1999) with the Moran I test replacing the LM test. Next, under the null of nonstationarity, H_0:ρ_ε=1,

$$\boldsymbol{\Delta\varepsilon} = \boldsymbol{\mu} \Leftrightarrow \boldsymbol{\varepsilon} = \boldsymbol{\Delta}^+ \boldsymbol{\mu}$$

follows from the spatial error process (3) with $\boldsymbol{\Delta} = \mathbf{I} - \mathbf{W}$ as the spatial difference operator. $\boldsymbol{\Delta}^+$ denotes the Moore-Penrose generalised inverse which satisfies the conditions $\boldsymbol{\Delta}^+\boldsymbol{\Delta\Delta}^+ = \boldsymbol{\Delta}^+$ and $\boldsymbol{\Delta\Delta}^+\boldsymbol{\Delta} = \boldsymbol{\Delta}$. By employing the spatial difference operator $\boldsymbol{\Delta}$ to (2) the transformed regression equation

(5) $$\boldsymbol{\Delta}\mathbf{y} = \boldsymbol{\Delta}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$$

is obtained. Equation (5) implies that a regression of $\boldsymbol{\Delta}\mathbf{y}$ on $\boldsymbol{\Delta}\mathbf{X}$ provides i.i.d. errors, so that the LM error test statistic for this spatially differenced model (DLME) will be close to zero. On the other hand, if the null of nonstationarity, H_0: $\rho_\varepsilon = 1$, does not hold, then the spatial differencing will bring about an error term of the form

$$\boldsymbol{\Delta\varepsilon} = (\mathbf{I}\text{-}\mathbf{W})(\mathbf{I}\text{-}\rho_\varepsilon\mathbf{W})^{-1}\boldsymbol{\mu} \Leftrightarrow \boldsymbol{\mu} = (\mathbf{I}\text{-}\rho_\varepsilon\mathbf{W})\boldsymbol{\varepsilon}.$$

The spatially autocorrelated errors resulting from spatial “overdifferencing” are expected to go along with a positive DLME value. Concluding, the test strategy consists of calculating and inspecting the LME and the DLME values, leading to one of four conclusions as shown in Table 1 (where the test result is designated “positive” if the LM test statistic differs significantly from zero and “zero” otherwise).

(Table 1 around here)

A further advantage of the LM test strategy is that it is quite flexible. Thus, it is possible to control for omitted inference inflating features insofar that these can be incorporated as part of the likelihood function. Especially, it is important to account for unobserved heteroscedasticity as demonstrated by Anselin (1988b). The remainder of the present paper consists of a generalisation of the above LM test strategy by incorporating control for unobserved heteroscedasticity along the lines of Anselin (1988b).

2.3. Spurious regression and heteroscedasticity.

In case of heteroscedasticity the disturbance variance is allowed to vary across the observations. Here we consider the case where the error variance is functionally related to a set of variables via some positive function. Denoting by $\mathbf{\Omega}$ the covariance matrix of $\boldsymbol{\varepsilon}$, absense of heteroscedasticity implies $\mathbf{\Omega} = \sigma^2\mathbf{I}$, while presense of heteroscedasticity implies that $\mathbf{\Omega} = \mathrm{diag}(\sigma_1^2, \sigma_2^2, .., \sigma_n^2)$, where $\sigma_i^2 = f^+(\mathbf{z}_i, \boldsymbol{\alpha}_Z)$, with $\mathbf{z}_i$ being a P by 1 vector of observations of exogenous variables for region i which are positively related to σ_i^2 via the P by 1 parameter vector $\boldsymbol{\alpha}_Z$. A basin of econometric literature discusses the nature and implications of heteroscedasticity (Prais and Houthakker 1955; Eicker 1967; Horn, Horn and Duncan 1975; Taylor 1977; White 1980; Cragg 1982; Engle 1982, 1983; Messer and White 1984; MacKinnon and White 1985) as well as tests for heteroscedasticity (Harvey 1976; Breusch and Pagan 1979, 1980; Koenker 1981; Koenker and Bassett 1982; Ohtani and Toyoda 1980; Ali and Giacotto 1984). One seminal result is that unbiasedness and consistency of estimation procedures for linear regression are unaffected by heteroscedasticity, but that the covariance matrix for these parameters is inconsistently estimated, and that this bias may be severe, but that the inconsistency matters only marginally in cases where $\mathbf{z}_i$ consists of variables that are not parts of the design, i.e. the **X** matrix (White 1980; Messer and White 1984; MacKinnon and White 1985)). Thus, consideration of heteroscedasticity can be restricted to the case where $\mathbf{z}_i$ is a subset of $\mathbf{x}_i$.

Anselin (1988a, 1988b) developed a Lagrange Multiplier test for spatially autocorrelated errors, adjusted for unobserved heteroscedasticity. This test reads as

(6) $$\text{LMEH} = (\mathbf{e}'\mathbf{W}\mathbf{e}/\sigma^2)^2 / \text{tr}(\mathbf{W}^2 + \mathbf{W}'\mathbf{W}) + \mathbf{f}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{f}/2 ,$$

with $\mathbf{f} = (\mathbf{e}\#\mathbf{e}/\sigma^2 - \mathbf{i})$, where # denotes the Hadamard product, and is asymptotically χ^2 distributed with P+1 degrees of freedom under $H_0 : \rho_e = 0, \boldsymbol{\alpha}_Z = \mathbf{0}$. It should be noticed that the first part of the sum in (6) is equal to the simple LME test statistic for residual autocorrelation, while the second part is equal to a standard Breusch-Pagan test for heteroscedasticity, the latter denoted by LMH.

The suggestion of the present study is to provide a test strategy for nonstationarity by supplying the LME statistic (4) with the LMH statistic derived from (6), and the DLME statistic calculated according to (4) with a DLMH statistic derived from (6). These are inspected together with the LMEH and DLMEH tests.

2.4. Spatial cointegration and heteroscedasticity.

Spatial cointegration denotes the case where two or more variables in a regression are nonstationary, while the errors are stationary (Lauridsen and Kosfeld, 2006). A simple data generating process which generates two nonstationary but possibly cointegrating series with heteroscedastic regression disturbances is the following system:

(7) $$\mathbf{x} + \beta\mathbf{y} = \mathbf{u} \text{ , } \mathbf{u} = \mathbf{W}\mathbf{u} + \mathbf{e}_1 \text{ , } \mathbf{e}_1 = \boldsymbol{\Omega}^{1/2}\, \mathbf{v}_1$$

(8) $\mathbf{x} + \alpha\mathbf{y} = \mathbf{e_2}$, $\mathbf{e_2} = \mathbf{\Omega}^{1/2}\, \mathbf{v_2}$

where $\mathbf{v_1}$ and $\mathbf{v_2}$ are white noise processes, and $\mathbf{\Omega}$ = diag(exp($\alpha_z\mathbf{v_2}$)). From these definitions, a simple rearrangement gives

$$\mathbf{x} = \alpha(\alpha-\beta)^{-1}\mathbf{u} - \beta(\alpha-\beta)^{-1}\mathbf{e_2}$$

$$\mathbf{y} = -(\alpha-\beta)^{-1}\mathbf{u} + (\alpha-\beta)^{-1}\mathbf{e_2}$$

from which it is clear that **x** and **y** are SI(1) but that they cointegrate for any α different from 0 and certain β values, because ($\mathbf{x}+\alpha\mathbf{y}$) is I(0). Specifically, the relation will be non-integrated if (i) $\alpha=0$ or (ii) $\alpha>0$ and $\beta>\alpha$. The latter case defines a grey zone between cointegration and non-integration, which we will denote as near-integration.

The system defined by (7) and (8) is a general version of the homoscedastic system suggested in Lauridsen and Kosfeld (2006), which can be obtained as a special case by setting $\alpha_z = 0$.

We suggest that the above LM strategy may apply to this situation. Specifically, a regression of **y** on **X** represents a cointegrating relation (if LME and LMEH are zero, and DLME and DLMEH are positive) or a non-integrating relation (if LME and LMEH are positive, and DLME and DLMEH are zero). The limiting case of "near integration" ($\alpha>0$, $\beta>\alpha$) will also be revealed (if LME and LMEH as well as DLME and DLMEH are positive). Further, discrepancies between LME and LMEH and/or between DLME and DLMEH will be ascribed to as a consequence of unobserved heteroscedasticity, and observed using the LMH and DLMH values.

3. Monte Carlo simulation studies: Designs and results.

In this section, the small-sample properties of the above suggested test strategies will be investigated using Monte Carlo simulation studies. The chosen Monte Carlo designs are outlined together with the results. All calculations are done using SAS/IML, including the software's standard routines for random number generation.

3.1. Spurious regression and heteroscedasticity.

To investigate the finite sample properties of the suggested LM test strategy for spurious regression adjusted for heteroscedasticity, the following Monte Carlo design were investigated:

For specific sample size n: Perform 10,000 iterations:

Generate $\mathbf{v_x}$ and $\mathbf{v_y}$ as independent N(0,1) series.

Let $\alpha = 0$ or 1.

Let $\sigma_i^2 = \exp(\alpha v_{xi})$.

Let $\mathbf{\Omega} = \text{diag}(\sigma_1^2, .., \sigma_n^2)$.

Let $\mathbf{e_x} = \mathbf{\Omega}^{1/2}\,\mathbf{v_x}$ and $\mathbf{e_y} = \mathbf{\Omega}^{1/2}\,\mathbf{v_y}$.

Let $\mathbf{x} = (\mathbf{I}-\rho_x\mathbf{W})^{-1}\mathbf{e_x}$.

Let $\mathbf{y} = (\mathbf{I}-\rho_y\mathbf{W})^{-1}\mathbf{e_y}$.

Regress $\mathbf{y}$ on $\mathbf{X} = [\mathbf{i}\ \mathbf{x}]$ and $\mathbf{\Delta y}$ on $\mathbf{\Delta X}$.

Obtain LME, LMH, LMEH, DLME, DLMH, and DLMEH.

Report the percentage of cases out of 10,000 where each test exceeds the 5 per cent critical value of $\chi^2(1)$ and $\chi^2(2)$

To investigate the impact of contiguity matrix type, we make use of the rook and queen type of regular contiguity matrices based on an $r\times r$ board (so that $n = r^2$) with r assumed to take the values 5, 10, 15, and 20. The rook matrix represents a square tesselation with a connectivity of 4 for the inner fields on the chessboard and 1 and 2 for the corner - and border fields, respectively.

The queen matrix represents an octogonal tesselation with a connectivity of 8 for the inner fields and 3 and 5 for the corner and border fields. Thus, these tesselations represent extremes for a number of patterns, including the hexagonal tesselation, which is of importance due to its application for empirical maps in vector and raster based GIS (Boots and Tiefelsdorf, 2000; Tiefelsdorf, 2000). Actually, the hexagonal tesselation can be constructed from the queen tesselation by deleting connections from any field to the fields vertically above and below this. Moreover, most empirically observed regional structures in spatial econometrics are made up of regions with a connectivity within the range of the rook and queen tesselations. Further, irregular matrices based on the 275 Danish municipalities are applied: An n=36 matrix based on the municipalities located on the island of Funen, an n=97 matrix made up of the municipalities located on Seeland together with the adjacent islands Lolland and Falster, an n=141 matrix created from the municipalities located on the peninsula of Jutland, and the full matrix of n=275 Danish municipalities, which consists of the above municipalities plus 5 municipalities located on the island of Bornholm. The map of the 275 municipalities, together with the above partitioning, is shown in Appendix Figure A.1.

The behaviour of the strategy under H_0 : nonstationarity as well as H_1 : stationarity (including the case of near nonstationarity) is investigated by assuming ρ_y to take the values (0.0, 0.1, 0.2, ..., 0.8, 0.9, 0.99, 1.00). For each of these, ρ_x is assumed to take the values (0.0, 0.1, 0.2, ..., 0.8, 0.9, 0.99, 1.00). For the cases of nonstationarity, we use the Moore-Penrose generalised inverse $(\mathbf{I}\text{-}\mathbf{W})^{+}$ instead of $(\mathbf{I}\text{-}\mathbf{W})^{-1}$. The results are provided in Figures 1-8.

For the homoscedastic case the results from Lauridsen and Kosfeld (2006) regarding the LME and DLME tests are confirmed. It is found that the procedure performs well, and that the performance of the procedure is acceptable, even for small sample sizes. That the case of near nonstationarity causes problems in identifying the “true” data generating process is well-known from time series analysis. However, contrary to time series analysis, spatial dependence of moderate size (i.e. values of about 0.5) in economic systems seems to be much more reasonable than the case of near-nonstationarity (see e.g. Rey and Montouri, 1999; Kosfeld, Eckey and Dreger, 2002). For the heteroscedastic case, it is observed that the LME and DLME tests are robust toward heteroscedasticity. Thus, a large LME value together with a small DLME value indicate spurious regression irrespective of whether heteroscadasticity is present or not.

Opposed to these, the LMH test is not robust towards spurious regression. According to the curves under homoscedasticity, the size of the LMH test is found to increase for increasing spurious regression, while the curves under heteroscedasticity shows that the power drops for increasing spuriousity under heteroscedasticity. This anomality is not shared by the DLMH test, which is found to have a size close to 0.05 and a power around 0.95. Thus, the DLMH test can be used as a proper indicator of heteroscedasticity irrespective of whether nonstationarity, stationary spatial autocorrelation or absense of spatial autocorrelation is present.

From the curves for the LMEH and the DLMEH tests, it is seen that these tests combine the properties of the LME and LMH tests and the DLME and DLMH tests, respectively. Although the joint tests could be applied as indications of spurious regression and/or heteroscedasticity, they convey no information that is not already provided by the simple LM tests.

Thus, combining the evidence for the LME, DLME and DLMH tests, a precise diagnosis on spurious regression and heteroscedasticity can be concluded. It is noticed that this diagnosis works well for medium (n=100 for the rook matrix and n=97 for the empirical matrix) to large sample sizes. For small samples (n=25 for the rook matrix and n=36 for the empirical matrix), the powers of the tests are seen to be considerably smaller. The diagnosis strategy for medium to large samples is outlined in Table 2.

(table 2 around here)

3.2. Spatial cointegration and heteroscedasticity.

To investigate the finite sample properties of the suggested test strategy for cointegration using the suggested example, the following Monte Carlo design was investigated:

For specific sample size n: Perform 10,000 iterations:

> Generate $\mathbf{v_1}, \mathbf{v_2}$ as independent N(0,1) series.
>
> Let $\mathbf{\Omega} = \mathrm{diag}(\exp(\alpha_z \mathbf{v_2}))$.
>
> Let $\mathbf{e_i} = \mathbf{\Omega}^{1/2}\, \mathbf{v_i}$, i=1,2
>
> Let $\mathbf{u} = (\mathbf{I}\text{-}\mathbf{W})^{-1}\, \mathbf{e_1}$.
>
> Let $\mathbf{x} = \alpha(\alpha\text{-}\beta)^{-1}\mathbf{u} - \beta(\alpha\text{-}\beta)^{-1}\, \mathbf{e_2}$ and $\mathbf{y} = -(\alpha\text{-}\beta)^{-1}\mathbf{u} + (\alpha\text{-}\beta)^{-1}\, \mathbf{e_2}$.
>
> Regress $\mathbf{y}$ on $\mathbf{x}$ and $\mathbf{\Delta y}$ on $\mathbf{\Delta x}$.
>
> Report LME, LMH, LMEH and DLME, DLMH, DLMEH.

Report the percentage of cases out of 10,000 where each test exceeds the 5 per cent critical value of $\chi^2(1)$ and $\chi^2(2)$.

To investigate the impact of contiguity matrix type, we again use the regular rook and queen type contiguity matrices based on an $r\times r$ board with r assumed to take the values 5, 10, 15, and 20 and the irregular matrices based on the Danish case. Further, the behaviour of the strategy under H_0: nonstationarity as well as H_1: stationarity (including the case of near nonstationarity) is investigated for varying α and β. Specifically, α was varied across the values (0, 0.1, 0.2, ..., 0.8, 0.9, 1.0). For each of these, β was varied across the same values, with an exception for the cases when $\alpha=\beta$. For these, β was set to (α+0.01), except for the α=1.0 cases, where β was set to 0.99. The results are shown in Figures 9-16.

For the case of homoscedasticity, the results are in agreement with results from Lauridsen and Kosfeld (2006): The procedure performs well, especially for fairly large n, and the performance of the procedure is acceptable, even for fairly small sample sizes. Especially, in the case of cointegration (α=1) and non-integration (α=0), the procedure works excellently, while the greyzone case of near-integration ($0<\alpha<1$, $\beta>\alpha$) is characterized by inconclusive test sizes. For the heteroscedastic case, the LME and DLME tests are found to be robust toward heteroscedasticity, with only a slight drop in power for α and β close to 1. Thus, they can be used as indication of non-integration problems irrespective of whether heteroscedasticity is present or not.

The LMH test, on the other hand, is found to be sensitive to non-integration, as the size of the test rises for the non-integrated case, while the the power is low under co-integration as well as under strong non-integration. The size of the DLMH test, on the other hand, is close to the desired value of 0.05 irrespective of integration status, but the power of the test is reduced under cointegration, so that the test works properly only in the case of non-integration. Thus, a low DLMH test value safely indicates homoscedasticity, while a medium to high DLMH value indicates heteroscedasticity.

Finally, the LMEH and DLMEH tests generally reflect a combination of the properties of the LME and LMH, and DLME and DLMH, respectively. It is especially noticed that the LMEH test uniformly has a size close to the expected 0.05 under cointegration and homoscedasticity, while the power towards heteroscedasticity is flawed under cointegration.

Thus, a test strategy may be summarized as follows: Decide, based on the LME and DLME whether cointegration or non-integration is present. In any case, a low DLMH test is a safe indication of homoscedasticity. In the case of near-integration or non-integration, a high DLMH test indicates heteroscedasticity. This is also partly the reflected for the case of cointegration, where a medium to high value of DLMH should be seen as an indication of heteroscedasticity. These guidelines are summarised in Table 3.

(table 3 around here)

These features are observed to hold well for medium sized (n=100 for the regular matrix and n=97 for the empirical matrix) to large samples. For small samples (n=25 for the regular matrix and n=36 for the empirical matrix), the powers of the tests are considerably flawed.

4. EMPIRICAL ILLUSTRATIONS

4.1. A commuting model

We elaborate on an empirical example investigated in Lauridsen and Nahrstedt (1999) and Lauridsen (2006). The model is concerned with determination of a regression model for outcommuting ratios as a function of unemployment, participation rate, density of working places and average household size. Data were from a 1994 census for 275 Danish municipalities. See

Table 4 for a description of the data.

(table 4 around here)

Table 5 presents the estimated model. In Lauridsen (2006) it was left as an open question whether the unexpected negative sign for the UNEMP coefficient was caused by spuriosity due to spatial nonstationarity. The LM tests clearly point to stationarity of the geo-referenced variables as well as of the residuals. It is concluded that the single variables as well as the entire regression are stationary. From this point of view, the negative sign for unemployment is rather due to structural properties than to spatial nonstationarity. However, estimation of the commuting model does not account for the unobserved heteroscedasticity that is revealed by the DLMH test.

(table 5 around here)

4.2. A growth model

The topic of this section is a model developed in Kosfeld and Lauridsen (2004). The model is concerned with determination of a regression model for regional labour productivity as a function of population growth, growth technology, depreciation of capital and physical and human capital accumulation. Data were from a 2000 census for 180 German labour markets of which 133 are located in West Germany and 47 in East Germany. See Table 6 for a description of the data.

(table 6 around here)

Table 7 presents the estimated models. For both models the LM tests support the conclusion of Kosfeld and Lauridsen (2004) that the regressions are stationary. Just as with the commuting model, the DLMH test points to some degree of unobserved heteroscedasticity.

(table 7 around here)

5. CONCLUSIONS

Recently, it was established how to separate the case of spatial nonstationarity from the case of stationary positive autocorrelation, leading to reliable diagnostics for spurious spatial regression and for the existence of spatial cointegrating relations. The present study contributes to existing knowledge by showing that the strategy for detecting spatial nonstationarity is robust towards unobserved heteroscedasticity. It is further shown that an LM test for heteroscedasticity based on a regression with spatially differenced variables properly diagnoses heteroscedasticity irrespective of nonstationarity or cointegration status. By means of Monte Carlo simulations it is demonstrated that the finite sample properties of the suggested methodology are as desired.

REFERENCES

Ali M, Giaccotto C (1984) A Study of Several New and Existing Tests for Heteroscedasticity in the General Linear Model. *Journal of Econometrics*, 26, 355-74.

Anselin L (1988a) *Spatial Econometrics: Methods and Models*. Kluwer.

Anselin L (1988b) Lagrange Multiplier Tests Diagnostics for Spatial Dependence and Spatial Heterogeneity. *Geographical Analysis*, 20, 1-17.

Anselin L, Florax RJGM (eds.)(1995) *New Directions in Spatial Econometrics*. Springer-Verlag.

Bivand R (1980) A Monte-Carlo Study of Correlation Coefficient Estimation with Spatially Autocorrelated Observations. *Quaestiones Geographicae*, 6, 5-10.

Breusch T, Pagan A (1979) A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47, 1287-94.

Breusch T, Pagan A (1980) The LM Test and Its Applications to Model Specification in Econometrics. *Review of Economic Studies*, 47, 239-54.

Cliff AD, Ord JK (1981) *Spatial Processes: Models and Applications*. Pion.

Clifford P, Richardson S (1985) Testing the Association between Two Spatial Processes. *Statistics and Decisions*, Supplement no. 2, 155-60.

Clifford P, Richardson S, Hémon D (1989) Assessing the Significance of the Correlation between Two Spatial Processes. *Biometrics*, 45, 123-34.

Cragg J (1982) Estimation and Testing in Time Series Regression Models with Heteroscedastic Disturbances. *Journal of Econometrics*, 20, 135-57.

Eckey H-F (2001) Der wirtschaftliche Entwicklungsstand in den Regionen des Vereinigten Deutschlands, Discussion Papers in Economics No. 20/01, University of Kassel, Dept. of Economics.

Eicker F (1967) Limit Theorems for Regression with Unequal and Dependent Errors. In LeCam L, Neyman J (eds.) *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 59-82.

Engle R (1982) Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflations. *Econometrica*, 50, 987-1008.

Engle R (1983) Estimates of the Variance of U.S. Inflation Based on the ARCH Model. *Journal of Money, Credit, and Banking* 15, 286-301.

Fingleton B (1999) Spurious Spatial Regression: Some Monte Carlo Results with a Spatial Unit Root and Spatial Cointegration. *Journal of Regional Science*, 39, 1-19.

Griffith DA (1992) Simplifying the Normalizing Factor in Spatial Autoregressions for Irregular Lattices. *Papers in Regional Science*, 71 71-86.

Haining RP(1990) *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press.

Harvey A (1976) Estimating Regression Models with Multiplicative Heteroscedasticity. *Econometrica*, 44, 461-5.

Horn D, Horn A, Duncan G (1975) Estimating Heteroscedastic Variances in Linear Models.- *Journal of the American Statistical Association*, 70, 380-5.

Koenker R (1981) A Note on Studentizing a Test for Heteroscedasticity. *Journal of Econometrics*, 17, 107-12.

Koenker R, Bassett G (1982) Robust Tests for Heteroscedasticity Based on Regression Quantiles. *Econometrica*, 50, 43-61.

Kosfeld R, Eckey H-F, Dreger C (2005) Regional Convergence in Unified Germany: A Spatial Econometric Perspective. In Dreger C, Galler HP (eds) *Advances in Macroeconometric Modeling*, Nomos, Baden-Baden.

Kosfeld R, Lauridsen J (2004) Dynamic Spatial Modelling of Regional Convergence Processes. *Empirical Economics*, 29, 705-22.

Lauridsen J (2006) Spatial autoregressively distributed lag models: equivalent forms, estimation, and an illustrative commuting model. Discussion paper, Department of Statistics and Demography, University of Southern Denmark. *Annals of Regional Science* (forthcoming).

Lauridsen J, Kosfeld R (2004) A Wald test for Spatial Nonstationarity. *Estudios de Economía Aplicada*, 22, 475-486.

Lauridsen J, Kosfeld R (2006) A Test Strategy for Spurious Spatial Regression, Spatial Nonstationarity, and Spatial Cointegration. *Papers in Regional Science* (forthcoming).

Lauridsen J, Nahrstedt B (1999) Spatial Patterns in Intermunicipal Danish Commuting. In G. Crampton (ed.): *Regional Unemployment, Job Matching, and Migration*, 49-62, Pion Limited.

MacKinnon J, White H (1985) Some Heteroscedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties. *Journal of Econometrics*, 19, 305-25.

Messer K, White H (1984) A Note on Computing the Heteroscedasticity Consistent Covarian-ce Matrix Using Instrumental Variable Techniques. *Oxford Bulletin of Economics and Statistics*, 46, 181-4.

Mur J, Trívez FJ (2003) Unit Roots and Deterministic Trends in Spatial Econometric Models. *International Regional Science Review*, 26, 289-312.

Ohtani K, Toyoda T (1980) Estimation of Regression Coefficients After a Preliminary Test for Homoscedasticity. *Journal of Econometrics*, 12,, 151-9.

Ord JK (1975) Estimation Methods for Models of Spatial Interaction. *Journal of the American Statistical Association*, 70, 120-6.

Prais S, Houthakker H (1955) *The Analysis of Family Budgets*. Cambridge University Press, NY.

Rey S, Montouri B (1999) US Regional Income Convergence: A Spatial Econometric Perspective. *Regional Studies*, 33, 146-156.

Richardson S (1990) Some Remarks on the Testing of Association between Spatial Processes. In Griffith DA (ed.) *Spatial Statistics: Past, Present and Future*. Image.

Richardson S, Hémon D (1981) On the Variance of the Sample Correlation between Two Independent Lattice Processes. *Journal of Applied Probability*, 18, 943-8.

Ripley BD (1981) *Spatial Statistics*. Wiley.

Taylor W (1977) Small Sample Properties of a Class of Two Stage Aitken Estimators. *Econometrica*, 45, 497-508.

Upton GJG, Fingleton B (1985) *Spatial Data Analysis by Example, vol. 1*. Wiley.

White H (1980) A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica*, 48, 817-38.

Whittle P (1954) On Stationary Processes in the Plane. *Biometrika*, 41, 434-49.

TABLE 1. OUTCOMES OF THE TWO-STAGE LM TEST

	DLME zero:	DLME positive:
LME zero:	-	Absence of spatial autocorrelation
LME positive:	Spatial nonstationarity (spurious regression)	Stationary spatial autocorrelation

TABLE 2. A STRATEGY FOR TESTING SPURIOUS REGRESSION AND/OR HETEROSCEDASTICITY

	Homoscedasticity	Heteroscedasticity
Absence of spatial autocorrelation	LME low DLME high DLMH low	LME low DLME high DLMH high
Stationary spatial autocorrelation	LME inconclusive DLME inconclusive DLMH low	LME inconclusive DLME inconclusive DLMH high
Nonstationarity (spurious regression)	LME high DLME low DLMH low	LME high DLME low DLMH high

Note. ‘Inconclusive’ refer to cases where 0.05<power<0.95

TABLE 3. A STRATEGY FOR TESTING COINTEGRATION AND/OR HETEROSCEDASTICITY

	Homoscedasticity	Heteroscedasticity
Cointegration	LME low	LME low
$(\alpha=1; 0<\alpha<1, \beta<\alpha)$	DLME high	DLME high
	DLMH low	DLMH inconclusive
Non-integration	LME high	LME high
$(\alpha=0)$	DLME low	DLME low
	DLMH low	DLMH high
Near-integration	LME inconclusive	LME inconclusive
$(0<\alpha<1, \beta>\alpha)$	DLME inconclusive	DLME inconclusive
	DLMH low	DLMH high

Note. ‘Inconclusive’ refer to cases where $0.05<\text{power}<0.95$

TABLE 4. VARIABLES USED FOR COMMUTING STUDY

Variable	Definition	Mean	S.D.	Min	Max
OUTCOM	Number of persons with residence in the municipality and workplace in another municipality in percentage of the number of workplaces in the municipality[a]	58.14	37.79	6.00	237.00
PSH1766	Population share of 17-66 year-olds (%)[a]	65.22	2.85	57.90	74.20
WORKPL	Number of workplaces per 100 inhabitants[a]	43.11	11.63	21.00	100.00
IPHOUS	Number of inhabitants per household[a]	2.39	0.16	1.74	2.77
UNEMP	Number of unemployed per 100 17-66 year-olds[a]	9.37	2.24	5.00	18.70
Proximity matrix:					
W1	Neighbourhood matrix for N=275 Danish municipalities[b]				
	Description of number of links per municipality:	4.59	1.68	1	8
	Density of $\mathbf{W_1}$ = .017				
W	Row standardization of $\mathbf{W_1}$				

Data collected 1994, for N=275 Danish municipalities.

Source: a : Statistics Denmark, Copenhagen.

b : Own construction.

TABLE 5. ESTIMATION OF THE COMMUTING MODEL

Dependent variable: OUTCOM.

Variable	Parameter	Standard Error	T value	Probability
Intercept	-264.63	33.60	-7.88	<.001
UNEMP	-3.39	0.53	-6.46	<.001
PSH1766	6.30	0.36	17.49	<.001
WORKPL	-2.33	0.10	-22.99	<.001
IPHOUS	18.79	8.08	2.32	0.021

Tests for nonstationarity of variables:

Variable	LME	Probability	DLME	Probability
OUTCOM	38.27	<0.001	69.62	<0.001
UNEMP	449.77	<0.001	46.53	<0.001
PSH1766	554.34	<0.001	47.12	<0.001
WORKPL	498.51	<0.001	69.25	<0.001
IPHOUS	547.90	<0.001	49.53	<0.001

Tests for residual nonstationarity and heteroscedasticity:

	LME		DLME	
	48.49	<0.001	54.03	<0.001
	LMH		DLMH	
	150.76	<0.001	77.17	<0.001

Table 6. VARIABLES USED IN THE GROWTH MODEL

Variable	Definition	Mean	S.D.	Min	Max
LGDPER	Log gross domestic product[a] per total employment 2000	10.72	0.16	10.29	11.12
LGDPCR	Log gross domestic product per capita 2000[a]	20.94	4.88	12.07	40.32
EAST	East-West Dummy[a]	0.26	0.44	0	1
LDTW	Log (depreciation rate + rate of technical Progress + growth rate of population) (Averages resp. representative values for 90ties)[a]	-2.89	0.12	-3.17	-2.60
LHUMAN	Log proportion of highly educated people per total employment 2000 (Secondary school + technical college + university degree)[a]	2.55	0.28	1.98	3.41
LNBF	Log newly founded business per 1000 inhabitants 2000[a]	1.90	0.17	1.51	2.34
Proximity matrix:					
W*	Neighbourhood matrix for N=180 German labour markets[b]				
	Number of links per labour market	5.22	1.90	1	12
	Density of **W*** = .029				
W	Row standardization of **W***				

Data constructed for N=180 German labour markets from districtional and state data

Source: a: Volkswirtschaftliche Gesamtrechnung der Länder (Statistical State Office Baden-Württemberg); Statistik regional, Statistisches Jahrbuch (Federal Statistical Office Germany); German statistical state offices; Own construction.

b: University of Kassel, Department of Economics (see Eckey, 2001).

TABLE 7. ESTIMATION OF THE GROWTH MODEL

	Income model: LGDPCR		Productivity model: LGDPER	
Variable	Coefficient	Stand. err.	Coefficient	Stand. err.
Intercept	10.018**	0.445	10.613**	0.291
EAST	-0.346**	0.037	-0.290**	0.024
LDTW	0.317*	0.144	0.110	0.094
LHUMAN	0.254**	0.039	0.168**	0.025
LNFB	0.138*	0.068	0.041	0.044
Tests for nonstationarity of variables:				
Variable	LME	Probability	DLME	Probability
LGDPCR	148.47	<0.001	29.24	<0.001
LGDPER	202.06	<0.001	26.11	<0.001
LDTW	431.00	<0.001	7.94	0.005
LHUMAN	433.28	<0.001	13.29	<0.001
LNFB	433.34	<0.001	26.51	<0.001
Tests for residual nonstationarity and heteroscedasticity:				
Income model:	LME		DLME	
	19.09	<0.001	40.60	<0.001
	LMH		DLMH	
	22.27	<0.001	21.36	<0.001
Productivity model:	LME		DLME	
	19.48	<0.001	41.02	<0.001
	LMH		DLMH	
	19.82	<0.001	46.87	<0.001

Note: **: 1% significance level; *: 5%: significance level.

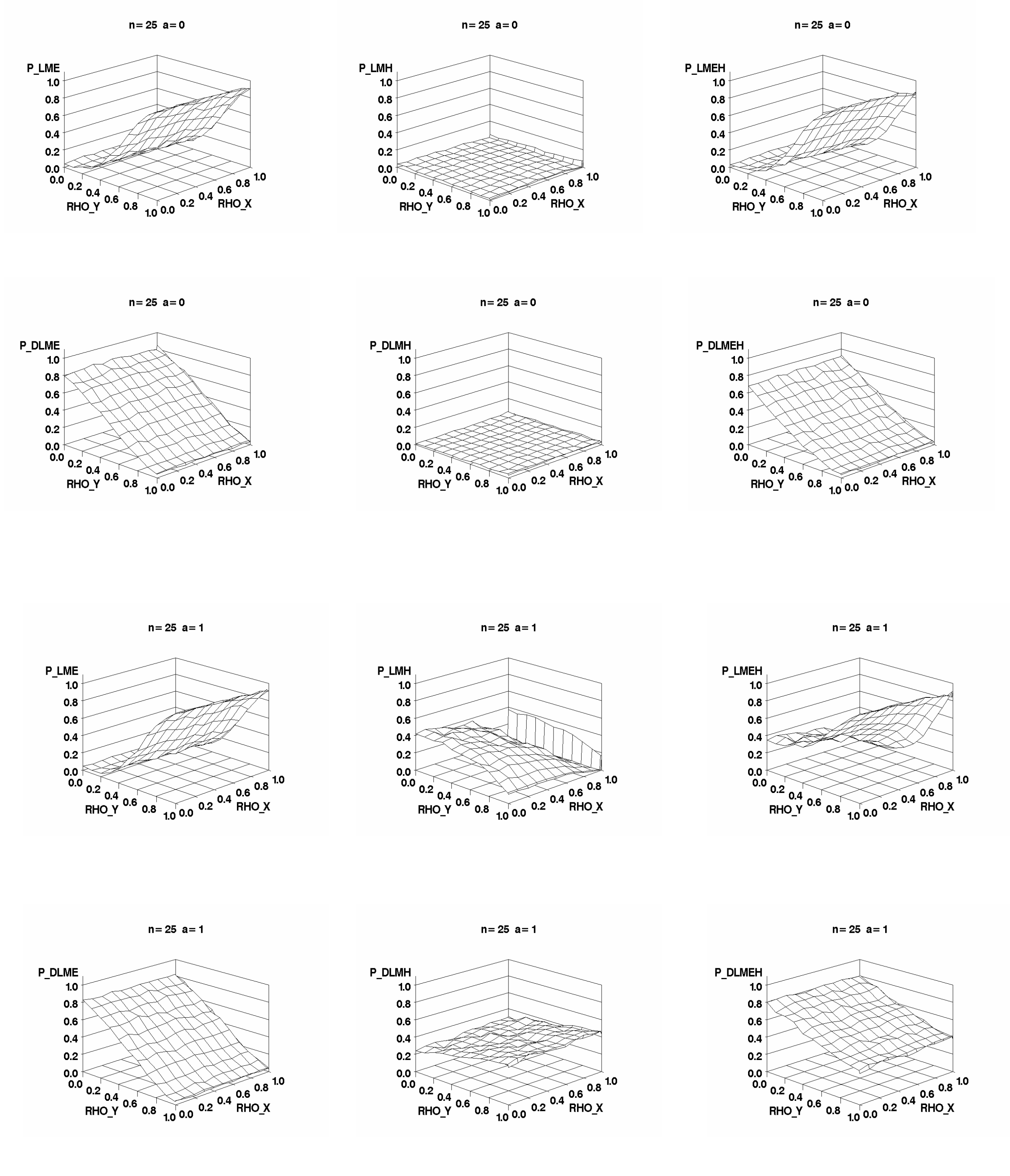

Figure 1. Monte Carlo results for spurious regression study: Rook matrix – n=25.

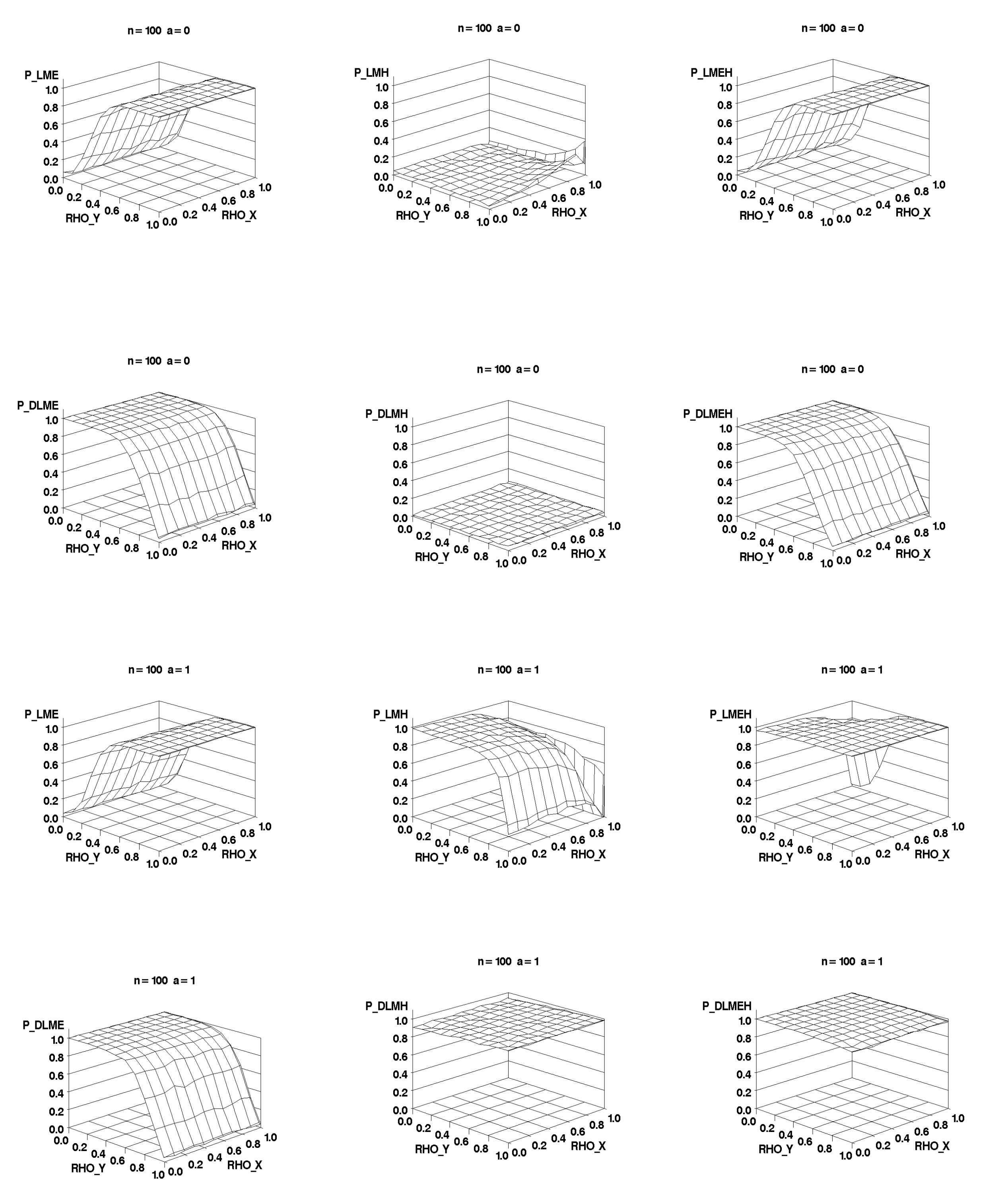

Figure 2. Monte Carlo results for spurious regression study: Rook matrix – n=100.

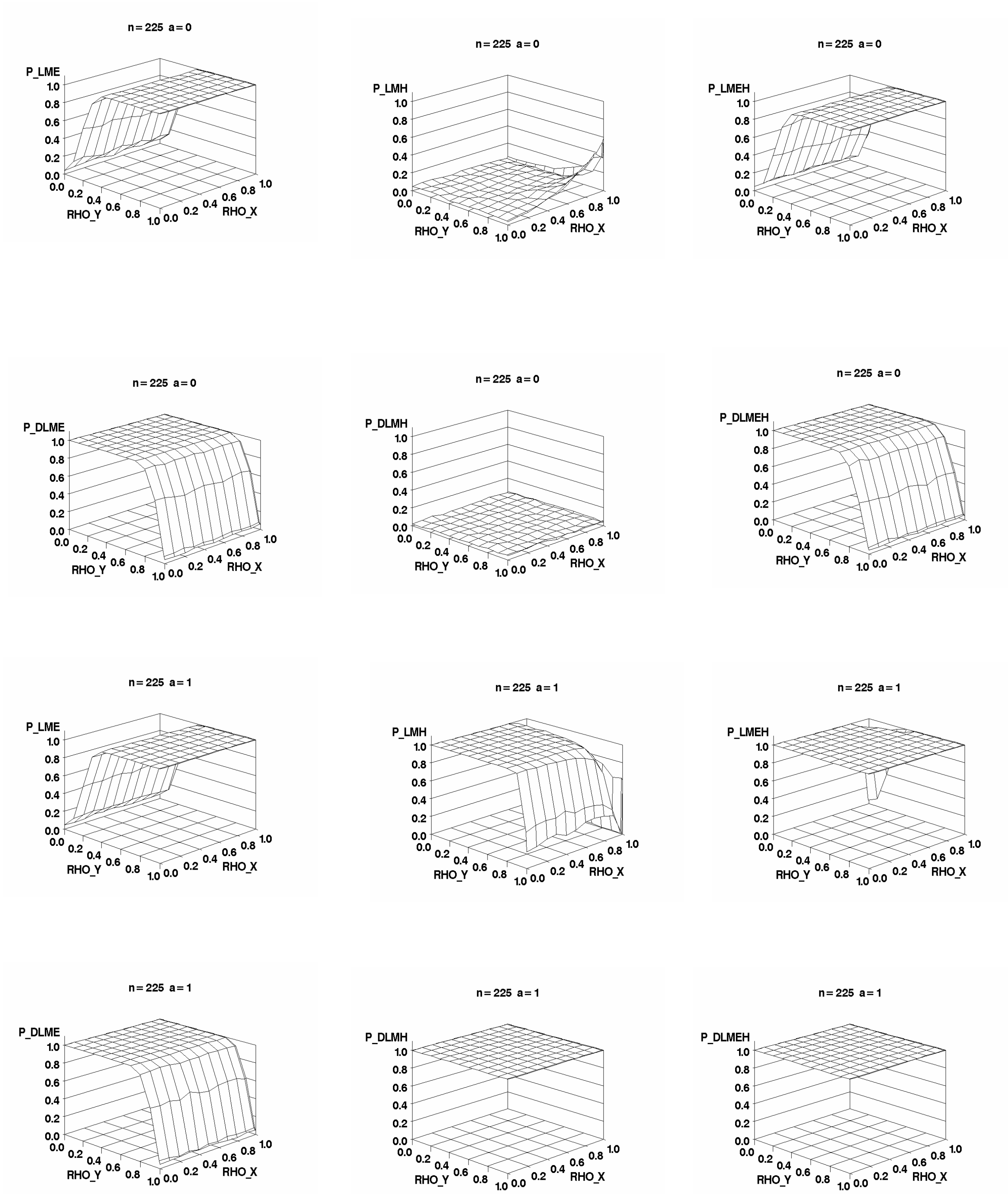

Figure 3. Monte Carlo results for spurious regression study: Rook matrix – n=225.

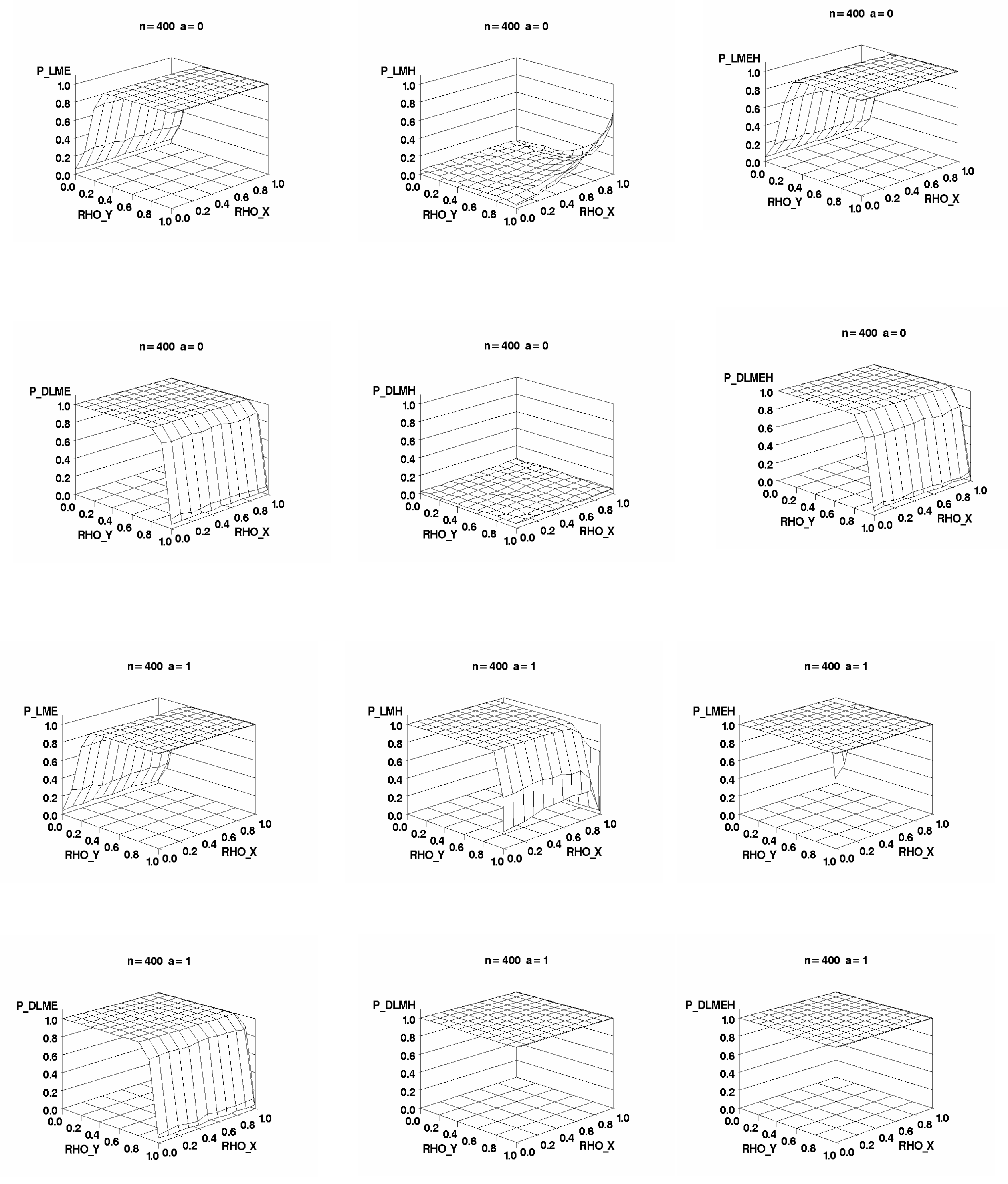

Figure 4. Monte Carlo results for spurious regression study: Rook matrix – n=400.

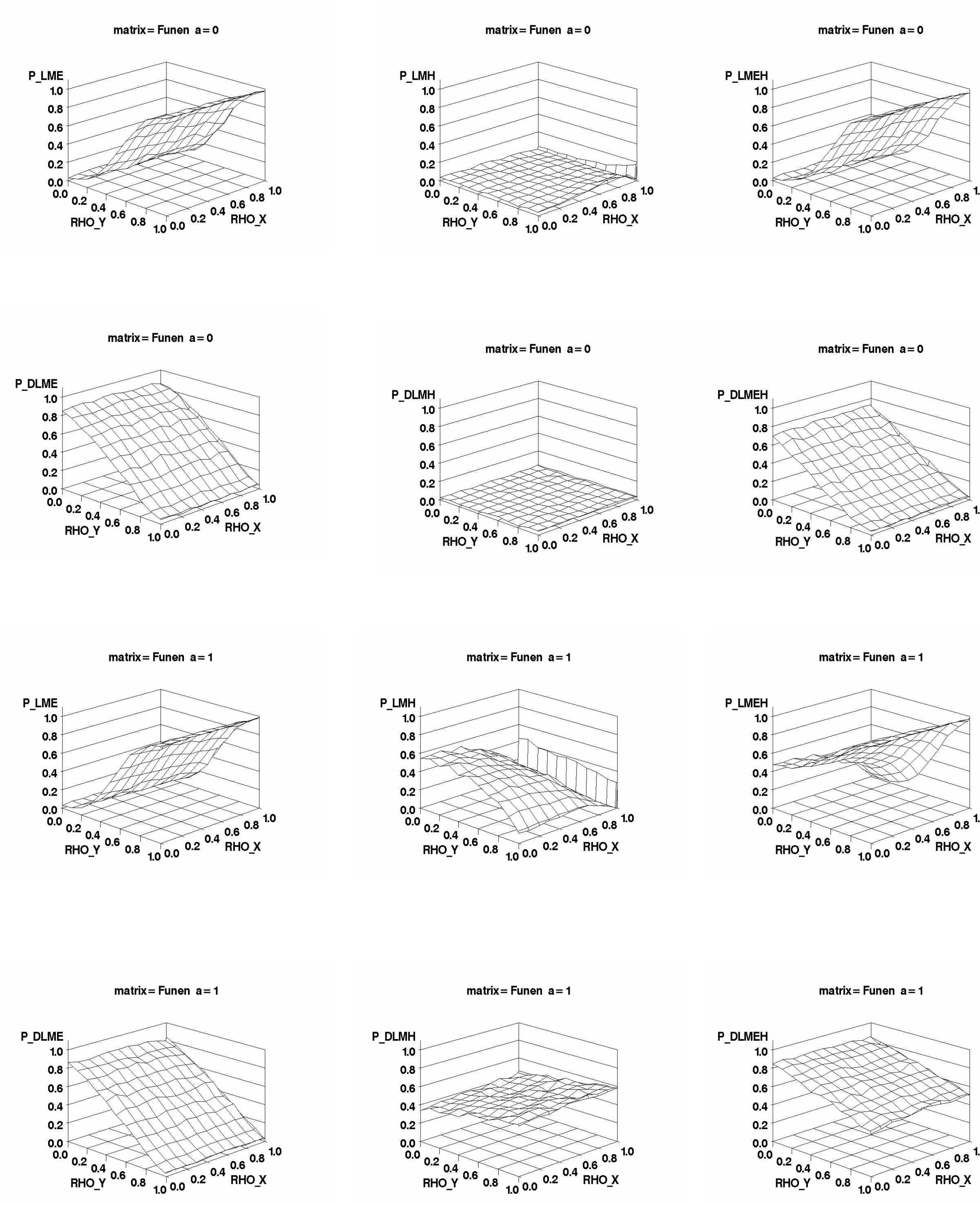

Figure 5. Monte Carlo results for spurious regression study: Empirical matrix – Funen (n=36).

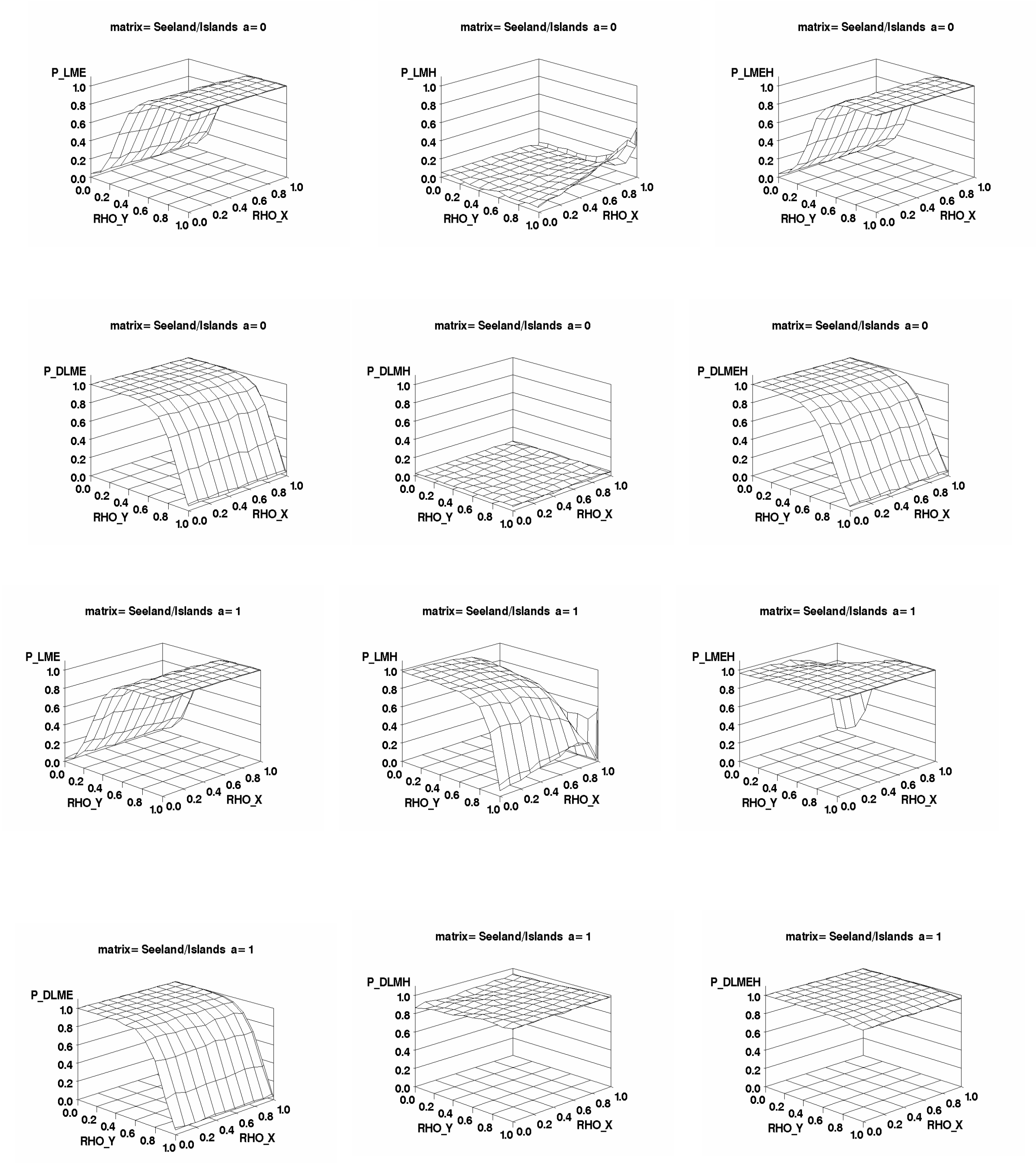

Figure 6. Monte Carlo results for spurious regression study: Empirical matrix – Seeland (n=97).

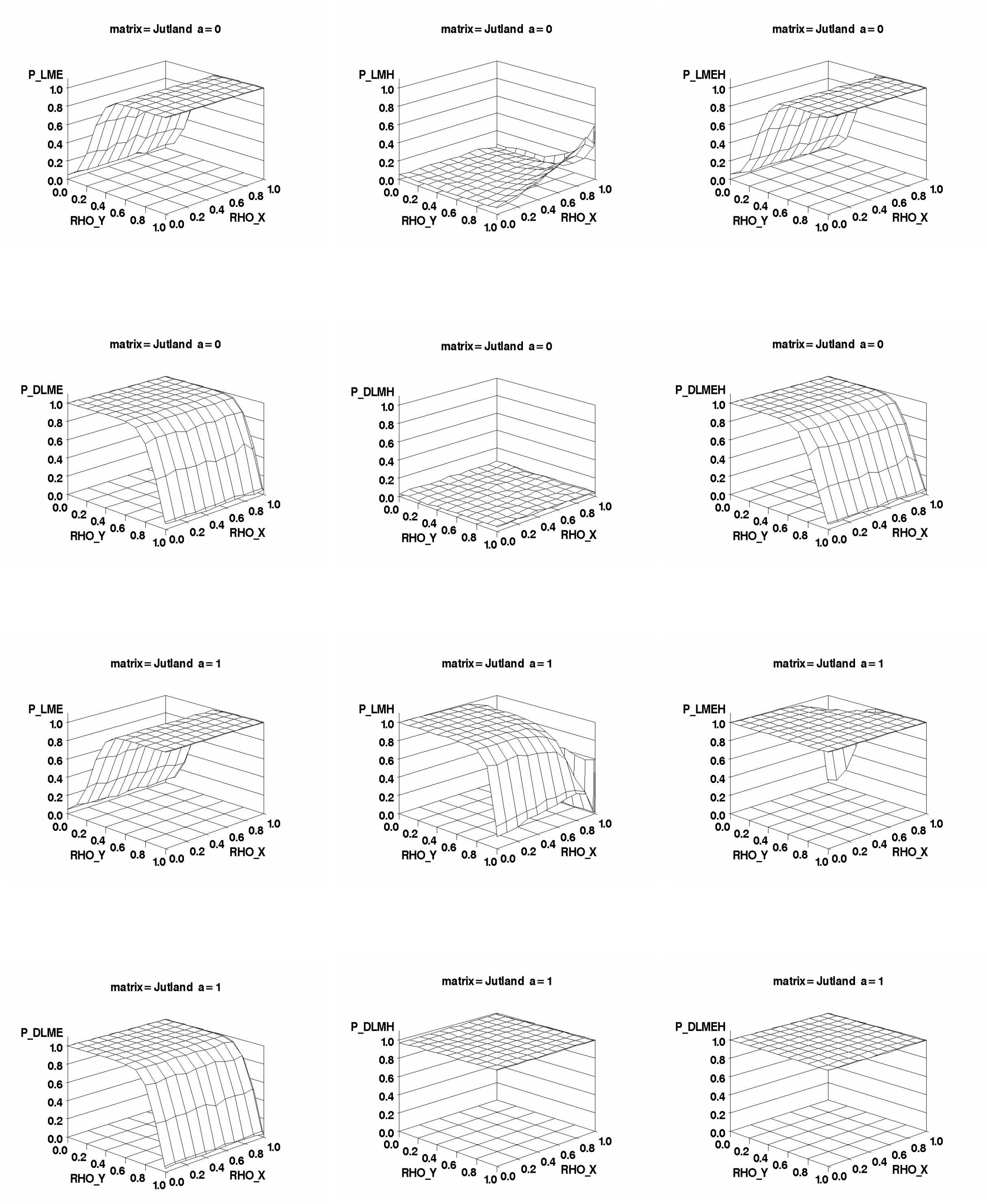

Figure 7. Monte Carlo results for spurious regression study: Empirical matrix – Jutland (n=141).

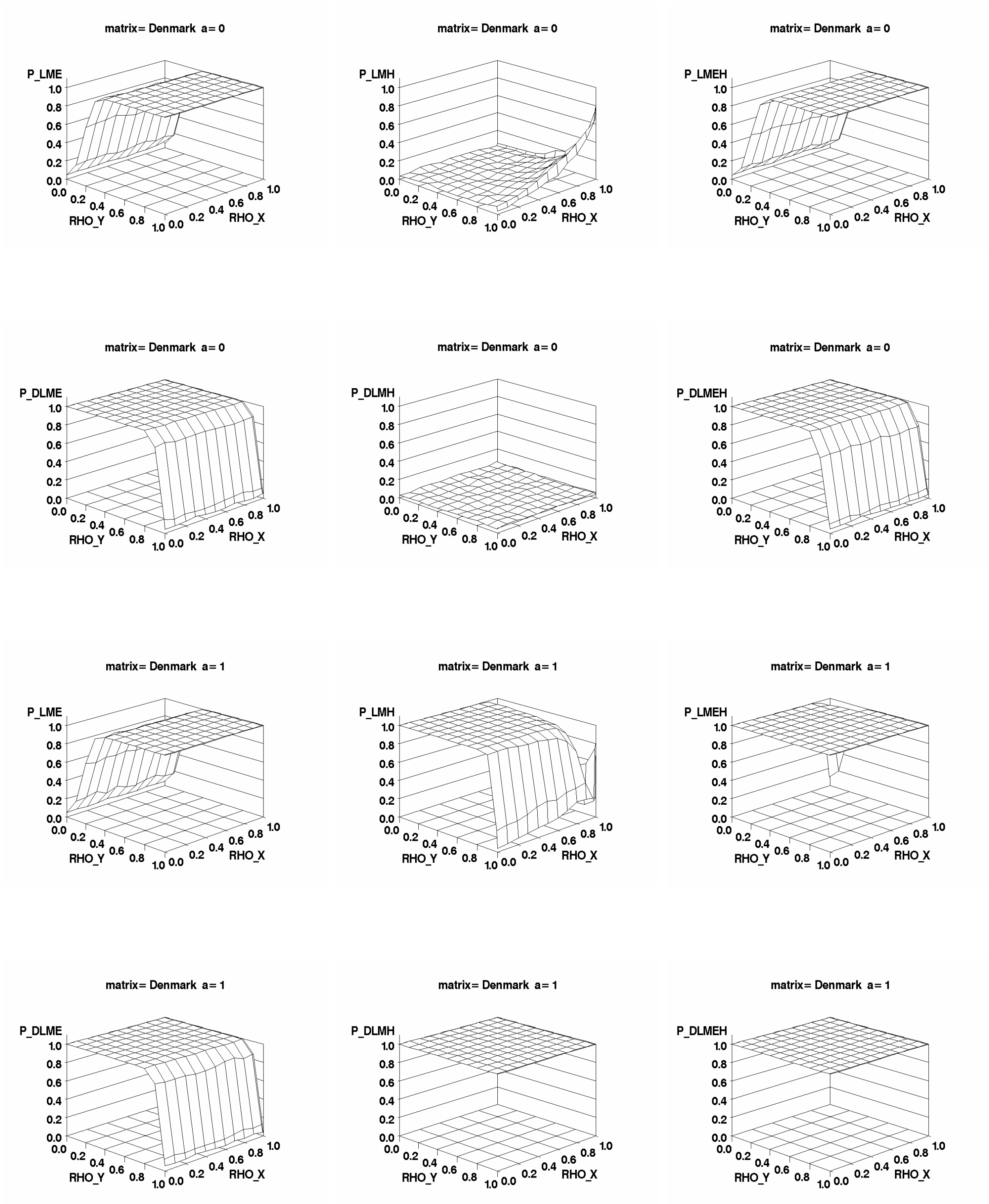

Figure 8. Monte Carlo results for spurious regression study: Empirical matrix – Denmark (n=275).

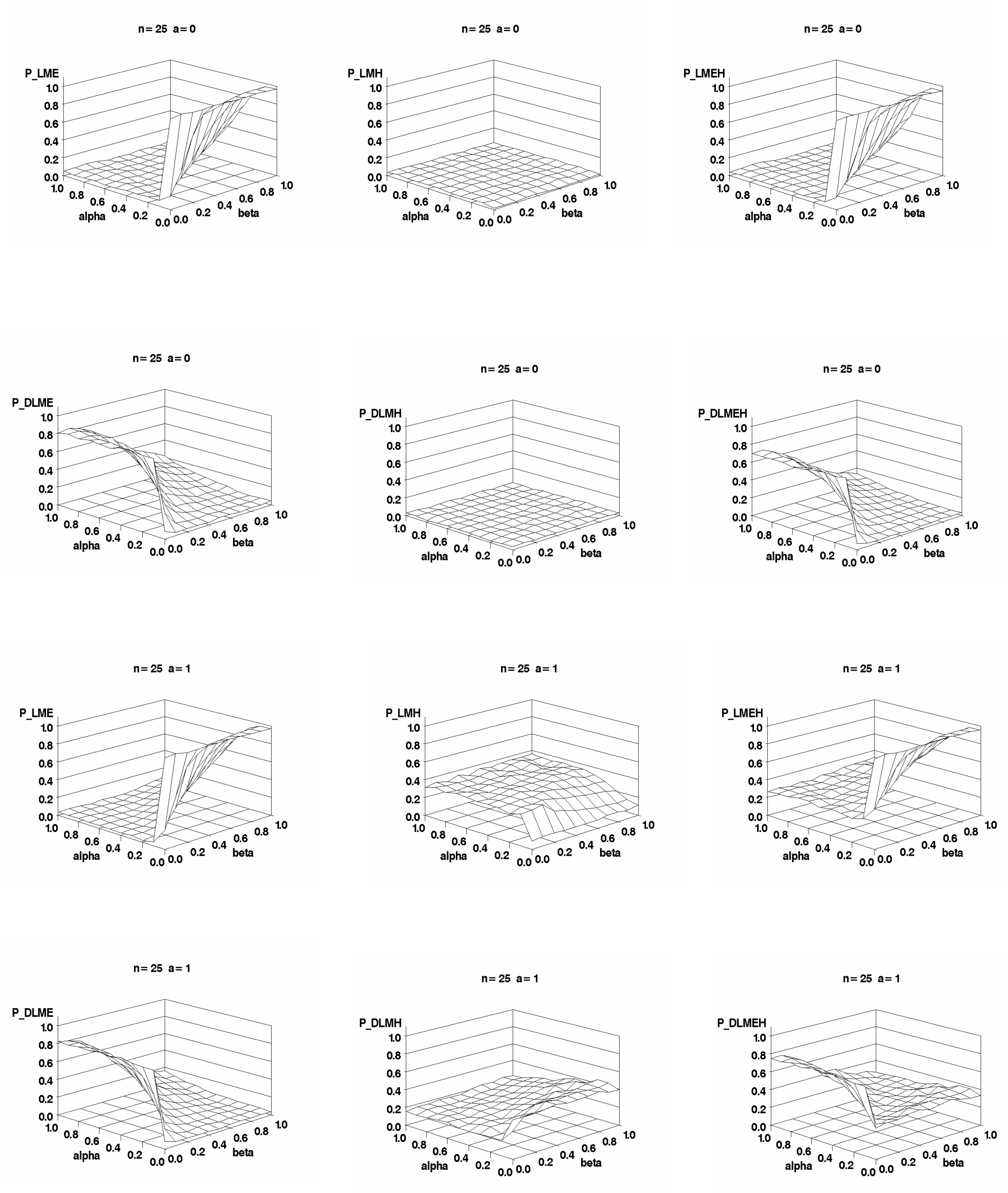

Figure 9. Monte Carlo results for cointegration study: Rook matrix – n=25.

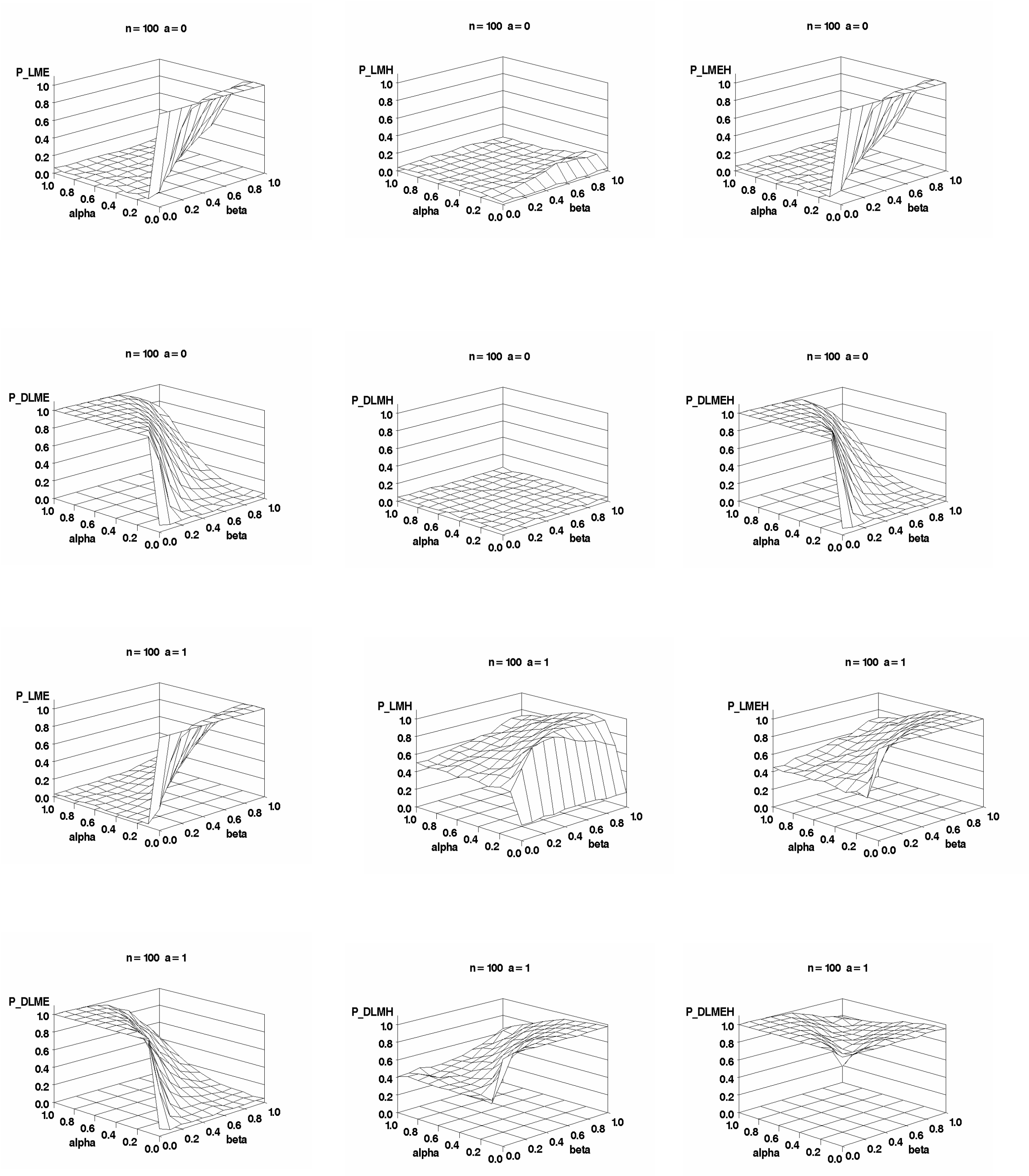

Figure 10. Monte Carlo results for cointegration study: Rook matrix – n=100.

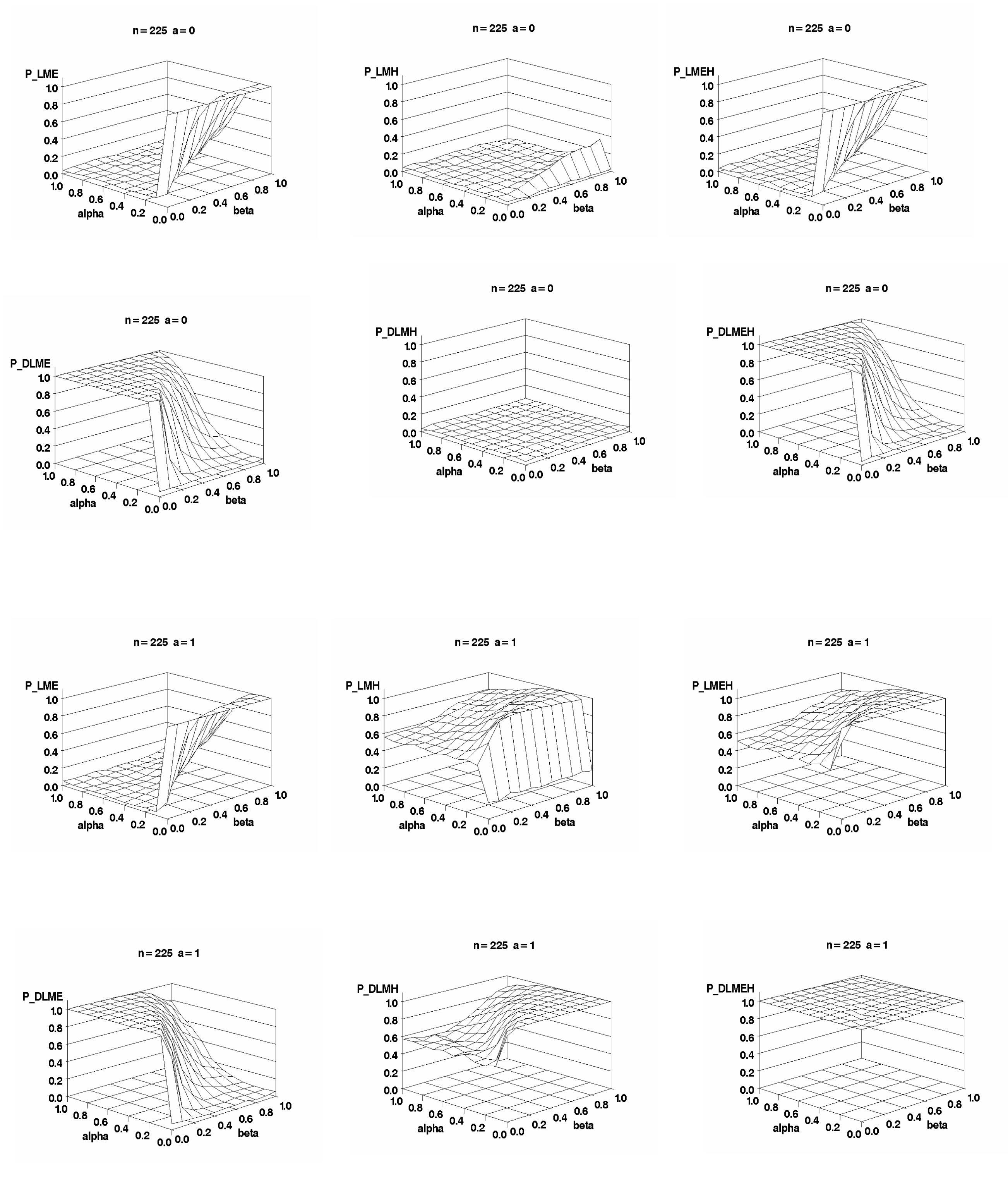

Figure 11. Monte Carlo results for cointegration study: Rook matrix – n=225.

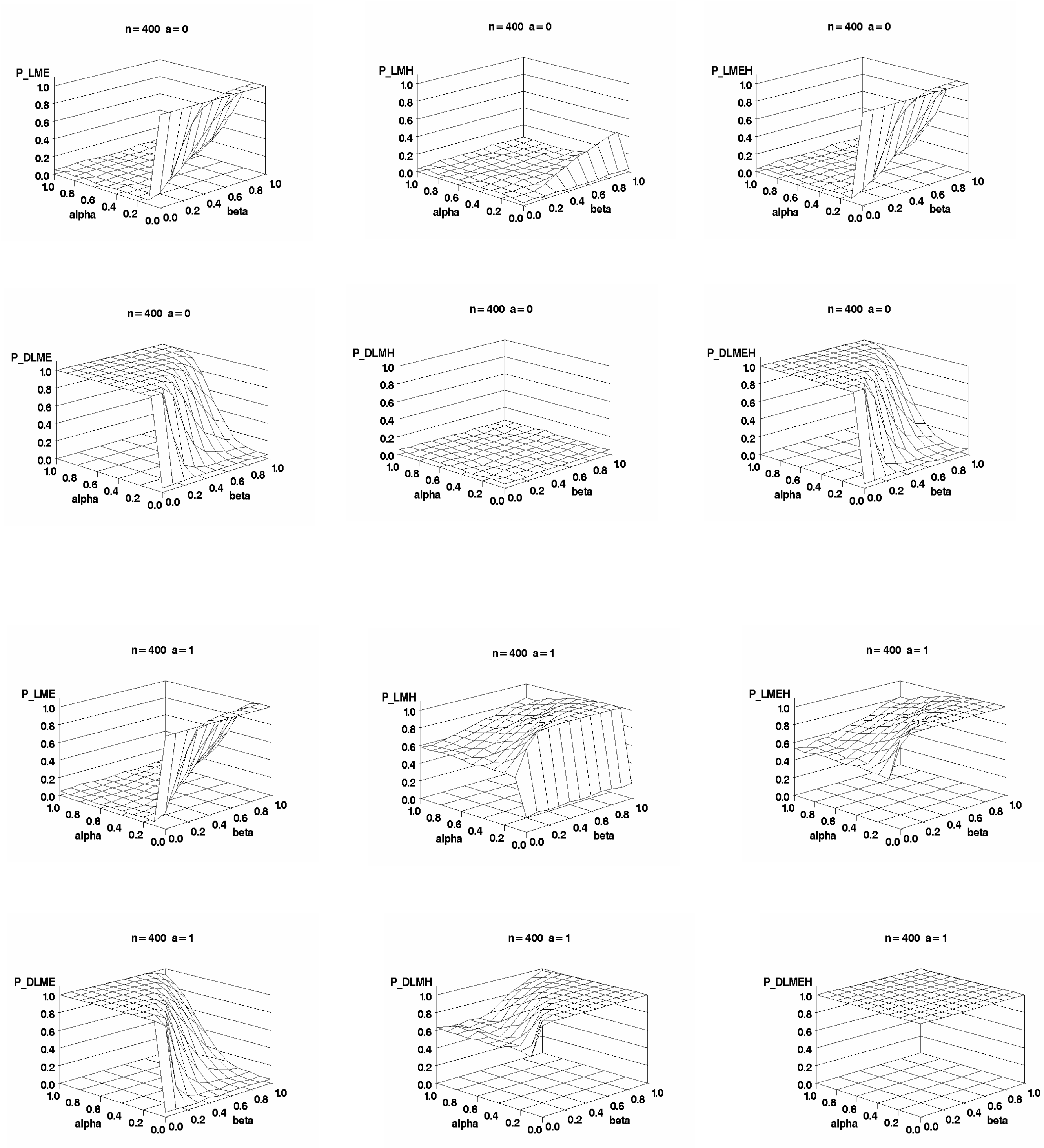

Figure 12. Monte Carlo results for cointegration study: Rook matrix – n=400.

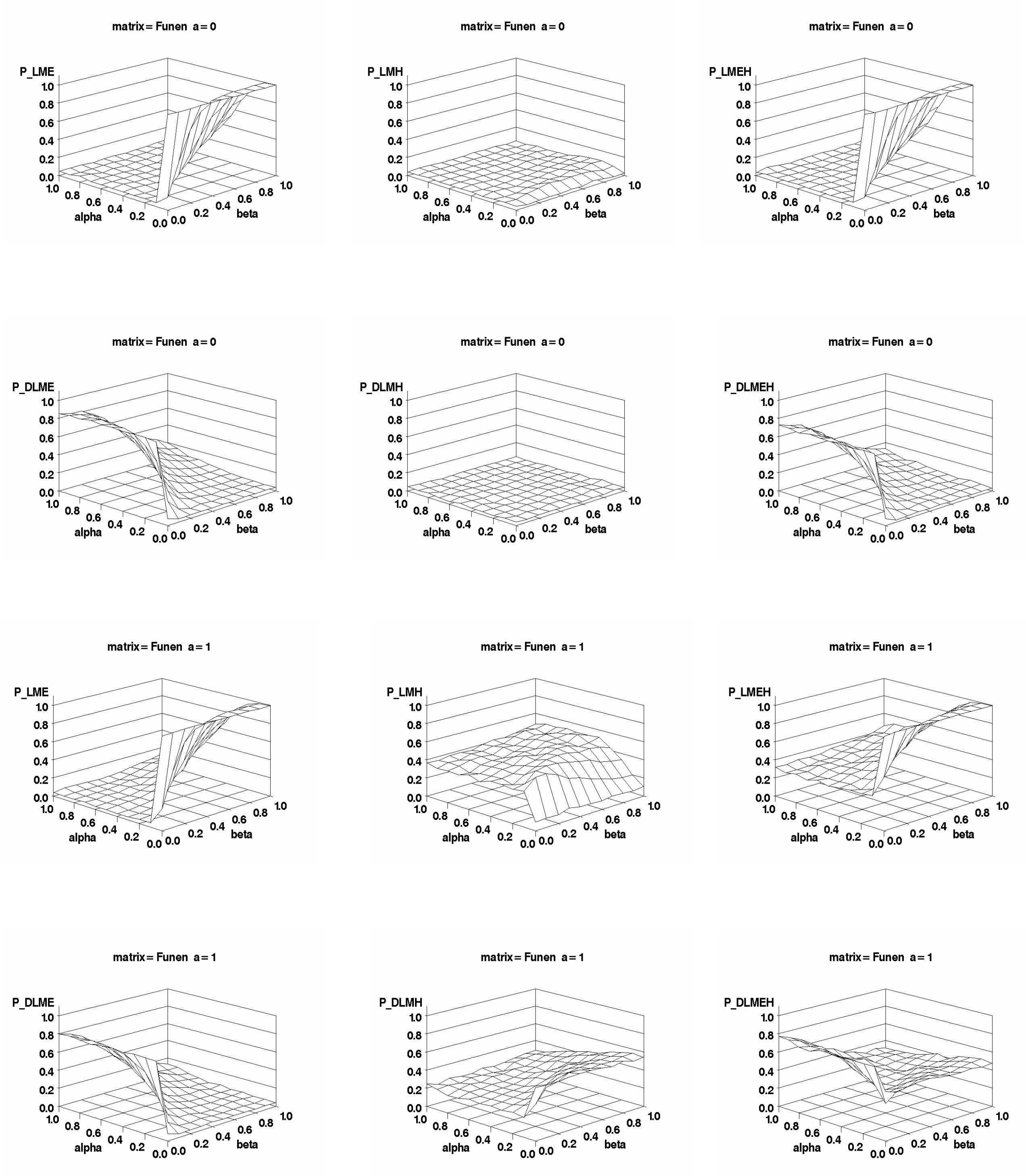

Figure 13. Monte Carlo results for cointegration study: Empirical matrix – Funen (n=36).

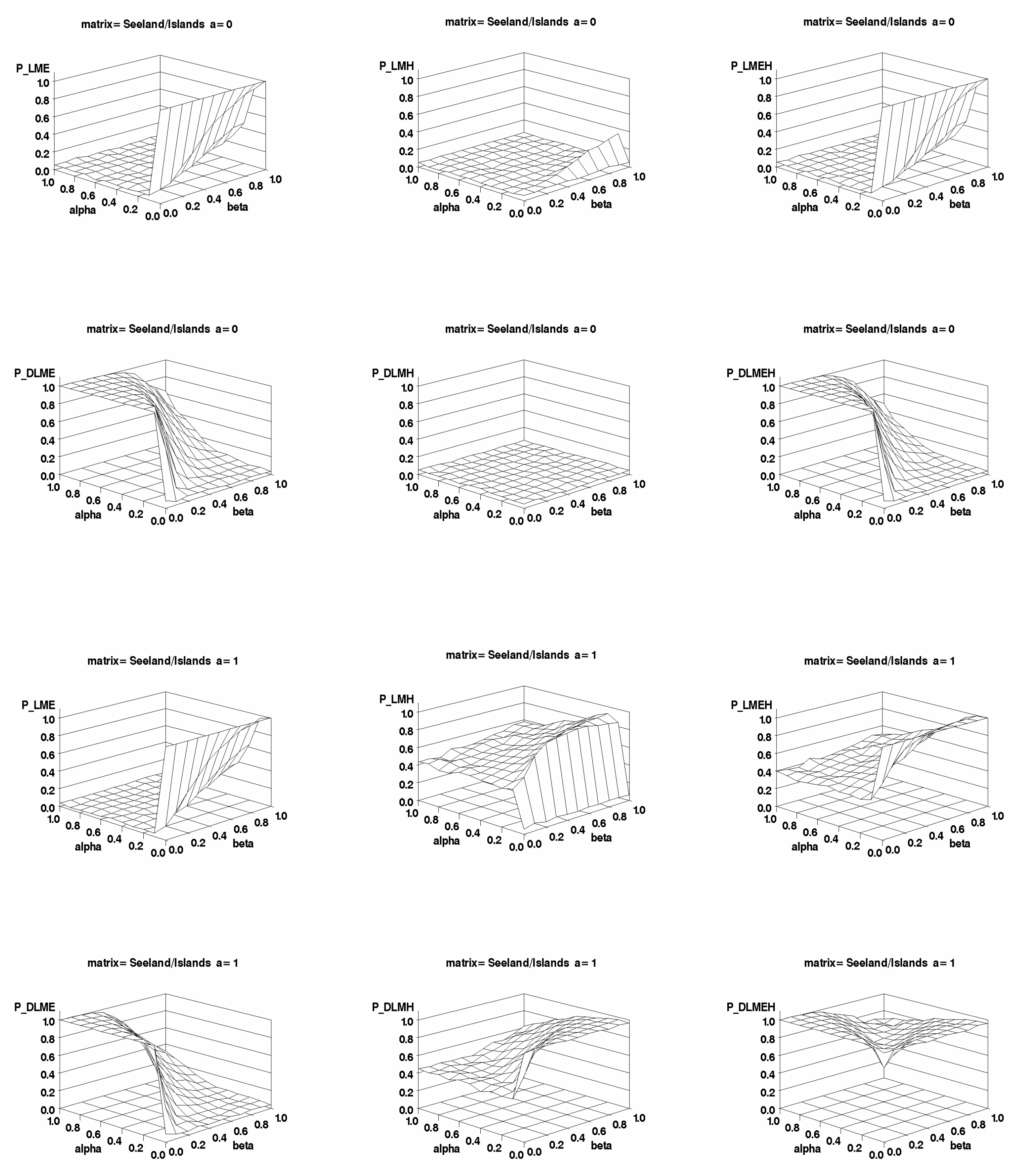

Figure 14. Monte Carlo results for cointegration study: Empirical matrix – Seeland (n=97).

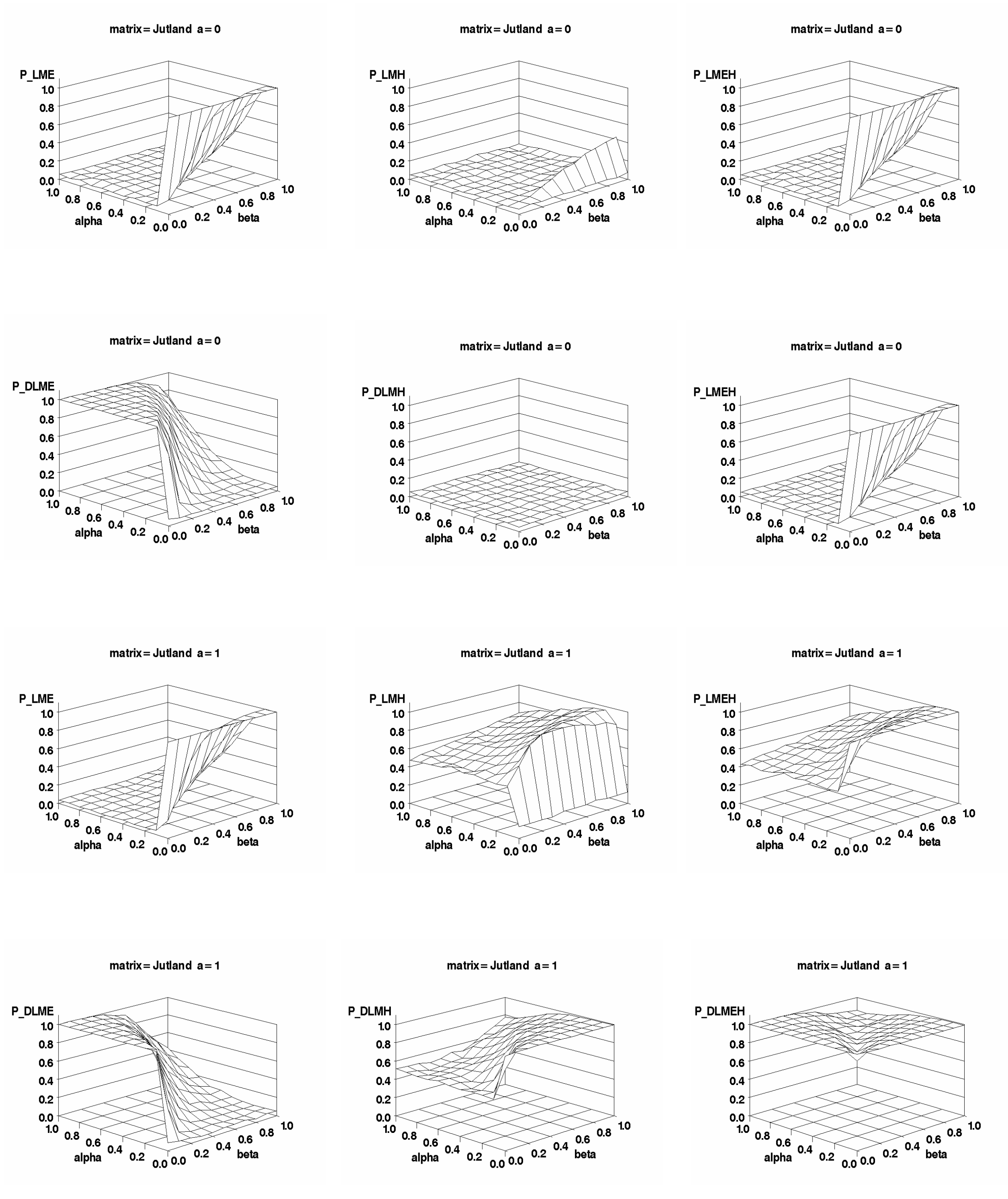

Figure 15. Monte Carlo results for cointegration study: Empirical matrix – Jutland (n=141).

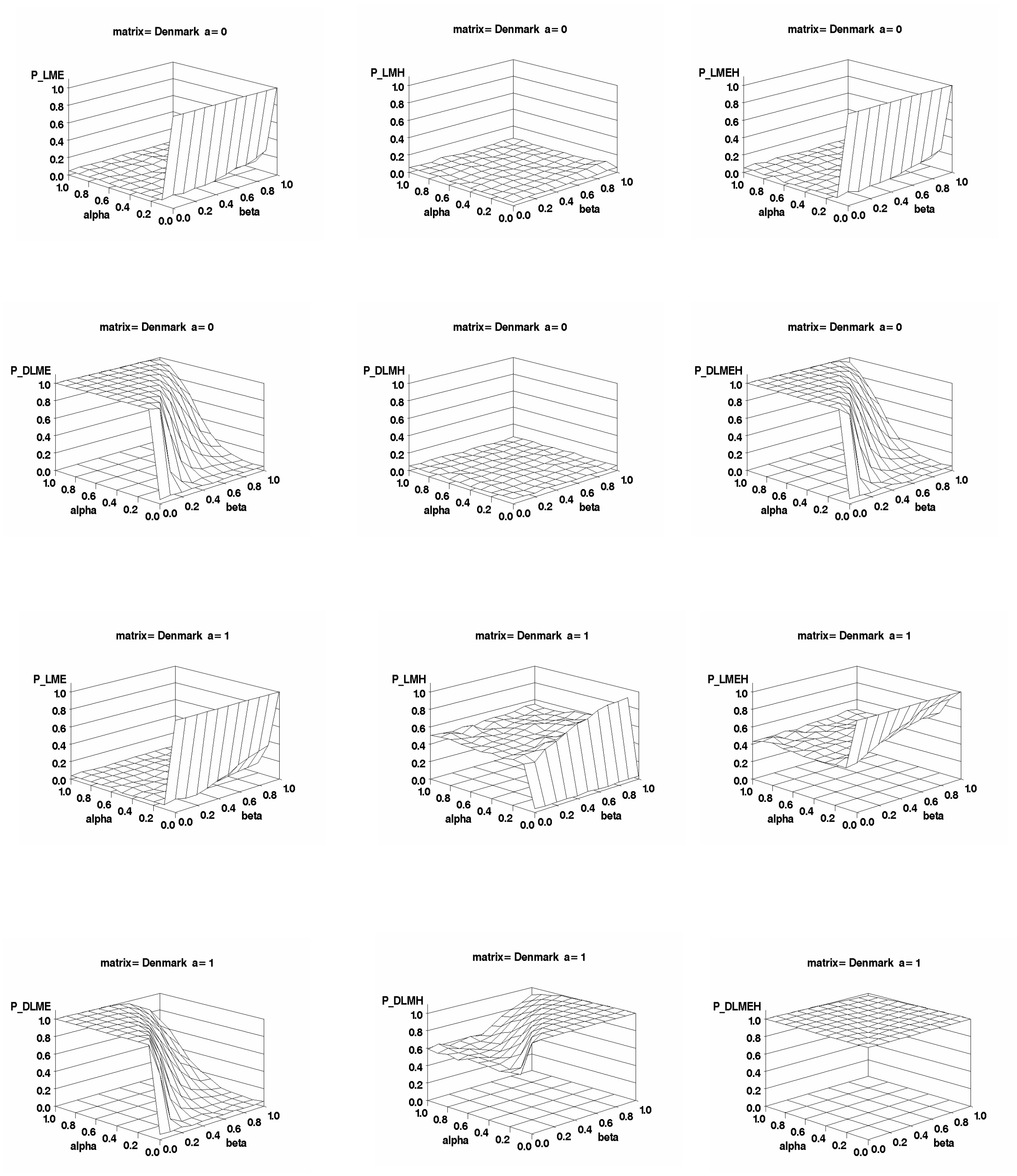

Figure 16. Monte Carlo results for cointegration study: Empirical matrix – Denmark (n=275).

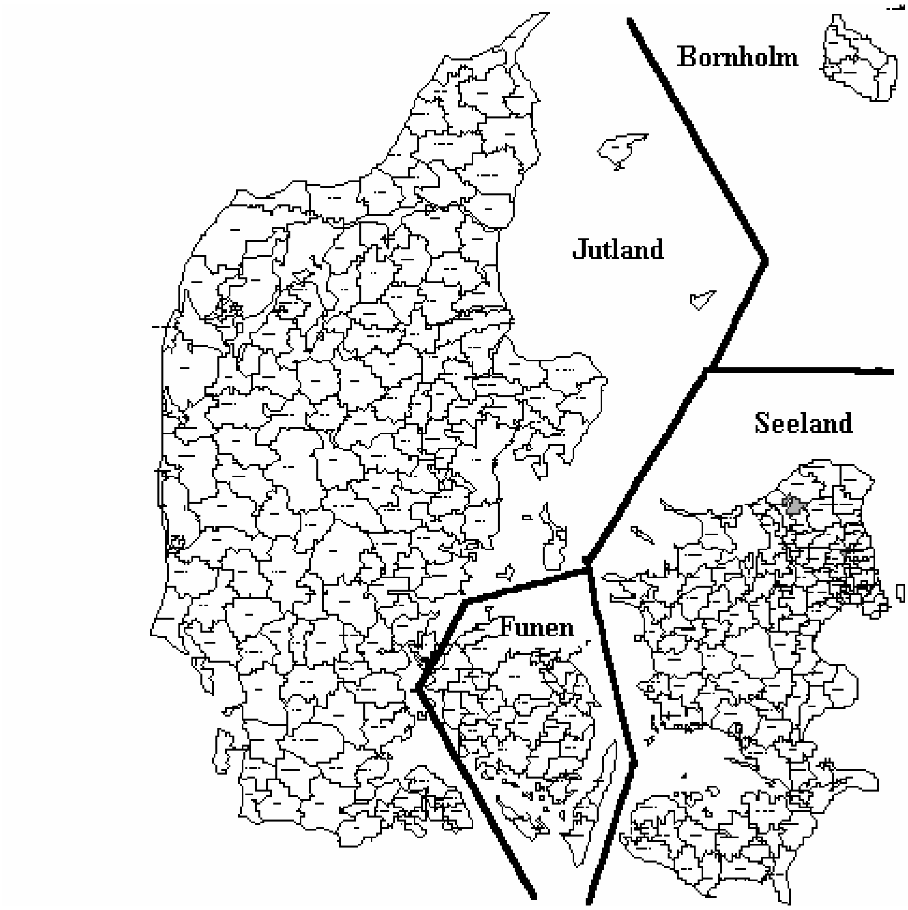

Figure A.1. The *n*=275 Danish municipalities