

İSTATİSTİK

JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION
TÜRK İSTATİSTİK DERNEĞİ DERGİSİ

ISSN: 1300-4077



- Characterizations motivated by the nexus between convolution and size biasing for exponential variables** 88
B.C. Arnold and J.A. Villasenor
- Pólya-Aeppli process of order k of the second type with an application** 98
S. Chukova, M. Lazarova and L. Minkova
- Fusion of geometric and texture features for side-view face recognition using SVM** 108
S.M. Jiddah, M. Abushakra and K. Yurtkan
- Robust outer product of gradients tests for testing spatial dependence** 120
O. Doğan, S. Taşpınar and B. Güloğlu

SUPPORTED AND PRINTED BY



İZMİR UNIVERSITY OF ECONOMICS

<https://dergipark.org.tr/en/pub/ijtsa>
**AVAILABLE
ONLINE**

İSTATİSTİK
JOURNAL OF THE TURKISH
STATISTICAL ASSOCIATION

ISSN: 1300-4077

<https://dergipark.org.tr/tr/pub/ijtsa>

EDITOR IN CHIEF

Ismihan Bayramoglu (I. Bairamov), Izmir University of Economics, Turkey E-mail:
ismihan.bayramoglu@ieu.edu.tr

FORMER EDITOR

Ömer L. Gebizlioglu, Kadir Has University, Turkey (Volume 1-3)

EDITORS

Serkan Eryilmaz, Atilim University, Turkey

Fatih Tank, Ankara University, Turkey

Gözde Yazgı Tütüncü, Izmir University of Economics, Turkey

EDITORIAL ASSISTANTS

Cihangir Kan, Xi'an Jiaotong-Liverpool University, China

Aysegül Erem, Cyprus International University, Nicosia, North Cyprus

TECHNICAL EDITOR ASSISTANT

Gökçe Özaltun, İzmir University of Economics, Turkey

ADVISORY EDITORS

F. Akdeniz, Çukurova University, Turkey

B. C. Arnold, University of California, United States
O. Ayhan, Middle East Technical University, Turkey

N. Balakrishnan, McMaster University, Hamilton, Ontario, Canada

E. Castillo, University of Castilla-La Mancha, Spain

Ch. A. Charalambides, University of Athens, Greece

J. C. Fu, University of Manitoba, Canada

H. Nagaraja, The Ohio State University, United States

V. B. Nevzorov, St. Petersburg State University, Russian Federation

J. S. Huang, University of Guelph, Ontario, Canada

G. D. Lin, Institute of Statistical Science, Academia Sinica, Taiwan

ASSOCIATE EDITORS

T. Mahmood, King Fahd University of Petroleum and Minerals, Saudi Arabia
V. Anisimov, University of Glasgow, Scotland, United Kingdom
M. Asadi, University of Isfahan, Iran, Islamic Republic Of
E. Cramer, RWTH Aachen University, Germany
J. Dhaene, Katholieke Universiteit Leuven, Belgium
U. Gurler, Bilkent University, Turkey
M. Kateri, RWTH Aachen University, Germany
N. Kolev, Department of Statistics, University of São Paulo, São Paulo, Brazil
M. Koutras, University of Piraeus, Greece
D. Kundu, Indian Institute of Technology, Kanpur, India
C. D. Lai, Massey University, New Zealand
W. Y. Wendy Lou, University of Toronto, Canada
J. Navarro, Facultad de Matematicas University of Murcia, Spain
O. Ozturk, The Ohio State University, Columbus, OH, United States
J. M. Sarabia, University of Cantabria, Spain
A. Stepanov, Immanuel Kant Baltic Federal University, Russia
R. Soyer, The George Washington University, United States
N. Papadatos, University of Athens, Greece

PRESIDENT OF THE TURKISH STATISTICAL ASSOCIATION

A. Apaydın, Ankara University, Turkey

SUPPORTED AND PRINTED BY



Izmir University of Economics Sakarya Cad. No:156 Balcova - Izmir/TURKEY

Phone: +90 (232) 488-8139 Fax: + 90 (232) 279-2626 www.ieu.edu.tr

AIM AND SCOPE

İSTATİSTİK, *Journal of the Turkish Statistical Association* is a refereed journal which publishes papers containing original contributions to probability, statistics and the interface of them with other disciplines where prosperity of scientific thought emerges through innovation, vitality and communication on interdisciplinary grounds. The Journal is published four-monthly. The media of Journal is English. The main areas of publication are probability and stochastic processes, theory of statistics, applied statistics, statistical computing and simulation, and interdisciplinary applications in social, demographic, physical, medical, biological, agricultural studies, engineering, computer science, management science, econometrics, etc.

In addition, the journal contains original research reports, authoritative review papers, discussed papers and occasional special issues or relevant conference proceedings.

ABSTRACTED/INDEXED: İSTATİSTİK, Journal of the Turkish Statistical Association is indexed in MathSciNet, Zentralblatt MATH and EBSCO.

SUBMISSION OF MANUSCRIPTS:

Authors should submit their papers electronically by using online submission system (<https://dergipark.org.tr/pub/ijtsa>).

As part of the submission process, authors are required to check their submission's compliance with all of the following items. Submissions that do not adhere to these guidelines may be returned to authors.

1. The submission has not been previously published, or is being considered for publication in another journal.
2. The submission file is in Portable Document Format (PDF) prepared in LaTeX (TEX) document file format.
3. Where available, URLs for the references have been provided.
4. The text is single-spaced; uses a 12-point font; employs italics, rather than underlining (except with URL addresses); and all illustrations, figures, and tables are placed within the text at the appropriate points, rather than at the end.
5. References should be listed in alphabetical order and should be in the following formats:

Cai, T. and Low, M. (2005). Non-quadratic estimators of a quadratic functional. *The Annals of Statistics*, 33, 2930-2956.

Meyer, Y. (1992). *Wavelets and Operators*. Cambridge University Press, Cambridge.

Cox, D. (1969). Some sampling problems in technology. In *New Developments in Survey Sampling* (N.L. Johnson and H. Smith, Jr., eds.). Wiley, New York, 506-527.

CHARACTERIZATIONS MOTIVATED BY THE NEXUS BETWEEN CONVOLUTION AND SIZE BIASING FOR EXPONENTIAL VARIABLES

Barry C. Arnold

Department of Statistics,
University of California,
Riverside, USA

Jose A. Villasenor*

Department of Statistics,
Colegio de Postgraduados,
Montecillo, Mexico

Abstract: For a continuous density $f(x)$ with support on the real interval $(0, \infty)$ and finite mean μ , its size biased density is defined to be of the form $(x/\mu)f(x)$. It is well known that for exponential variables, the convolution of two copies of the density yields the size biased form. This is the basis of the so-called inspection paradox. We verify that this agreement between size biasing and convolution actually characterizes the exponential distribution. We next consider the case in which the addition of one more term in a sum of independent identically distributed (i.i.d.) positive random variables also coincides with size biasing. Some related conjectures are also introduced. We then consider the problem of characterizing the class of all pairs of densities that can be called size-bias convolution pairs in the sense that their convolution is just a size biased version of one of them. We then consider discrete analogs to the size bias convolution results. It turns out that matters are more easily dealt with in the case of non-negative integer valued variables. Related geometric and Poisson characterizations are provided. Next, denote the sum of n i.i.d non-negative integer valued random variables $\{X_i\}$, $i = 1, 2, \dots$ by S_n . We verify that the ratio of the densities of S_{n_1} and S_{n_2} determines the distribution of the X 's. The absolutely continuous version of this result, though judged to be plausible, can only be conjectured at this time.

Key words: Continuous density, Size biased density, Convolution, Non negative integer valued variables, Exponential distribution, Size-bias convolution pairs.

1. Introduction

If we consider the convolution of two identical exponential distributions, the resulting density is just a size biased version of the exponential density involved in the convolution. This observation, discussed below in Section 2, provides a characterization of the exponential density. This characterization is so simply verified that it seems inevitable that it must have been proved in some earlier paper, but we have not been able to find a reference. In fact, in the exponential case, we can observe that the n -fold convolution of the exponential distribution also produces a weighted version of the common density of the convolutants. This too will be shown to be a characteristic property of the exponential density. In Section 3, we investigate the problem of identifying all pairs of densities corresponding to positive random variables that can be called size-bias convolution pairs in the sense that their convolution is just a size-biased version of one of the densities in the pair. If we turn to consider non-negative integer valued random variables, as we shall in Section 4, not unexpected parallel results involving geometric variables can be formulated. Analogous Poisson characterizations can also be identified. In fact, in Sub-section 4.3, very general characterization results will be proved for any distribution with support equal to the non-negative integers. Similar

*Corresponding author. E-mail address: jvillasr@colpos.mx

results are obtained for bounded non-negative integer valued random variables. It is tempting to propose that parallel general results will be available for general absolutely continuous positive random variables. This ambitious conjecture remains open, except for a few exponential cases.

2. Exponential characterizations

Suppose that X_1 and X_2 are i.i.d exponential random variables and that we define $S_2 = X_1 + X_2$. A comprehensive survey of distributional properties of exponential variables may be found in the volume dedicated to the exponential distribution that includes [1]. The density function of S_2 is that of a gamma distributed random variable and thus is a size biased version of the density of the X_i 's. That this is a characteristic property of the exponential distribution is readily confirmed as follows.

THEOREM 1. *Suppose that X_1 and X_2 are i.i.d positive absolutely continuous random variables and that $S_2 = X_1 + X_2$. Suppose that, for some positive c we have*

$$f_{S_2}(x) = cx f_{X_1}(x), \quad x > 0. \tag{2.1}$$

It follows that X_1 has an exponential distribution with mean $1/c$.

PROOF. Denote the Laplace transform of a positive random variable X by $L_X(s) = E(e^{-sX})$, $s > 0$. Chapter 13 of [3] will provide adequate discussion of Laplace transforms for our current purposes. Since $L_{S_2}(s) = [L_{X_1}(s)]^2$, we can conclude from (2.1) that

$$[L_{X_1}(s)]^2 = \int_0^\infty e^{-sx} cx f_{X_1}(x) dx = -c(d/ds)L_{X_1}(s).$$

However, this is a simple “variables-separable” differentiable equation with general solution of the form $L_{X_1}(s) = (k + s/c)^{-1}$. Since $L_{X_1}(0) = 1$ it follows that $k = 1$, and that X_1 has an exponential density with $\lambda = c$.

Several closely related characterizations can be formulated. A sample of five such possibilities follows. Proofs will be supplied for three of them, while the other two at present lack proofs and are labeled as (plausible) conjectures.

THEOREM 2. *Let $\{X_i\}_{i=1}^\infty$ be i.i.d. positive absolutely continuous random variables and for each n define $S_n = \sum_{i=1}^n X_i$. If, for some $c > 0$ and some positive integer k , we have*

$$f_{S_{k+1}}(x) = cx f_{S_k}(x) \tag{2.2}$$

then X_1 has an exponential distribution with mean $1/ck$.

PROOF. Using Laplace transforms we may rewrite (2.2) in the form

$$L_{X_1}^{k+1}(s) = -c(d/ds)L_{X_1}^k(s) = -ckL_{X_1}^{k-1}(s)(d/ds)L_{X_1}(s)$$

Dividing both sides by $L_{X_1}^{k-1}(s)$ yields an equation identical to that encountered in the proof of Theorem 1 with c replaced by ck . It follows that X_1 has an exponential density with $\lambda = ck$.

In the next items we will use the standard notation for a convolution of two densities, f_1 and f_2 , namely $f_1 * f_2$.

THEOREM 3. *If $f * g(x) = cx f(x)$ where g is an exponential density with intensity λ , then f is a gamma density.*

PROOF. Using Laplace transforms we have, by hypothesis,

$$L_f(s)(1 + s/\lambda)^{-1} = -c(d/ds)L_f(s).$$

The general solution to this differential equation is $L_f(s) = (1 + s/\lambda)^{-\alpha}$ where $\alpha > 0$, indicating that f is a gamma density.

Instead of using x as a weighting or biasing function, we may ask what happens when x is replaced by a power of x . For the case involving x^2 , we have the following result.

THEOREM 4. Suppose that X_1 and X_2 are i.i.d positive absolutely continuous random variables and that $S_2 = X_1 + X_2$. Suppose that, for some positive c we have

$$f_{S_2}(x) = cx^2 f_{X_1}(x), \quad x > 0. \tag{2.3}$$

Provided that $\text{var}(X_1) = (1/2)E^2(X_1)$, it follows that X_1 has a gamma distribution with shape parameter 2.

PROOF. First, note that it is readily verified that if $X_1 \sim \Gamma(2, \beta)$ then (2.3) holds. Suppose now that (2.3) holds. Evidently we must have $c = 1/\mu_2 = 1/E(X_1^2) < \infty$. Rewriting this in terms of $L(s)$, the Laplace transform of X_1 , we have

$$L''(s) = \mu_2 L^2(s). \tag{2.4}$$

Multiplying both sides of this equality by $2L'(s)$, we have

$$2L'(s)L''(s) = \frac{2\mu_2}{3}3L^2(s)L'(s)$$

Integrating over the interval $(0, t)$ and recalling that $L(0) = 1$ and $L'(0) = -\mu = -E(X_1)$, we have

$$[L'(t)]^2 - (-\mu)^2 = \frac{2\mu_2}{L} \int_0^t L^3(s) ds - \frac{2\mu_2}{3}t.$$

This will simplify when we apply the condition, stated in the hypothesis of the theorem, that $\text{var}(X_1) = (1/2)E^2(X_1)$, equivalently that $\mu_2 - (3/2)\mu^2 = 0$. Under this assumption, we have

$$[L'(t)]^2 = \frac{2\mu_2}{3}L^3(t).$$

However, from (2.4) we can write

$$L''(t)L(t) = \mu_2 L^3(t),$$

and consequently

$$[L'(t)]^2 = (2/3)L''(t)L(t).$$

we may rearrange this to obtain

$$\frac{3}{2} \frac{L'(t)}{L(t)} = \frac{-L''(t)}{-L'(t)}.$$

Integrating with respect to t over the interval $(0, s)$ yields

$$(3/2) \log L(s) = \log[-L'(s)] - \log \mu,$$

so that $-\mu = [L(s)]^{-3/2}L'(s)$. Integrating with respect to s over the interval $(0, t)$ produces

$$-\mu t = -2\{[L(t)]^{-(3/2)-1} - 1\},$$

so that $L(t) = (1 + \mu t/2)^{-2}$ and consequently $X_1 \sim \Gamma(2, \mu/2)$. Since in this expression, μ can take on any positive value, we conclude that, if (2.3) holds and if $\text{var}(X_1) = (1/2)E^2(X_1)$, then $X_1 \sim \Gamma(2/\beta)$ for some $\beta > 0$. After viewing this result, it is inevitable that one would consider the following unproved conjecture.

CONJECTURE 1. If $f * f(x) = cx^\alpha f(x)$ then, subject to regularity conditions involving certain moments of the density, f is a gamma density with shape parameter α .

Motivated by the fact that for the exponential case, convolution corresponds to size biasing (when $n_1 = 1$ and $n_2 = 2$ below) we have the following quite general conjecture.

CONJECTURE 2. Let $\{X_i\}_{i=1}^\infty$ be i.i.d. positive absolutely continuous random variables and for each n define $S_n = \sum_{i=1}^n X_i$. Claim: If for a fixed pair $1 \leq n_1 < n_2$ we have

$$f_{S_{n_2}}(x) = cx^{n_2 - n_1} f_{S_{n_1}}(x)$$

$\forall x$, and for some $c > 0$ then X_1 has an exponential distribution. Here too, it is likely that it will be necessary to invoke regularity conditions involving certain moments of the density of X_1 .

A proof or disproof of this last conjecture has eluded us. However, as we shall see below, better results are available in discrete cases. To introduce the discussion of non-negative integer valued random variables, we will first consider geometric and Poisson examples. But before leaving the absolutely continuous case, we will consider the general problem of identifying all cases, not just exponential and gamma cases, in which convolution is equivalent to size biasing.

3. Size-bias convolution pairs of densities

Throughout this section we will be dealing with density functions corresponding to positive absolutely continuous random variables which are positive throughout the interval $(0, \infty)$. If f is such a density, we will denote its Laplace transform by $L_f(s)$, thus

$$L_f(s) = \int_0^\infty e^{-sx} f(x) dx, \quad s \in (0, \infty).$$

We know that if f is a gamma density with shape parameter α and scale parameter $1/\lambda$ and if g is an exponential density with mean $1/\lambda$, then the convolution $f * g$ is again a gamma density. In fact we have the following situation:

$$f * g(x) = cx f(x), \quad x > 0. \tag{3.1}$$

for some positive c . In the particular case just mentioned we have $c = [\int_0^\infty x f(x) dx]^{-1}$.

If a pair of densities (f, g) satisfies equation (3.1), we will call it a size-bias-convolution (or sbc) pair. We have seen one example of an sbc pair. The name comes from the fact that when (3.1) holds then the convolution of f and g produces a size biased version of f .

Our goal is to characterize all valid sbc pairs of densities.

3.1. Laplace transforms corresponding to a size-bias-convolution pair of densities

If f and g are legitimate densities satisfying the sbc equation (3.1), then the corresponding Laplace transforms can readily be shown to be related by

$$L_g(s) = -c \left[\frac{d}{ds} \log L_f(s) \right], \tag{3.2}$$

or, equivalently

$$L_f(s) = \exp \left[-\frac{1}{c} \int_0^s L_g(t) dt \right]. \tag{3.3}$$

Note that in (3.2), in order that $L_g(0) = 1$ we must set $c = 1/\mu_f$ where μ_f is the necessarily finite mean of the density f .

Uniqueness considerations: For a given density f it is clear that if there exists a density g with (f, g) being an sbc pair. then g is the unique density with this property. Likewise, for a given density g it is clear that if there exists a density f with (f, g) being an sbc pair. then f is the unique density with this property.

3.2. Identifying sbc pairs

It might be hoped that every density f , with Laplace transform $L_f(s)$ will form part of an sbc pair. We may consider a candidate choice of g to be that density with a Laplace transform given by equation (3.2). This will be a solution provided that the expression on the right side of (3.2) is a valid Laplace transform, i.e., if it is completely monotone. Alternatively, it might be possible to recognize the right hand side of (3.2) as the Laplace transform of some well-known density. It is not at all obvious that the right hand side of (3.2) will always be completely monotone. We know it is for certain choices for f of the gamma form. But are there other cases ?

It turns out that the key result that allows us to resolve our identification problem is a characterization of infinite divisibility of distributions on $(0, \infty)$ provided by [3]. The result in question is as follows. The Laplace transform $L_f(s)$ corresponds to an infinitely divisible f if and only if the function $-\log(L_f(s))$ has a completely monotone derivative. However this is precisely the condition necessary for $L_g(s)$ defined by (3.2) to be a valid Laplace transform.

Note [2] made use of this characterization to verify the infinite divisibility of generalized inverse Gaussian densities.

We are able then to characterize the set of all valid sbc pairs (f, g) to consist of all pairs in which f is infinitely divisible and a corresponding g has its Laplace transform determined by (3.2).

EXAMPLE 1. If we choose f to correspond to a gamma density, which is infinitely divisible, then from (3.2) we can identify the choice of g to yield a valid sbc pair will be an exponential density. This observations (and analogous observations involving different sbc pairs) can be rephrased as characterizations of distributions. For example, we might wish to identify all possible densities g such that (f, g) constitutes an sbc pair with f being a gamma density. It follows that g must be an exponential density. This particular characterization appeared in [4], see also [5] and [6].

EXAMPLE 2. If f is taken to correspond to an inverse Gaussian distribution with parameters μ and λ denoted by $IG(\mu, \lambda)$, which is known to be infinitely divisible, then its Laplace transform is of the form

$$L_f(s) = \exp[(\lambda/\mu)(1 - \sqrt{1 + 2\mu^2\lambda^{-1}s})].$$

Differentiating with respect to s yields

$$L'_f(s) = \exp[(\lambda/\mu)(1 - \sqrt{1 + 2\mu^2\lambda^{-1}s})] \left\{ -\mu \left(1 + \frac{2\mu^2}{\lambda}s \right)^{-1/2} \right\}.$$

The corresponding density g to form an sbc pair is, from (3.2), one with Laplace transform given by

$$L_g(s) = -c \frac{d}{ds} \log L_f(s) = -c \frac{L'_f(s)}{L_f(s)} = c\mu \left(1 + \frac{2\mu^2}{\lambda}s \right)^{-1/2} = \left(1 + \frac{2\mu^2}{\lambda}s \right)^{-1/2}.$$

where c has been chosen equal to $1/\mu$ to ensure that $L_g(0) = 1$. Thus g is a gamma density with shape parameter $\alpha = 1/2$ and scale parameter $(2\mu^2/\lambda)$, i.e. corresponding to a random variable $Y = (\mu^2/\lambda)U$ where U has a chi-squared distribution with one degree of freedom.

EXAMPLE 3. If f is taken to correspond to a generalized inverse Gaussian distribution with parameters a, b and p denoted by $GIG(a, b, p)$, which is also known to be infinitely divisible, then its Laplace transform is of the form

$$L_f(s) \propto (a + 2s)^{-p/2} K_p(\sqrt{b(a + 2s)}),$$

where $K_p(u)$ is a modified Bessel function of the second kind. Differentiating with respect to s yields

$$L'_f(s) \propto (-p)(a + 2s)^{-p/2-1} K_p(\sqrt{b(a + 2s)}) + b(a + 2s)^{-(p+1)/2} K'_p(\sqrt{b(a + 2s)}).$$

The corresponding density g to form an sbc pair is, from (3.2), one with Laplace transform given by

$$L_g(s) = -c \frac{L_f'(s)}{L_f(s)}$$

where c is chosen to ensure that $L_g(0) = 1$. We can then recognize the density g as a linear combination of gamma densities.

4. Analogous discrete characterizations

We now turn to consider a selection of discrete characterizations suggested as natural analogs of the absolutely continuous results in Section 2.

4.1. Geometric characterizations

Parallel to the situation for exponential variables, in the geometric case, convolution essentially corresponds to size biasing. For a sample of size two, we have the following geometric characterization.

THEOREM 5. *Let X_1 and X_2 be i.i.d. non-negative integer valued random variables. Suppose that for each k and some $c > 0$*

$$P(X_1 + X_2 = k) = c(k + 1)P(X_1 = k), \tag{4.1}$$

it follows that X_1 has a geometric distribution.

PROOF. Let $P(s)$ be the probability generating function of X_1 . Then, from (4.1) we have

$$P^2(s) = csP'(s) + cP(s).$$

Rearranging this becomes:

$$\frac{P'(s)}{P(s)[P(s) - c]} = \frac{1}{cs}$$

i.e., writing dP/ds for $P'(s)$ and P for $P(s)$, as is usual in differential equations,

$$\frac{dP}{P(P - c)} = \frac{ds}{cs}.$$

Using partial fractions applied to $1/P(P - c)$ this is equivalent to

$$\frac{ds}{s} = \frac{dP}{P - c} - \frac{dP}{P}.$$

Integrating we get

$$\log(s) = \log(P - c) - \log(P) + k.$$

Thus

$$s = \tilde{k} \frac{P - c}{P}.$$

From this we have

$$P = \frac{c}{1 - \frac{s}{\tilde{k}}}.$$

However, we know that $P(0) = p_0$ and that $P(1) = 1$, so that finally we get

$$P(s) = \frac{p_0}{1 - (1 - p_0)s},$$

i.e., X_1 has a *geometric*(p_0) distribution.

A more general result is available. First note that if we have i.i.d. *geometric*(p) random variables, then for any $n \geq 2$, we have

$$P(S_n = k) = p \left(1 + \frac{k}{n-1} \right) P(S_{n-1} = k). \quad k = 0, 1, 2, \dots$$

where $S_n = \sum_{i=1}^n X_i$.

THEOREM 6. *Let $\{X_i\}_{i=1}^\infty$ be i.i.d. non-negative integer valued random variables. For each n define $S_n = \sum_{i=1}^n X_i$. If for a fixed integer $n \geq 2$ and for every k we have*

$$P(S_n = k) = c \left(1 + \frac{k}{n-1} \right) P(S_{n-1} = k). \quad k = 0, 1, 2, \dots \tag{4.2}$$

for some positive c , then the X_i 's have a common geometric distribution.

PROOF. Let $P(s)$ be the probability generating function of X_1 , so that the generating function of S_n is $[P(s)]^n$. From (4.2) we then have

$$[P(s)]^n = c[P(s)]^{n-1} + \frac{cs}{n-1} \frac{d}{ds} [P(s)]^{n-1} = c[P(s)]^{n-1} + cs[P(s)]^{n-2} P'(s).$$

Consequently we have

$$[P(s)]^2 = c[P(s)] + csP'(s).$$

But this is exactly the equation solved in the case $n = 2$ and we can conclude that

$$P(s) = \frac{p_0}{1 - (1 - p_0)s},$$

i.e., X_1 has a *geometric*(p_0) distribution.

In fact, in the geometric case, we are able to prove an even more general result which is parallel to the conjectured exponential characterization described in the previous section.

THEOREM 7. *Let $\{X_i\}_{i=1}^\infty$ be i.i.d. non-negative integer valued random variables. For each n define $S_n = \sum_{i=1}^n X_i$*

If for a fixed pair $1 \leq n_1 < n_2$ we have

$$P(S_{n_2} = k) = c \frac{(n_2 + k - 1)!}{(n_1 + k - 1)!} P(S_{n_1} = k) \tag{4.3}$$

$\forall k$, for some $c > 0$ then X_1 has a geometric distribution.

We will defer proving this result until Section 5, where will prove an even more general result as follows.

Consider a sequence of non-negative integer valued random variables $\{X_i\}_{i=1}^\infty$ and define

$$A(n, k) = \frac{P(\sum_{i=1}^n X_i = k)}{P(\sum_{i=1}^{n-1} X_i = k)}.$$

Claim : For any fixed $n \geq 2$, the sequence $A(n, k)$ determines the common distribution of the X_i 's.

We will illustrate a special case of this claim in the following Section where a Poisson sequence is considered.

4.2. Poisson characterizations

As usual, let the X_i 's be i.i.d. non-negative integer valued r.v.'s and for each n define $S_n = \sum_{i=1}^n X_i$. It is readily verified that if the X_i 's have a common *Poisson*(λ) distribution then for a fixed pair $1 \leq n_1 < n_2$ we have

$$P(S_{n_2} = k) = c \left(\frac{n_2}{n_1} \right)^k P(S_{n_1} = k), \tag{4.4}$$

$\forall k$, for $c = e^{-(n_2 - n_1)\lambda}$.

This observation leads to the following characterization of the Poisson distribution.

THEOREM 8. *Let the X_i 's be i.i.d. non-negative integer valued random variables and for each n define $S_n = \sum_{i=1}^n X_i$. If for a fixed pair $1 \leq n_1 < n_2$ we have*

$$P(S_{n_2} = k) = c \left(\frac{n_2}{n_1} \right)^k P(S_{n_1} = k) \tag{4.5}$$

$\forall k$, for some $c > 0$, i.e., if (4.4) holds for some $c > 0$, then X_1 has a *Poisson* distribution.

PROOF. It is tempting to try to resolve this issue by using probability generating functions. The generating function of X_1 , denoted by $P(s)$ must satisfy

$$P^{n_2}(s) = cP^{n_1}\left(\frac{n_2}{n_1}s\right).$$

However, it is not obvious how to solve this equation, even in the case in which $n_1 = 1$ and $n_2 = 2$.

We can make progress by considering equation(4.5) for a series of values of k . We will denote $P(X_1 = i)$ by p_i for $i = 0, 1, 2, \dots$. Next denote the ratio between p_1 and p_0 by λ . The case $k = 2$ of (4.5) simplifies to yield $p_2 = p_1^2 / (2p_0) = \lambda^2 p_0 / 2$. Next if we consider $k = 3$ we obtain an equation for p_3 as a function of p_0, p_1 and p_2 which can be solved to yield $p_3 = \lambda^3 p_0 / 3!$. We may then conclude that $p_i = \lambda^i p_0 / i!$ for every i by using an induction argument whereby we assume that for $j < i$ we have $p_j = \lambda^j p_0 / j!$ and, inserting these values in equation (4.5) for $k = i$, we may verify that $p_i = \lambda^i p_0 / i!$. The value of p_0 is then determined by the requirement that $\sum_{i=0}^{\infty} p_i = 1$. Thus we find $p_0 = e^{-\lambda}$ and confirm that X_1 has a *Poisson*(λ) distribution.

4.3. General characterizations of discrete distributions

Conjecture 2 in Section 2 was an instance in which for a sequence of i.i.d. X_i 's with sums defined by $S_n = \sum_{i=1}^n X_i$, it was felt to be plausible that the ratio of densities of S_{n_1} and S_{n_2} would determine the density of the X_i 's. In the absolutely continuous case, the conjecture remains open. However, progress can be made in the case in which the X_i 's are non-negative integer valued random variables.

Suppose that X_i^* 's are i.i.d. random variables with $P(X_i^* = k) = p_k^* > 0, k = 0, 1, 2, \dots$. The corresponding sums will be denoted by $S_n^* = \sum_{i=1}^n X_i^*$. For $1 \leq n_1 < n_2$ the corresponding ratio of densities of sums will be denoted by

$$A^*(n_1, n_2, k) = \frac{P(\sum_{i=1}^{n_2} X_i^* = k)}{P(\sum_{i=1}^{n_1} X_i^* = k)} \tag{4.6}$$

We claim that if another sequence $\{X_i\}_{i=1}^{\infty}$ has the same ratio of densities of sums as do the X_i^* 's and if $P(X_1 = 1) = P(X_1^* = 1)$ then $X_1 \stackrel{d}{=} X_1^*$.

THEOREM 9. Let $\{X_i\}_{i=1}^\infty$ be a sequence of i.i.d non-negative integer valued random variables with $P(X_1 = k) = p_k > 0$, $k = 0, 1, 2, \dots$ and with $p_1 = p_1^*$ as defined above. Suppose that for some pair n_1, n_2 with $1 \leq n_1 < n_2$ and every $k = 0, 1, 2, \dots$ we have

$$P(S_{n_2} = k) = A^*(n_1, n_2, k)P(S_{n_1} = k). \tag{4.7}$$

It follows that $X_1 \stackrel{d}{=} X_1^*$.

PROOF. Note that (4.7) holds for the S_n^* 's as well as for the S_n 's.

Consider the case in which $k = 0$, we have

$$p_0^{n_2} = P\left(\sum_{i=1}^{n_2} X_i = 0\right) = A^*(n_1, n_2, 0)P\left(\sum_{i=1}^{n_1} X_i = 0\right) = A^*(n_1, n_2, 0)p_0^{n_1},$$

so that p_0 is determined by $A^*(n_1, n_2, 0)$, and indeed $p_0 = p_0^*$.

Next consider $k = 1$, we have

$$n_2 p_1 p_0^{n_2-1} = P\left(\sum_{i=1}^{n_2} X_i = 1\right) = A^*(n_1, n_2, 1)P\left(\sum_{i=1}^{n_1} X_i = 1\right) = A^*(n_1, n_2, 1)n_1 p_1 p_0^{n_1-1}.$$

Note that p_1 cancels and is not determined by this equation. However, by one of our hypotheses, $p_1 = p_1^*$. Next consider $k = 2$,

$$\begin{aligned} [n_2 p_2 p_0^{n_2-1} + n_2(n_2 - 1)p_1^2 p_0^{n_2-2}] &= P\left(\sum_{i=1}^{n_2} X_i = 2\right) = A^*(n_1, n_2, 2)P\left(\sum_{i=1}^{n_1} X_i = 2\right) \\ &= A^*(n_1, n_2, 2)[n_2 p_2 p_0^{n_2-1} + n_2(n_2 - 1)p_1^2 p_0^{n_2-2}]. \end{aligned}$$

This gives p_2 as a linear function with coefficients that are functions of p_0 and p_1 . Thus p_2 is determined by $A^*(n_1, n_2, 2)$ and indeed $p_2 = p_2^*$.

Now each successive value of k will introduce a new p_k which will be a linear function with coefficients that are known functions of the preceding p_i 's. By an inductive argument the full sequence $p_0, p_1, p_2, p_3, \dots$ is determined by the sequence $A^*(n_1, n_2, k)$. Thus we conclude that $X_1 \stackrel{d}{=} X_1^*$.

COROLLARY 1. If the random variables, the X_i 's have bounded support say $0, 1, 2, \dots, M$, then for any fixed $1 \leq n_1 < n_2$, the finite sequence $\{A^*(n_1, n_2, k)\}_{k=0}^M$ determines the common distribution of the X_i 's.

PROOF. Just the same as in the theorem, except that we only need to consider values of k that are less than or equal to M .

5. Conclusions

Almost inevitably, when characterization results are presented to a statistical audience, the question of possible application of the results is raised. One strong argument for the study of characterizations is that they often enable researchers to realize interesting consequences of distributional assumptions that they routinely make. Characterizations often can be used to apply quick preliminary tests of certain distributional assumptions. In reliability settings, it will be of interest to know whether a size biased version of the lifetime distribution (the lifetime of an item in service) really behaves like a sum of two independent device lifetimes. If it doesn't, then a desirable assumption of exponentially distributed lifetimes must be set aside. If it does, then we can be more comfortable about the common distributional assumption.

References

- [1] Balakrishnan, N. and Basu, A.P. (1995). Basic distributional results and properties. in Balakrishnan, N. and Basu, A.P. (Eds.) *The Exponential Distribution: Theory, Methods and Applications*. (7-15). Gordon and Breach, Amsterdam.
- [2] Barndorff-Nielsen, O. and Halgreen, C. (1977). Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 38, 309-311.
- [3] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications. Vol. II. Second edition*, Wiley, New York.
- [4] Khattree, R. (1989). Characterization of inverse-Gaussian and gamma distributions through their length-biased distributions . *IEEE Transactions on Reliability*, 38, 610-611.
- [5] Pakes, A. G. and Khattree, R. (1992). Length-biasing characterization of laws and the moment problem. *Australian Journal of Statistics*, 34, 307-322.
- [6] Sen, A. and Khattree, R. (1996). Length biased distribution, equilibrium distribution and characterization of probability laws. *Journal of Applied Statistical Science*, 3, 239-252.

PÓLYA-AEPPLI PROCESS OF ORDER k OF THE SECOND TYPE WITH AN APPLICATION

Stefanka Chukova

School of Mathematics and Statistics,
Victoria University of Wellington of Wellington,
Wellington, NZ

Meglana Lazarova

Faculty of Applied Mathematics and Informatics,
Technical University of Sofia,
Sofia, Bulgaria

Leda Minkova*

Faculty of Mathematics and Informatics,
Sofia University,
Sofia, Bulgaria

Abstract: In this paper we propose and study the so called Pólya-Aeppli process of order k of the second type. Firstly, the process is defined using probability generating function, followed by its definition as a birth process. The distribution of the related counting process is presented by recursion formulae. The Pólya-Aeppli process of order k of the second type is considered within the framework of the risk process and corresponding probability of ruin is studied. Using simulation, some interesting results for the probability of ruin are obtained. Also, a comparison between the Pólya-Aeppli process of order k and Pólya-Aeppli process of order k of the second type is discussed.

Key words: Pólya-Aeppli distribution, Distributions of order k , Compound distributions, Ruin probability.

1. Introduction

Our motivation is based on the risk process, $\{X(t), t \geq 0\}$, and its use as a main tool in modeling of the surplus of an insurance company. In details, the risk process is given by

$$X(t) = ct - \sum_{i=1}^{N(t)} Z_i, \quad (1.1)$$

where c is a premium income per unit time, $N(t)$ is a counting process, $\{Z_i\}_{i=1}^{\infty}$ is a sequence of independent identically distributed, positive random variables, independent of $N(t)$, with Z_i representing the size of the i th claim. We assume that the individual claim amount has a continuous distribution with distribution function F , $F(0) = 0$, and mean value $\mu = EZ_1 < \infty$. In the classical risk model the process $N(t)$ is assumed to be a homogeneous Poisson process.

Let us consider the following stochastic process $N(t) = X_1 + \dots + X_{N_1(t)}$, where X_1, X_2, \dots are mutually independent random variables and also independent of the process $N_1(t)$.

It is well known that if the compounding random variable X has a discrete distribution with a finite support and truncated at 0, the random variable $N(t)$ has a distribution of order k , see for example [1], [3], [8] and [2]

Pólya-Aeppli distribution of order k was introduced by [10], and applied as a counting distribution in the risk model considered in [4]. There, the random variable $N_1(t)$ is Poisson distributed

* Corresponding author. E-mail address: leda@fmi.uni-sofia.bg

with parameter λ and X_i are truncated geometrically distributed with probability mass function (PMF) and probability generating function (PGF) given by

$$P(X = i) = \frac{1 - \rho}{1 - \rho^k} \rho^{i-1}, \quad i = 1, 2, \dots, k, \quad (1.2)$$

and

$$\psi_X(s) = \frac{(1 - \rho)s}{1 - \rho^k} \frac{1 - \rho^k s^k}{1 - \rho s},$$

where $k \geq 1$ is a fixed integer number and $\rho \in [0, 1)$. As a result, the above process $N(t)$ is called Pólya-Aeppli process of order k , denoted by $PA_k(\lambda, \rho)$.

In this paper we introduce another Pólya-Aeppli process of order k and call it Pólya-Aeppli process of order k of the second type, and denote it by $PA_{kII}(\lambda, \rho)$. The two Pólya-Aeppli processes of order k are different due to the difference of the compounding distributions included in their definitions. In the truncated geometric distribution in (1.2) the mass from $k + 1$ to infinity is uniformly distributed over the points $1, 2, \dots, k$. Here, we consider the case when the mass from $k + 1$ to infinity is clumped at point k .

So, what is the motivation for this new model $PA_{kII}(\lambda, \rho)$, and what is the difference between the current model and the model $PA_k(\lambda, \rho)$ in [4]? As mentioned above, our motivation for both modeling approaches comes from the risk process presented in (1.1). For the first model (the model in [4]) we consider (1.1) embedded in an usual operational environment for the insurance company, i.e., an environment without any major natural disasters or calamities, say storms, hurricanes, floods, earthquakes and so on. The only restriction we impose in this model is a limitation on the maximum possible number of simultaneous claims at any time, say k , which reasonably represents the reality faced by the insurance company in its everyday operations. Also, an environment with no major natural disasters, suggests no preference on any of the allowed integer numbers within $[1, k]$ of simultaneous claims, which means that the tail probability should be uniformly distributed over the domain $[1, k]$, leading to the truncated distribution considered in [4]. So, what is the motivation for the current model? The modeling in this study is for (1.1) embedded in an operational environment for the insurance company under the occurrence of a major natural disaster. As before, we preserve the limitation on the maximum possible number of simultaneous claims k at any time, but due to the external disastrous conditions we expect to have high number of simultaneous claims, so we place the tail probability at the maximum possible number of simultaneous claims to reflect the severity of the disaster, which leads to our model in this study.

The paper is organize as follows. In Section 2, we introduce the Pólya-Aeppli process of order k of second type. In Section 3, we define this process as a birth process. Some applications of this process to the risk model are given in Section 4. In Section 5, we present and discuss some simulation results related to Pólya-Aeppli process of order k of second type and Section 6 concludes this study.

2. Pólya-Aeppli process of order k of the second type

In this section we introduce the distribution of the Pólya-Aeppli process of order k of the second type as a compound Poisson distribution. The distribution of the compounding random variables X_i is given by the following PMF, which clumps the right tail of the distribution at point k :

$$P(X = i) = \begin{cases} (1 - \rho)\rho^{i-1}, & i = 1, 2, \dots, k - 1 \\ \rho^{i-1}, & i = k. \end{cases} \quad (2.1)$$

The corresponding PGF is given by

$$\psi_X(s) = \frac{(1 - \rho)s + (1 - s)(\rho s)^k}{1 - \rho s}. \quad (2.2)$$

Definition 1. The distribution defined by (2.1) or (2.2) is called a clumped geometric distribution with parameters k and $1 - \rho$, and it is denoted by $CGe(k, 1 - \rho)$.

In this case, the PGF of the $N(t)$ is given by

$$\psi_{N(t)}(s) = e^{-\lambda t \left(1 - \frac{(1-\rho)s + (1-s)(\rho s)^k}{1-\rho s}\right)}. \quad (2.3)$$

Definition 2. The process defined by the PGF in (2.3) is called a Pólya-Aeppli process of order k of the second type with parameters $\lambda > 0$ and $\rho \in [0, 1)$, and denoted by $PA_{kII}(\lambda, \rho)$.

If $k \rightarrow \infty$, the clumped geometric distribution approaches the usual geometric distribution with parameter $1 - \rho$.

If $k \rightarrow \infty$, the Pólya-Aeppli process of order k of second type, approaches the usual Pólya-Aeppli process, see [9] and [5]. If $\rho = 0$, it is the usual homogeneous Poisson process.

The mean and the variance functions of the $PA_{kII}(\lambda, \rho)$ are given by

$$EN(t) = \lambda t \frac{1 - \rho^k}{1 - \rho}$$

and

$$Var(N(t)) = \frac{\lambda t}{(1 - \rho)^2} [1 + \rho - (2k + 1)\rho^k + (2k - 1)\rho^{k+1}].$$

For the Fisher index, we obtain

$$FI(N(t)) = \frac{Var(N(t))}{E(N(t))} = \frac{1 + \rho}{1 - \rho} - 2k \frac{\rho^k}{1 - \rho^k}.$$

The Fisher index of the distribution of the Pólya-Aeppli process is equal to $\frac{1+\rho}{1-\rho}$, see [5]. Hence, the distribution of the counting process $PA_{kII}(\lambda, \rho)$ is underdispersed with respect to the distribution of the Pólya-Aeppli process.

Let us denote by $P_n(t) = P(N(t) = n)$, $n = 0, 1, \dots$. The following proposition gives an extension of the Panjer recursion formulas, see [11].

PROPOSITION 1. *The PMF of the $N(t) \sim PA_{kII}(\lambda, \rho)$ satisfies the following recursion formulae:*

$$P_1(t) = \lambda t(1 - \rho)P_0(t),$$

$$P_n(t) = (2\rho + \frac{\lambda t(1-\rho)-2\rho}{n})P_{n-1}(t) - (1 - \frac{2}{n})\rho^2 P_{n-2}(t), \quad n = 2, 3, \dots, k - 1$$

$$P_n(t) = (2\rho + \frac{\lambda t(1-\rho)-2\rho}{n})P_{n-1}(t) - (1 - \frac{2}{n})\rho^2 P_{n-2}(t) + \lambda t \rho^k \frac{k}{n} P_{n-k}(t)$$

$$- \lambda t \rho^k [\frac{k+1}{n} + \frac{k-1}{n} \rho] P_{n-k-1}(t) + \lambda t \rho^{k+1} \frac{k}{n} P_{n-k-2}(t), \quad n = k, k + 1, k + 2, \dots$$

and $P_{-1}(t) = P_{-2}(t) = 0$.

PROOF. Differentiation in (2.3) leads to

$$(1 - \rho s)^2 \frac{\partial}{\partial s} \psi_{N(t)}(s) = \lambda t [1 - \rho + k \rho^k s^{k-1} - \rho^k ((k + 1) + (k - 1)\rho) s^k + k \rho^{k+1} s^{k+1}] \psi_{N(t)}(s), \quad (2.4)$$

where $\psi_{N(t)}(s) = \sum_{n=0}^{\infty} P_n(t) s^n$ and $\frac{\partial}{\partial s} \psi_{N(t)}(s) = \sum_{n=0}^{\infty} (n+1) P_{n+1}(t) s^n$. The recursions are obtained by equating the coefficients of s^n on both sides of (2.4) for fixed $n = 0, 1, 2, \dots$

□

3. Pólya-Aeppli process of order k of the second type as a birth process

Suppose that $N(t) \sim PA_{kII}(\lambda, \rho)$. The properties of this process are specified by the following assumptions: For any small $h > 0$

$$P(N(t+h) = n \mid N(t) = m) = \begin{cases} 1 - \lambda h + o(h), & n = m, \\ (1 - \rho)\rho^{i-1}\lambda h + o(h), & n = m + i, \\ & i = 1, 2, \dots, k - 1, \\ \rho^{k-1}\lambda h + o(h), & n = m + k, \end{cases} \quad (3.1)$$

for every $m = 0, 1, \dots$, where $o(h) \rightarrow 0$ as $h \rightarrow 0$. Note that the assumptions imply that for $i = k + 1, k + 2, \dots$, $P(N(t+h) = m + i \mid N(t) = m) = o(h)$.

The above assumptions yield the following Kolmogorov forward equations:

$$\begin{cases} P'_0(t) = -\lambda P_0(t), \\ P'_n(t) = -\lambda P_n(t) + (1 - \rho)\lambda \sum_{j=1}^n \rho^{j-1} P_{n-j}(t), & n = 1, 2, \dots, k - 1, \\ P'_n(t) = -\lambda P_n(t) + (1 - \rho)\lambda \sum_{j=1}^{k-1} \rho^{j-1} P_{n-j}(t) + \lambda \rho^{k-1} P_{n-k}(t), & n = k, k + 1, \dots, \end{cases} \quad (3.2)$$

with the conditions

$$P_0(0) = 1 \quad \text{and} \quad P_n(0) = 0, \quad n = 1, 2, \dots \quad (3.3)$$

Multiplying the n th equation of (3.2) by s^n and summing for all $n = 0, 1, 2, \dots$ we get the following differential equation

$$\frac{\partial \Psi_{N(t)}(s)}{\partial t} = -\lambda[1 - \psi_X(s)]\Psi_{N(t)}(s). \quad (3.4)$$

The solution of (3.4) with the initial condition

$$\Psi_{N(1)}(s) = 1$$

is given by (2.3), which is the PGF of the distribution of $PA_{kII}(\lambda, \rho)$. This leads to the following definition for the Pólya-Aeppli process of order k of second type, namely:

Definitin 3. The process defined by (3.2) and (3.3) is the Pólya-Aeppli process of order k of second type.

4. Application to risk model

We consider the risk model (1.1), where $N(t) \sim PA_{kII}(\lambda, \rho)$. We call this model a Pólya-Aeppli of order k of second type risk model. In this case the relative safety loading θ is defined by

$$\theta = \frac{EX(t)}{E \sum_{i=1}^{N(t)} Z_i} = \frac{c(1 - \rho)}{\lambda\mu(1 - \rho^k)} - 1.$$

To ensure that $\theta > 0$, the premium income per unit time c should satisfy the following inequality

$$c > \frac{\lambda\mu(1 - \rho^k)}{1 - \rho}.$$

Denote by $\tau = \inf\{t : X(t) < -u\}$ the time to ruin of an insurance company having initial capital $u \geq 0$, and by

$$\Psi(u) = P(\tau < \infty) \quad (4.1)$$

the related ruin probability. Let $G(u, y)$ be the probability of the following event: {ruin occurs with initial capital u and deficit, immediately after ruin occurs, is at most y } with $u \geq 0$ and $y \geq 0$. Hence

$$G(u, y) = P(\tau < \infty, D \leq y), \quad (4.2)$$

where $D = |u + X(\tau)|$ is the deficit immediately after ruin occurs. Therefore

$$\lim_{y \rightarrow \infty} G(u, y) = \Psi(u). \quad (4.3)$$

Using the assumptions in (3.1), and for any small $h > 0$, we have

$$\begin{aligned} G(u, y) &= (1 - \lambda h) G(u + ch, y) + \\ &+ (1 - \rho) \lambda h \sum_{i=1}^{k-1} \rho^{i-1} \left[\int_0^{u+ch} G(u + ch - x, y) dF^{*i}(x) + (F^{*i}(u + ch + y) - F^{*i}(u + ch)) \right] + \\ &+ \rho^{k-1} \lambda h \left[\int_0^{u+ch} G(u + ch - x, y) dF^{*k}(x) + (F^{*k}(u + ch + y) - F^{*k}(u + ch)) \right] + o(h), \end{aligned} \quad (4.4)$$

where $F^{*i}(x)$, $i = 1, 2, \dots$ is the distribution function of $Z_1 + Z_2 + \dots + Z_i$.

Let us denote by

$$H(x) = (1 - \rho) \sum_{i=1}^{k-1} \rho^{i-1} F^{*i}(x) + \rho^{k-1} F^{*k}(x) \quad (4.5)$$

the non defective probability distribution function of the claims with

$$H(0) = 0, \quad H(\infty) = 1.$$

Rearranging the terms in (4.4) and letting $h \rightarrow 0$ we obtain the following differential equation

$$\frac{\partial G(u, y)}{\partial u} = \frac{\lambda}{c} \left[G(u, y) - \int_0^u G(u - x, y) dH(x) - [H(u + y) - H(u)] \right]. \quad (4.6)$$

In terms of the safety loading the equation has the form

$$\frac{\partial G(u, y)}{\partial u} = \frac{1 - \rho}{\mu(1 - \rho^k)} \frac{1}{1 + \theta} \left[G(u, y) - \int_0^u G(u - x, y) dH(x) - [H(u + y) - H(u)] \right]. \quad (4.7)$$

4.1. Ruin probability

THEOREM 1. *The probability of ruin $\Psi(u)$ satisfies the equation*

$$\frac{d\Psi(u)}{du} = \frac{\lambda}{c} \left[\Psi(u) - \int_0^u \Psi(u - x) dH(x) - [1 - H(u)] \right], \quad u \geq 0. \quad (4.8)$$

PROOF. The result follows from (4.6) and (4.3). □

Similarly to [4], we obtain the function $G(0, y)$ given by

$$G(0, y) = \frac{\lambda}{c} \int_0^y [1 - H(u)] du, \quad (4.9)$$

and for the ruin probability with no initial capital we obtain

$$\Psi(0) = \frac{\lambda \mu}{(1 - \rho)c} (1 - \rho^k). \quad (4.10)$$

4.2. Exponentially distributed claims

Let us consider the case of exponentially distributed claim sizes with mean μ , i.e. $F(x) = 1 - e^{-\frac{x}{\mu}}$, $x \geq 0$, $\mu > 0$. In this case, the function

$$F^{*i}(x) = 1 - \sum_{j=0}^{i-1} \frac{\left(\frac{x}{\mu}\right)^j}{j!} e^{-\frac{x}{\mu}}, \quad x \geq 0$$

is an Erlang distribution function. Then, the distribution function $H(x)$ in (4.5) is given by

$$H(x) = 1 - \sum_{i=0}^{k-1} \frac{\left(\frac{\rho x}{\mu}\right)^i}{i!} e^{-\frac{x}{\mu}}.$$

The density function $h(x)$ has the form

$$h(x) = \frac{1}{\mu} \left[(1 - \rho) \sum_{i=0}^{k-2} \frac{\left(\frac{\rho x}{\mu}\right)^i}{i!} + \frac{\left(\frac{\rho x}{\mu}\right)^{k-1}}{(k-1)!} \right] e^{-\frac{x}{\mu}}.$$

So, the initial condition (4.9) in the case of exponential distribution is

$$G(0, y) = \frac{\lambda \mu}{c} \sum_{i=0}^{k-1} \frac{\rho^i}{i!} \gamma(i+1, y/\mu),$$

where $\gamma(\alpha, x) = \int_0^x t^{\alpha-1} e^{-t} dt$ is the incomplete Gamma function.

5. Simulation

In what follows, we apply the simulation approach for calculating the probability of ruin suggested in [6] for the case of exponentially distributed claims with initial capital $u = 0$. We confirm the validity of our simulated results by matching them with the value of the ruin probability computed analytically using (4.10). Then, using our simulator, we provide results for the case of non-zero initial capital not only for exponentially distributed claims but also for claims with gamma and Weibull distributions. For a summary of the simulation approach for calculating the probability of ruin see [4]. All of our simulation results are based on 3 000 000 runs. Next, we provide some results regarding the probability of ruin for different scenarios of the claim distribution as well as the value of the initial capital.

5.1. Results

We consider the case of exponentially distributed claims and no initial capital $u = 0$. We verify the correctness of our simulator by comparing the results for the probability of ruin for fixed model parameters, produced in two different ways : (i) by the simulator, given in column “simulated”, and (ii) computed using (4.10) given in column “analytical”. These are given in Table 1.

λ	k	ρ	simulated Exp(1)	analytical Exp(1)
1.0	15	0.6	0.208117	0.208235
1.5	4	0.8	0.316531	0.316286
2.0	10	0.4	0.256365	0.256383
2.5	3	0.9	0.423526	0.423437
3.0	6	0.2	0.288426	0.288443

Table 1: Simulated and analytical Exp(1)

As it is easy to see, the “analytical“ and “simulated“ results are very close. So, we use our simulator, written in MATHEMATICA, to compute a reasonable approximation of the probability of ruin for non-exponentially distributed claims and non-zero initial capital ($u \neq 0$) and a summary of our results is given in subsection 5.1.1.

5.1.1. Case 1: Exponentially distributed claims

Here, we present some simulation results for the case of exponentially distributed claims with non-zero initial capital.

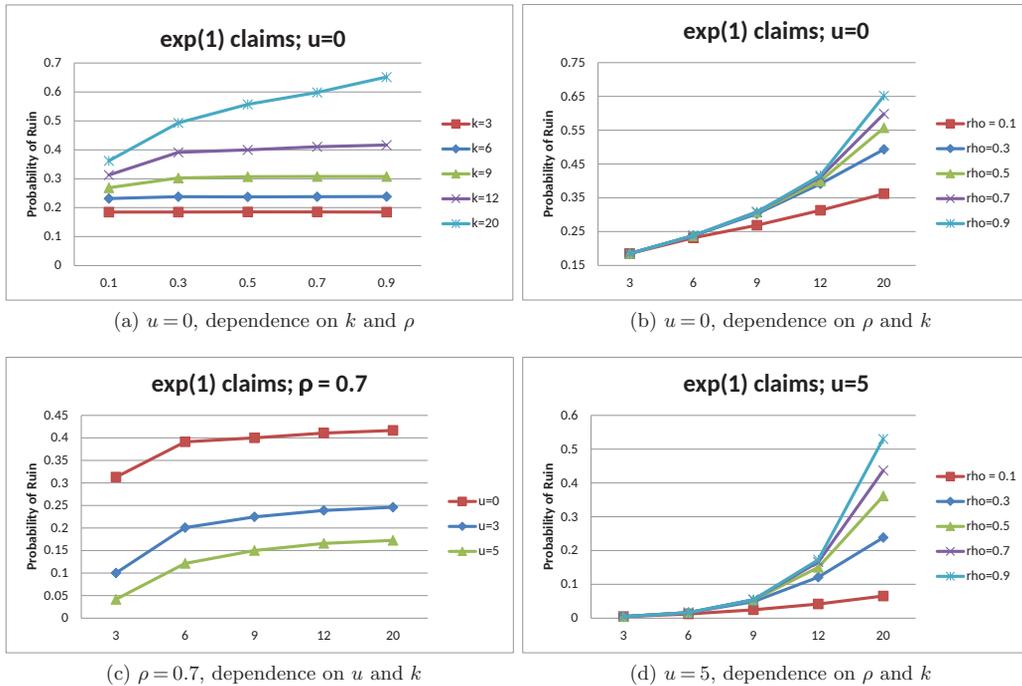


FIGURE 1. Probability of ruin: exponentially distributed claims

Comparing part(b) and part(d) of Figure 1, both with x-label k , it is easy to see that the probability of ruin is shifted downwards as the initial capital increases. If the initial capital is $u = 0$, the smallest values for the probability of ruin is just above 0.35 for $\rho = 0.1$, whereas the analogous value for $u = 5$ is just below 0.1. The depicted overall dependence on ρ , regardless of the value of the initial capital, is as expected, the probability of ruin increases as ρ increases. The overall trends depicted in part(a), with x-label ρ , and part(c), with x-label k , of Figure 1 also agree with our intuition. Namely, for a fixed value of ρ , the probability of ruin is higher for low values of the initial capital and it increases on k . It is worth to point out the sharp increase of the probability of ruin for large values of ρ and large k , as shown in part(a) of Figure 1.

5.1.2. Case 2: Gamma distributed claims

Next, we consider gamma distributed claims with parameters α and β , i.e., the density function of the claim sizes is

$$f(x) = \frac{x^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\frac{x}{\beta}}, \quad x \geq 0,$$

where $\Gamma(\alpha)$ is the Gamma function. Suppose that $\alpha = 2$ and $\beta = 0.5$. In this case the mean values of the claims are $EZ_i = \alpha\beta = 1$. We present results for different values of the model parameters u , k and ρ .

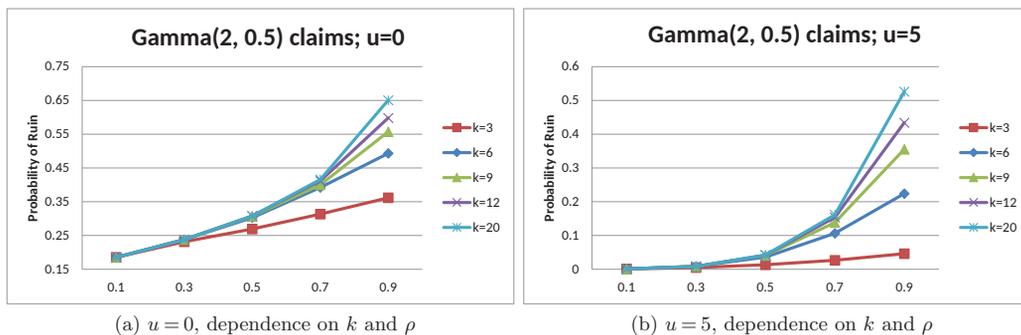


FIGURE 2. Probability of ruin: gamma distributed claims

The trends observed for the gamma distributed claims are similar to the one we have presented and discussed for the case of exponentially distributed claims in subsection 5.1.1. Here, in Figure 2, with x-label ρ , we depict the dependence of the probability of ruin from u , for similar ρ and k . Overall the probability of ruin for lower value of the capital u is higher, similar to what we have observed in the exponential case. In addition we see that for high values of u and ρ , k have a strong impact on the probability of ruin, e.g., see for $u = 0$ and $\rho = 0.9$, the range of the probability of ruin is approximately $(0.35, 0.65)$, whereas for $u = 5$ this range is much larger, approximately $(0.05, 0.53)$.

5.1.3. Case 3: Weibull distributed claims

Next, we consider the Weibull distribution with parameters $\alpha = 1.43552259$ and $\beta = 1.1013206$ distributed claims. Here α is the shape parameter and β is the scale parameter. The parameters of the Weibull and gamma distributions were selected so that the three claim size distributions considered in sections 5.1.1, 5.1.2 and 5.1.3 have the same expectation $\mu = 1$ and the Weibull and gamma claim sizes have the same variances.

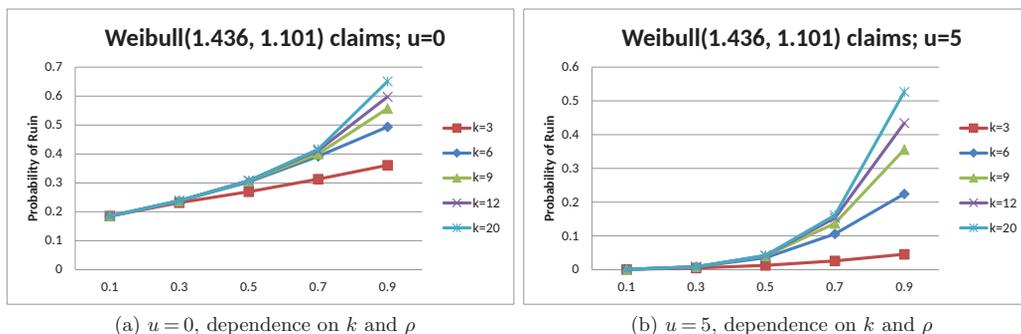


FIGURE 3. Probability of ruin: Weibull distributed claims

We were quite surprised to see that the behavior of the probability of ruin under Weibull distributed claims, part(a) and part(b) in Figure 3, with x-label ρ , mimics quite closely the behavior of this probability for gamma distributed claims. So, then the natural question is: under a risk model based on the Pólya-Aeppli process of order k , are the mean value and the variance of the claim distribution what determines the probability of ruin, i.e., the actual form of the claim size distribution does not have an effect on the probability of ruin. Interestingly, similar observations were made in [4]. Again, observing these results is a good motivation for future research because at this point we are not able to answer this question.

5.2. Comparison between M1 and M2

For brevity we will refer to the current model as M2 and to $PA_k(\lambda, \rho)$ from [4] as M1. Here we provide a brief comparison between the probabilities of ruin for the two models.

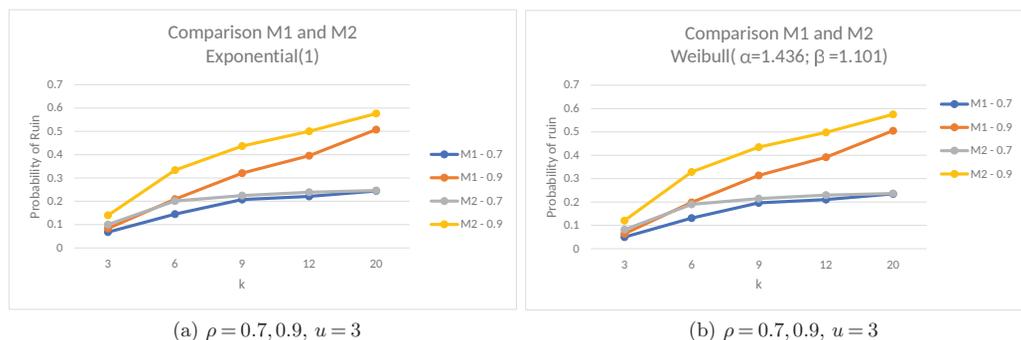


FIGURE 4. Probability of ruin: comparison between M1 and M2

In part(a) and part(b) in Figure 4, we fix the value of the parameter $u = 3$, and illustrate the dependence of the probability of ruin for M1 and M2 for two different values of $\rho = 0.7, 0.9$. Again, the probability of ruin for M1 and M2 is similar for the selected exponential and Weibull claim size distributions. The probability of ruin is an increasing function of k and its value is shifted upwards for higher values of parameter ρ . As expected, the probability of ruin for M2 is higher than for M1 and this is exactly what we expect to observe as an outcome for the insurance company at the time of severe natural disaster. Having model $PA_{kII}(\lambda, \rho)$ in place provides a reasonable theoretical background for the company to plan accordingly for natural calamities. From the observations above a natural question arises: are there any condition on the mean and the variance of the claim size distribution that will guaranty the satisfaction of some inequalities on the related ruin probabilities. These inequalities will be very useful in the sense that, even at the time of calamity, the probability of ruin would not exceed a known value. Again, further numerical and theoretical studies are needed to gain some insight on this question.

6. Conclusions

In the present study we have defined and studied the Pólya-Aeppli process of order k of second type as a compound Poisson process with clumped geometric compounding distribution with success probability equal to $1 - \rho > 0$. We have discussed some possible application of this process in risk theory. We have studied the probability of ruin for the related risk model and have derived an exact expression for the ruin probability in the particular case of zero initial capital. Also, we have adopted a simulation approach, given in [6] for our particular model. Using this simulation

approach we have provided results for general cases of the model, such as non-exponential claim distribution and non-zero initial capital. The simulation results have opened for discussion several very interesting questions related to the probability of ruin for Pólya-Aeppli of order k second type risk model. Also, a motivation for $PA_k(\lambda, \rho)$ studied in [4] and $PA_{kII}(\lambda, \rho)$ is outlined and a comparison between these two models is discussed.

Acknowledgment

The research of the first author was partially supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (C) Grant Number 18K04621, Japan. The second and the third authors are partially supported by Grant DN12/11/20.dec.2017 of the Ministry of Education and Science of Bulgaria.

References

- [1] Aki, S., Kuboku, H. and Hirano, K. (1984). On discrete distributions of order k . *Annals of the Institute of Statistical Mathematics*, 36, Part A, 431-440.
- [2] Balakrishnan, N. and Koutras, M.V. (2002). *Runs and Scans with Applications*. Wiley Series in Probability and Statistics.
- [3] Charalambides, Ch. (1986). On discrete distributions of order k . *Annals of the Institute of Statistical Mathematics*, 38, Part A, 557-568.
- [4] Chukova, S. and Minkova, L.D. (2015). Pólya-Aeppli of order k risk model. *Communications in Statistics - Simulation and Computation*, 44, 551-564.
- [5] Chukova, S. and Minkova, L.D. (2013). Characterization of the Pólya-Aeppli process. *Stochastic Analysis and Applications*, 31, 590-599.
- [6] Dufresne, F. and Gerber, H.U. (1989). Three methods to calculate the probability of ruin. *Astin Bulletin*, 19(1), 71-90.
- [7] Gerber, H.U. and Shiu, E.S.W. (1998). On the time value of ruin. *North American Actuarial Journal*, 2, 48-72.
- [8] Hirano, K. (1986). Some properties of the distributions of order k . in *Fibonacci Numbers and Their Applications*, A. N. Philippou et al.(eds), 43-53.
- [9] Minkova, L.D. (2004). The Pólya-Aeppli process and ruin problems. *Journal of Applied Mathematics and Stochastic Analysis*, 3, 221-234.
- [10] Minkova, L.D. (2010). Pólya-Aeppli distribution of order k . *Communications in Statistics - Theory and Methods*, 39, 408-415.
- [11] Panjer, H. (1981). Recursive evaluation of a family of compound discrete distributions. *Astin Bulletin*, 12, 22-26.

FUSION OF GEOMETRIC AND TEXTURE FEATURES FOR SIDE-VIEW FACE RECOGNITION USING SVM

Salman Mohammed Jiddah*, Main Abushakra and Kamil Yurtkan

Department of Computer Engineering,
Cyprus International University,
99258, Nicosia, Cyprus

Abstract: Biometric recognition systems have been getting a lot of attention in both academia and the industrial sector, one of such aspects of biometrics attracting interest is side-view face recognition, the side-view of the face is known to hold unique biometric information of subjects. This study embarks on contributing to the research of side-view face biometrics by proposing the fusion of geometric and texture features of the side-view face. Local Binary Pattern (LBP) was used for the extraction of texture features and the application of Laplacian filter was used for the extraction of geometric features, both features were tested in side-view face recognition individually before fusion of the two features in order to observe and note the effect of fusing the two features has on the performance of side-view face recognition, the experiments carried out in the proposed recognition system utilized Support Vector Machine (SVM) for classification, the training of the system was done using the histograms of the texture and geometric features extracted and labelled for every individual subject in the dataset. All experiments were done on the National Cheng Kung University (NCKU) faces dataset.

Key words: Side-view face recognition, Local binary pattern, Histogram fusion, SVM, Laplacian filter.

1. Introduction

It is clear that the processes of identification and verification has seen an evolution in the way these processes are being carried out, the traditional methods of these processes are slowly being replaced and being automated and integrated with biometric systems. As these systems grow and become sophisticated so is the growing need of security, biometric systems are being used in providing robust systems due to their growing sophistication and efficiency. Biometric systems use a process known as pattern recognition to carry out authentication or identification processes, biometric systems are mainly categorized based on the modalities used for the systems, which are physiological and behavioural modalities [2]. One of the most popular and well accepted biometric systems in both the research sector and the industrial sector is the face biometric systems, systems based on the face biometric modality use the human face for the pattern recognition process. Most facial biometric systems have been designed and developed for the frontal facial view, these kinds of systems require a relatively controlled scenario for an efficient recognition or identification process to be carried out, the control environment usually means a system is limited in its robustness, where a factor such as the viewing angle may pose a challenge to a recognition or identification system [12]. One of the most popular recognition systems for the human face recognition is the Viola-Jones technique of facial recognition systems, this system uses a process which involves the extraction of a specific feature from a detected facial image input which has to be a full-frontal view image, the specified feature is extracted through what is known as a window which is automatically scaled based on the size of the detected face, figure 1 below shows the steps of this algorithm through a typical recognition system flow chart.

* Corresponding author. E-mail address: salman.m.jiddah@gmail.com

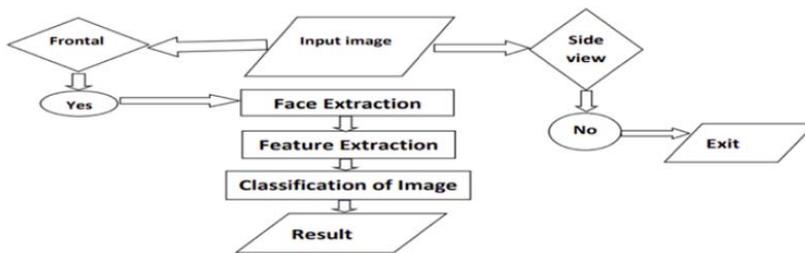


FIGURE 1. Typical Facial Recognition Flowchart[2]

As most systems and algorithms of facial recognition are designed and built to work with full frontal facial images, this makes the use of facial images with non-full-frontal pose images such as facial side view images in such systems challenging for the systems to recognize and authenticate [1]. Both two dimensional (2D) and three dimensional (3D) facial images have been used in the field of facial image recognition systems, and studies and reports have shown the use of 3D facial images yielding more performing systems when compared to 2D facial recognition systems, however 2D facial images recognition systems are also significantly well performing, and it is also noteworthy that 3D systems require a lot more computing resources than their 2D systems counterpart which require a lot less computing power to work efficiently [13]. There are several 2D face recognition feature analysis systems that have been prominently used over the years and some of them are as follows: eigenfaces approach and geometric features-based approach, this study attempts to utilize two different approaches to facial features in side view, which are the texture based feature and the geometric based feature, the two chosen features will be applied using Local Binary Patterns (LBP) and Laplacian filters for texture and geometric features respectively.

2. Literature Review

Studies in facial recognition have been significantly pushed by the progress of computational technology of the past few decades, these technological advancements have made it possible to model mechanisms in such a way that they could possibly be as good or even more robust than the human visual perception. Facial recognition has seen a huge rise in interest due to its efficiency and non-invasive characteristics of the modality, this makes it suitable in many kinds of system and even in systems which do not require the cooperation of a subject for a recognition or identification process to be performed on them. Facial recognition systems go as far back as 1964 where Woodrow Wilson Bledsoe conducted computerized facial recognition experiments with an objective of simply identifying a specific image from a large number of images, in this study Bledsoe reported challenges faced by the study which were as follows: variation in accuracy with respect to variation in inclination and facial poses of the subjects, the intensity/angle of lighting, and the variation in facial expressions of subjects [17]. In a study by [6], subject markers were utilized by an operator which were measured and compared by the system to perform the facial recognition in each input photograph image, despite the earlier start of studies on facial recognition systems, automated facial recognition systems only really started in a 1971 [10], the study used an automated system to automatically extract and analyse facial features which included measurements of the nose, chin, and eye region, the study used the Euclidean distance classifier [23]. Over the years there have been proposal and utilization of several methods for facial recognition systems, the classification of the existing methods of facial recognition systems is complex, however they can still be categorized into three major categories which are: local based methods, holistic based methods and hybrid-based methods of facial recognition.

Holistic methods according to [11] are the systems which implement a method which uses the entire facial regions in their recognition process, this method includes the eigenface and fisher facial recognition techniques, local based facial recognition methods are the methods of facial recognition which specify local characteristics from a facial image to analyse such as the region of the mouth, eyes or nose.

local based methods have been very successful in facial recognition systems especially when compared with holistic based methods of facial recognition systems because they have an advantage whereby they are not as sensitive to variations in lamination and poses of the facial images been analysed but its accuracy performance is highly dependent on the efficiency of the features extraction method applied. Hybrid based recognition methods are as the name implies, a combination of both local based methods and holistic based methods in an attempt to use the best of what both of them have to offer as a single method in a facial recognition system.

2.1. Side View Face Recognition

This reported side view face recognition was in the 1970s [18], where the study was carried out by profiling side view facial images' silhouettes and was used for recognition, the study reported an impressive performance of up to 90 percent accuracy while auto correlations and K nearest neighbour was used as the classification algorithm of the facial recognition system, the study used a total of ten subjects. Another study [20] used facial labelling of facial landmarks on side view facial images as shown in figure 2 below, they used this labelling to wrap and register the facial features which they then applied Principal Component Analysis (PCA), Local Binary Pattern (LBP), and Linear Discriminant Analysis (LDA), and compare them, the study showed LBP having the better performance of the three applied methods with a performance accuracy of 91.1%.



FIGURE 2. Manually labelled facial landmarks

Another study [15] carried out an innovative technique whereby both left and right-side view facial images of subjects were taken and used to detect and select the eye region of the subjects using an algorithm to create a single facial image from the two-side view facial images of subjects, the processed image is used to train the system after a median filter is applied to rid of possible noise, this study concluded the region of interest in side view facial images are the nose and eye regions of the facial images. [14] a study which successfully developed an algorithm to detect the eye brow region of facial images for both right and left side view face images made it possible to identify whether an input side view image was either left side or right side view face image, this enabled their facial recognition system to specify which facial images and their features to analyse in the automatic recognition process, the study also utilized local features for the recognition process of

the side view facial images, and the vectors of LBP and Grey Level Co-occurrence Matrix (GLCM) were formed. [16] carried out a study which investigated facial recognition with respect to facial expressions in a multi view facial image which was a study motivated with handling non full frontal face images in a face recognition system, the study utilized local descriptors in the face images in the form of histograms, they used LBP for feature extraction in the form of grid set uniform sampling which were a division of 46 sub blocks of the facial image, the multi pose variation of the subjects was handled through the use of viola jones for facial regions extraction from the input images, the pose variation were in 0° , 15° , 30° , 45° , 60° , 75° and 90° as can be seen for the respective angle of pose respectively in figure 3 below.



FIGURE 3. Subject face image in multiple pose variation [16]

The dataset used for this study had 4200 images, and the classification method used for the study in the experiments was Support Vector Machine (SVM).

2.2. Local Binary Pattern (LBP)

Local Binary Pattern (LBP) is an algorithm which is also an operator which is best described as a texture descriptor, LBP is used to extract and provide the texture information from the contours in an image of any kind of object. Unlike the visual perception of the human eye, LBP can distinguish between colours from the contours of an image, LBP is a very powerful descriptor because the texture information of an image is capable of giving information about the outline of an image [4], LBP descriptor was developed with working with images that are monoscriptal in as a feature, LBP uses eight closest pixel neighbours, where the LBP resulting pixel value depends on the value of the neighbouring pixels surrounding it, LBP also works by highlighting the edges of an image which in turn gives it a chance of obtaining better description of the texture of an image. LBP as an operator uses the following steps to carry out its function: it first starts by dividing the input image into n parts, however the most advised number of divisions of an image for an LBP operator is 16 because it gives efficiency for both accuracy and time taken to carry out the LBP operation on an image. LBP then uses a 3×3 mask with respect to the centre pixel of the mask, and then proceeds to apply the following formula as seen in formula 2.1 below.

$$LB(px_t - px_c) = \begin{cases} 1, & px_t \geq px_c \\ 0, & px_t < px_c \end{cases} \quad (2.1)$$

Where px_t is the pixel being analysed, and px_c is the centre pixel of the matrix, This is used to calculate and obtain new pixel values for the matrix in a manner where all neighbouring pixels are compared to the centre pixel, where the centre pixel is greater than the neighbouring pixel a binary value of 0 is assigned to the pixel as its new value else it is assigned 1 and the new pixel value [4], the new pixel value are then extracted as a vector of binary values which is then used to generate an LBP histogram from the acquired binary values, the total histograms gotten for an image are then concatenated into a spatially enhanced histogram as defined by formula 2.2 below.

$$H_{ij} = \sum_{x,y} I \{f_t(x,y) = i\} I \{(x,y) \in R_j\}, \quad i = 0, \dots, n-1, j = 0, \dots, m-1 \quad (2.2)$$

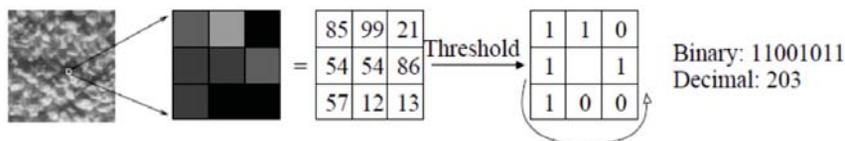


FIGURE 4. Basic LBP operator function on image

The entire LBP operator function on an image can be seen as described in figure 4 below

LBP as an operator has been used over the years in several studies and has also been proven to be a very powerful texture feature extractor, the use of LBP in facial recognition system can be seen in [4] where the authors achieved a performance rate of up to 90% accuracy in the facial recognition system in which they applied LBP to the FERET facial dataset. Another study which used LBP for side view face recognition is [21] where they applied LBP on the National Cheng Kung University (NCKU) dataset, they applied LBP on both left and right side face images of the dataset, after which they used a distance based classifier to gauge the performance of their recognition system, they achieved different accuracy performance for the right side images and left side images of the dataset, with 67% performance for the right side face images and 74.19% accuracy performance for the left side view images.

2.3. Support Vector Machine (SVM)

Support vector machine is a classification algorithm which is discriminative, SVM as a classifier has also seen applications in regression challenges, however the major function of the SVM classifier lies in its used a function which performs classification using hyperplanes, using an N-dimensional plane SVM is capable of distinctively identifying data, where the features of class type is denoted by N, and every class type is situated on a different side of the hyperplane as can be seen in figure 5 below. SVM also has multidimensional hyperspace functions, SVM has kernel functions which enables it to map classification regions within a space [5].

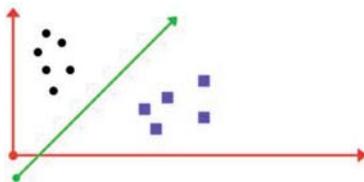


FIGURE 5. SVM hyperplane showing two distinct classes

2.3.1. SVM in Facial Recognition

Zhang et al. [22] carried out a facial expression recognition in which they applied SVM for classification in their experiments, they used a variation of SVM known as the fuzzy multi class SVM for their classification phase of their study and achieved an accuracy performance of 96.77%. another study which used SVM in their facial recognition system classification is a study by Richhariya and Gupta [7], this study used a variation of SVM known as the iterative universum twin SVM in which they used datapoints in their datasets which were not associated with any of the classes being trained to supervise the training of the classes data. Another study which used the SVM classifier in a facial recognition system study is study by Wang et al. [19] which they used in on

extracted LBP features from a facial image dataset, their system was a facial recognition system which operates in real-time which showed promising recognition accuracy performance. Julina and Sharmila [9] used the SVM classification algorithm on Histogram of Gradients (HOG) features which they extracted from the AT & T face dataset, this study took on the challenge of multi variation in pose and lighting with their study to which they achieved an impressive accuracy performance of up to 90.2 %.

2.4. Feature Fusion

There have been studies over the years which have carried on the researches to fuse multiple features of a dataset in an attempt to gain better performance than otherwise using a single of the multiple selected features in an experiment. A study by Santemiz et al. [18] carried out a study in which they used HOG and LBP features together and classified them using SVM after they have fused them using sum rule fusion, this study arrived at a performance accuracy of 89%. Another similar study is that of Chen et al. [3] which used HOG-TOP fusion on geometric warp features and acoustic features of a face image dataset. A study which used the same multiple feature fusion as proposed by this study is that of Jiddah and Yurtkan [8], however they used the Euclidean distance classifier on human ear dataset to which they also reported an improvement in accuracy when compared to using any of the two features individually in the facial recognition system

3. Methodology

This study has been proposed to fuse two features; geometric and texture features of side view face images, and using the SVM classifier for classification. Our methodology for this study seeks to use LBP to extract the texture features from our sideview facial images, and use Laplacian filters to extract the geometric features from the side view face images. The methodology of study follows an outline as follows: the NCKU dataset images were pre-processed to rid the image of redundant data, after which LBP and Laplacian filter were used to extract texture and geometric features of the images respectively, the histogram of the images were then extracted and fused together using histogram concatenation, the concatenated histograms were then classified using SVM and tested for recognition accuracy. Figure 6 below shows a general outline of the methodology using a block diagram.



FIGURE 6. Methodology block diagram

3.1. National Cheng Kung University (NCKU) dataset

The National Cheng Kung University (NCKU) dataset is a side view face dataset which is publicly available to the research community courtesy of the National Cheng University, the dataset contains images of a total of 90 subjects with 12 female subjects and 78 male subjects, with each subject having a total of 37 images making a total of 6660 images for all subjects, with a 50:50 ratio for left side face and right side face images, a sample of the NCKU dataset can be seen in figure 7 below. Each image is captured with a resolution of 640x480. For the purpose of the experiments carried out in this study the images were pre-processed to rid the original images of any redundant data in order to speed up the computational process and rid the images of noise, during the pre-processing the images were resized to 128x128 pixels, a sample of a processed image

can be seen in figure 8 below. Also, only images of the right side with pose variations that are actually side view face images were used for our experiments which are the 70 to 90 degrees pose variations, which brings our experiment images to a total of 450 images.

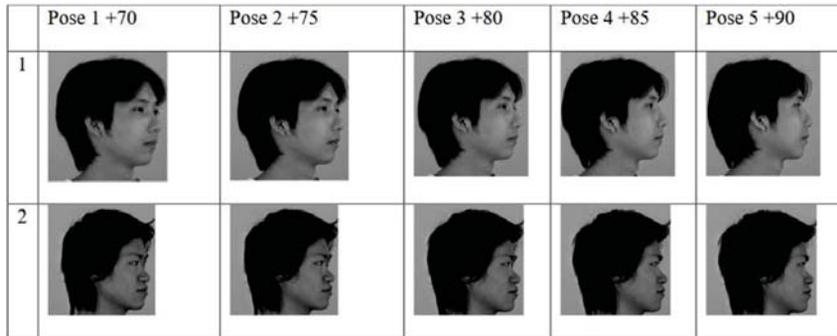


FIGURE 7. NCKU dataset sample images

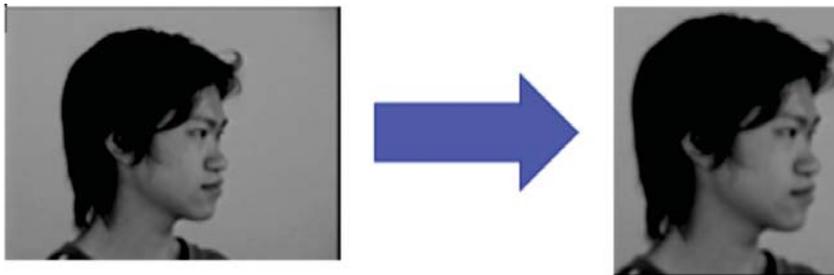


FIGURE 8. Original image (left), processed image (right)

3.2. Feature Extraction

This section explains the methodology used for the feature extraction procedure for both of our geometric and texture feature in order to achieve the proposed fusion of geometric and texture features in our side view facial image recognition system.

3.2.1. Texture Feature Extraction

The texture features used in the experiments of this study were extracted using the LBP algorithm with the LBP procedure explained in section 2.2 of this paper as proposed by Ojala et al. [16]. figure 9 below shows a sample image from our experiments after the LBP operator has been applied on an image to highlight the texture features of the image.



FIGURE 9. Greyscale Image (Left), Corresponding LBP (right)

3.2.2. Geometric Feature Extraction

The geometric features used in the course of the experiments of this study were extracted by the utilization of the Laplacian filter, Laplacian filter is a known image filter which is known to highlight and extract the geometric features of an image. Figure 10 below shows a sample image from our experiments after a Laplacian filter has been applied on a side view facial image and the geometric features of the image have been highlighted.

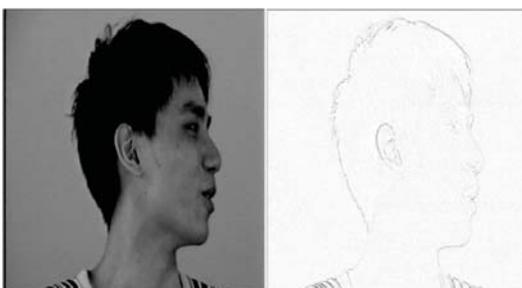


FIGURE 10. Greyscale image(left), Laplacian image (right)

3.3. Image Histograms

Part of the proposed methodology of this study is the use of image histograms of the extracted features from the image, it has been identified as one the efficient ways for the fusion of features and hence the choice to use the image histograms of the extracted features. These extracted image histograms were concatenated for the purpose of the fusion of the features, figure 11 below shows a sample image histogram from our study.

3.4. Classification

The classification phase of this study used SVM for its classification, the training and testing of the classification process used a ratio of 80:20 for training and testing respectively, which brings our training images to a total of 360 and testing images to 90. Also, k-fold cross validation was used to obtain the average of the accuracy performance of our proposed facial recognition system.

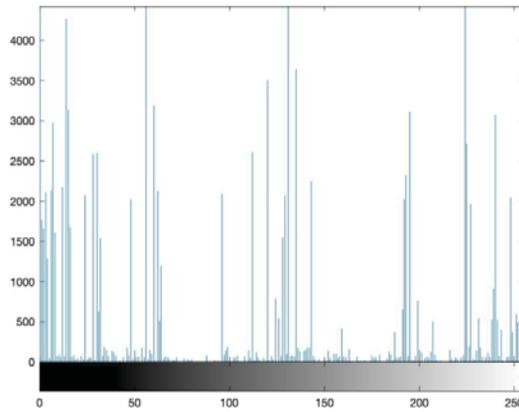


FIGURE 11. Sample Image Histogram

4. Results

The results of our experiments are reported in this section of this paper, the results are reported to show experiments carried out on individual features before their fusion, doing so enabled us to compare and contrast the performance of both geometric and texture features individually and when combined as one in the side view facial recognition system. Table 1 below shows the performance of the texture features when used alone in our proposed facial recognition system, as can be seen in the table after the k-fold cross validation the texture features achieved a recognition accuracy of 70%.

TABLE 1. Texture features accuracy performance

Iteration	Training Images	Testing Images	Accuracy (%)
1	360	90	77
2	360	90	74
3	360	90	67
4	360	90	81
5	360	90	55
Average			70

Table 2 below shows the accuracy performance of the geometric features used alone in the proposed side view face recognition system, the geometric feature significantly out performs the texture features with an increase of 15% accuracy in the performance, as can be seen after the k-fold cross validation the geometric feature achieved 85% accuracy performance in the proposed recognition system

Table 3 as shown below shows the performance of our proposed side view facial recognition system when both texture and geometric features are fused together and used as a single feature in the recognition system. The fusion of the features has shown promising results as expected, there was a significant increase in performance accuracy, the fusion of the features outperforms both texture and geometric features when used individually, the fusion of both features achieved an impressive 90% accuracy performance.

TABLE 2. Geometric features accuracy performance

Iteration	Training Images	Testing Images	Accuracy (%)
1	360	90	95
2	360	90	94
3	360	90	87
4	360	90	70
5	360	90	83
Average			85

TABLE 3. Geometric features accuracy performance

Iteration	Training Images	Testing Images	Accuracy (%)
1	360	90	96
2	360	90	88
3	360	90	87
4	360	90	87
5	360	90	76
Average			90

5. Results Discussion

The aim of this paper is to propose a side-view facial recognition system based on the fusion of both texture and geometric histograms rather than study they separately. Based on the results of the study we can see that histogram fusion outperform if conducted separately both textural and geometrical methods with an accuracy of around 90% noting that study utilized only 5 images per subject using the SVM classifier. Authors in [24] ave used the NCKU dataset where they have trained the first 37 images of all subjects where they have divided them into 3 subsets 2 for training and 1 for testing. They have proposed the use of Improved Random Regression Forests classifier with an optimal accuracy of 88.32% using the HOG method. Thus, even with a relatively smaller training set conducted in this study histogram fusion yields better results by utilizing the use of SVM classifier and would yield further greater results if trained with more images.

6. Conclusion and Recommendation

The experiments carried out to implement our proposed side view face recognition system have shown promising results as we have hoped it will based on the literature review done prior to the experiments carried out. This study paves a way for further studies with more comprehensive experiments to be carried out in order to produce an increase in accuracy and robustness in side view facial recognition systems, this study concludes by recommending this methodology to be carried out on a much more larger scale as it is known that SVM classifier increases performance with more data in its training phase, perhaps the use of other geometric and texture feature extractors to compare and contrast the performance gotten from the feature extraction methods used in this study.

References

- [1] Aloysius, N. and Geetha, M. (2017). A review on deep convolutional neural networks. In *2017 International Conference on Communication and Signal Processing (ICCSP)*, 0588-0592. IEEE.
- [2] Akanksha, Kaur, J. and Singh, H. (2018). Face detection and recognition: A review. In *6th International Conference on Advancements in Engineering and Technology (ICAET)*, 138-140.
- [3] Chen, J., Chen, Z., Chi, Z. and Fu, H. (2016). Facial expression recognition in video with multiple feature fusion. In *IEEE Transactions on Affective Computing*, 9(1), 38-50.
- [4] Ahmed, K.T., Ummesafi, S. and Iqbal, A. (2019). Content based image retrieval using image features information fusion. *Information Fusion*, 51, 76-99.
- [5] Foody, G.M. and Mathur, A. (2004). Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment*, 93(1-2), 107-117.
- [6] Goldstein, A.J., Harmon, L.D. and Lesk, A.B. (1971). Identification of human faces. *Proceedings of the IEEE*, 59(5), 748-760.
- [7] Gupta, D. and Richhariya, B. (2018). Entropy based fuzzy least squares twin support vector machine for class imbalance learning. *Applied Intelligence*, 48(11), 4212-4231.
- [8] Jiddah, S.M. and Yurtkan, K. (2018). Fusion of geometric and texture features for ear recognition. In *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 1-5. IEEE.
- [9] Julina, J.K.J. and Sharmila, T.S. (2017). Facial recognition using histogram of gradients and support vector machines. In *2017 IEEE International Conference on Computer, Communication and Signal Processing (ICCCSP)*, 1-5. IEEE.
- [10] Kanade, T. (1974). *Picture processing system by computer complex and recognition of human faces*. [Doctoral dissertation, Kyoto University].
- [11] Kaufman, G.J. and Breeding, K.J. (1976). The automatic recognition of human faces from profile silhouettes. *IEEE Transactions on Systems, Man, and Cybernetics*, 2, 113-121.
- [12] Lal, M., Kumar, K., Arain, R.H., Maitlo, A., Ruk, S.A. and Shaikh, H. (2018). Study of face recognition techniques: A survey. *IJACSA International Journal of Advanced Computer Science and Applications*, 9(6), 42-49.
- [13] Masi, I., Rawls, S., Medioni, G. and Natarajan, P. (2016). Pose-aware face recognition in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4838-4846.
- [14] Moore, S. and Bowden, R. (2010). Multi-view pose and facial expression recognition. *The British Machine Vision Conference (BMVC)*, 2, 1-11.
- [15] Naik, R.K. (2014). Advanced face recognition using reconstruction of 2-D frontal face images from multi angled images. *International Journal of Advanced Research in Computer Science & Technology (IJARCST)*, 3(1), 129-132.
- [16] Ojala, T., Pietikäinen, M. and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1), 51-59.
- [17] Patel, R., Rathod, N. and Shah, A. (2012). Comparative analysis of face recognition approaches: a survey. *International Journal of Computer Applications*, 57(17), 50-61.
- [18] Santemiz, P., Spreeuwens, L.J. and Veldhuis, R.N. (2013). Automatic landmark detection and face recognition for side-view face images. In *2013 International Conference of the BIOSIG Special Interest Group (BIOSIG)*, 1-4. IEEE.
- [19] Wang, D., Hoi, S.C., He, Y., Zhu, J., Mei, T. and Luo, J. (2013). Retrieval-based face annotation by weak label regularized local coordinate coding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3), 550-563.
- [20] Wang, J., Zheng, J., Zhang, S., He, J., Liang, X. and Feng, S. (2016). A face recognition system based on local binary patterns and support vector machine for home security service robot. *IEEE 9th International Symposium on Computational Intelligence and Design (ISCID)*, 2, 303-307 .

- [21] Yang, C.S. and Yang, Y.H. (2016). A robust feature descriptor: Signed LBP. In *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (ICIP)*. 316. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
- [22] Zhang, Y.D., Yang, Z.J., Lu, H.M., Zhou, X.X., Phillips, P., Liu, Q.M. and Wang, S.H. (2016). Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. *IEEE Access*, 4, 8375-8385.
- [23] Zhao, W., Chellappa, R., Phillips, P.J. and Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys (CSUR)*, 35(4), 399-458.
- [24] Zhu, R., Sang, G., Cai, Y., You, J. and Zhao, Q. (2013). Head pose estimation with improved random regression forests. In *Chinese Conference on Biometric Recognition*, 457-465. Springer, Cham.

ROBUST OUTER PRODUCT OF GRADIENTS TESTS
FOR TESTING SPATIAL DEPENDENCE

Osman Doğan*

Department of Economics,
University of Illinois,
Urbana, U.S.A

Süleyman Taşpınar

Department of Economics,
Queens College, CUNY,
New York, U.S.A.

Bülent Güloğlu

Department of Economics,
Istanbul Technical University,
Istanbul, Turkey

Abstract: In this paper, we suggest outer-product-of-gradient (OPG) variants of the Lagrange multiplier (LM) test statistic for testing spatial dependence under the local parametric misspecification in spatial models. Our OPG statistic for testing the presence of a spatial lag in the disturbance term remains valid irrespective of whether or not there is spatial dependence in the dependent variable. Similarly, our suggested OPG statistic for testing the presence of a spatial lag in the dependent variable is robust to the presence of spatial dependence in the disturbance term. We also suggest the OPG variants that are robust to the presence of an unknown form of heteroskedasticity in the disturbance terms. The computations of all suggested tests only require the least squares estimates from a linear regression model. In a Monte Carlo simulation, we investigate the finite sample properties of our tests and some alternative tests suggested in the literature. The simulation results show that our tests work well in finite samples.

Key words: Spatial autoregressive models, SARAR, Heteroskedasticity, Testing, Robust tests, LM tests, OPG tests, Inference.

1. Introduction

In this paper, we propose outer-product-of-gradient (OPG) variants of Lagrange multiplier (LM) test statistic for testing spatial dependence in a spatial model that has spatial dependence in both the dependent variable and the disturbance term. Our OPG test for testing one type of spatial dependence (the spatial lag in the dependent variable or the spatial lag in the disturbance term) is valid whether or not the other type of spatial dependence is present. We show how such robust OPG tests can be systematically constructed in the quasi maximum likelihood (QML) framework for spatial models that have homoskedastic or heteroskedastic disturbances. Importantly, the computation of suggested tests only requires the ordinary least squares (OLS) estimates from a linear regression model. Thus, our approach provides robust alternative tests that can be easily adopted by applied researchers.

The OPG variants of LM tests for testing spatial dependence are based on the fact that the score type-functions of a spatial model, which can be written in terms of linear and quadratic forms of disturbance terms, form a martingale difference array. This inherent martingale structure is explored in Kelejian and Prucha [13, 14] to develop a central limit theorem (CLT) for spatial processes. Born and Breitung [7] show how the variance of linear and quadratic forms can be

* Corresponding author. E-mail: odogan@illinois.edu

estimated by the OPG method, and suggest simple test statistics for testing the presence of a spatial lag in the spatial auto-regressive (SAR) and spatial error (SE) models.¹ Their test statistics are robust to heteroskedasticity, and are equivalent to the one-directional LM tests derived in [8] and [1] under the homoskedastic case. Following Born and Breitung [7], Baltagi and Yang [4, 5] suggest the standardized OPG variants of one-directional LM tests that correct for both the mean and variance of the standard LM test statistics for improving the finite sample properties of tests. In the context of standard panel data models, there is evidence that standardizing an LM test improves its performance, especially when the asymptotic critical values are used to implement the test [3, 15, 20]. The simulation results in Baltagi and Yang [4, 5] show that the standardized OPG variants for testing spatial dependence can also perform relatively better in finite sample.

The OPG variants suggested in [7, 4, 5] are one directional tests in the sense that they are designed for testing one type of spatial dependence (spatial lag or spatial error dependence) in the absence of the other type of spatial dependence. However, it is well known that the one directional LM test for one type of spatial dependence will be invalid in the presence of other type of spatial dependence, i.e., under the local parametric misspecification [23, 9, 6, 2]. Anselin et al. [2] study systematically the consequences of testing one type of spatial dependence at a time, and use the approach suggested in [6] to develop adjusted one-directional LM test for testing the presence of spatial dependence in the dependent variable (spatial dependence in the disturbance term) in the possible presence of spatial dependence in the disturbance term (spatial dependence in the dependent variable). Though the test suggested in [2] are valid under the local parametric misspecification, they may not be valid under the heteroskedastic case, since the (quasi) likelihood function is misspecified when the disturbance terms of the model are heteroskedastic. Recently, following Born and Breitung [7], Jin and Lee [12] suggest the OPG variants of $C(\alpha)$ -type tests in the ML and generalized method of moments (GMM) settings under both homoskedastic and heteroskedastic cases. In comparison with our suggested tests, the tests suggested in [12] are computationally intensive as they require estimation of the respective null models by a \sqrt{n} -consistent constrained estimator, and therefore they do not share the simplicity of tests based on the OLS estimator.

The rest of this study is organized as follows. In Section 2, we state the spatial model and derive its quasi-likelihood function. In Section 3, we show how the OPG variants of LM test can be systematically derived under both homoskedastic and heteroskedastic cases. We provide test statistics for detecting spatial error and spatial lag dependence in a SARAR (1,1) model. In Section 4, we describe our Monte Carlo design and report the simulation results. In Section 5, we provide an empirical illustration. In Section 6, we conclude. Some technical details and simulation results are collected in an appendix.

2. Model Specification and ML Estimation Approach

We consider the following cross-sectional SARAR(1,1) specification

$$Y = \lambda_0 WY + X\beta_0 + U, \quad U = \rho_0 MU + V, \quad (2.1)$$

where $Y = (y_1, \dots, y_n)'$ is the $n \times 1$ vector of a dependent variable, X_n is the $n \times k_x$ matrix of non-stochastic exogenous variables with a matching parameter vector β_0 . W and M are the $n \times n$ spatial weights matrices of known constants with zero diagonal elements. The scalar parameters λ_0 and ρ_0 are called the spatial autoregressive parameters. In (2.1), $U = (u_1, \dots, u_n)'$ is the $n \times 1$ vector of regression disturbance terms and $V = (v_1, \dots, v_n)'$ is the $n \times 1$ vector of disturbance terms which are independent and identically distributed (iid) with mean zero and variance σ_0^2 .

¹ On the taxonomy and estimation of spatial models, see [1, 17, 11].

The model is stated with the true parameter vector $\theta_0 = (\beta'_0, \sigma_0^2, \lambda_0, \rho_0)'$ and we use $\theta = (\beta', \sigma^2, \lambda, \rho)'$ to denote any other arbitrary value in the parameter space. For notational simplicity, let $S(\lambda) = (I_n - \lambda W)$, $R(\rho) = (I_n - \rho M)$, $S = S(\lambda_0)$ and $R = R(\rho_0)$, where I_n denotes the $n \times n$ identity matrix. Under the assumption that the disturbance terms are iid with mean zero and variance σ_0^2 , the quasi log-likelihood function of the model can be written as

$$\begin{aligned} \ln L(\theta) &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) + \ln |S(\lambda)| + \ln |R(\rho)| \\ &\quad - \frac{1}{2\sigma^2} (S(\lambda)Y - X\beta)' R'(\rho)R(\rho) (S(\lambda)Y - X\beta), \end{aligned} \quad (2.2)$$

where $|\cdot|$ denotes the determinant operator. The QMLE is the extremum estimator defined by $\hat{\theta} = \arg \max_{\theta \in \Theta} \ln L(\theta)$. Under some regularity conditions, it can be shown that the QMLE is consistent and has asymptotic normal distribution [16].

3. Robust OPG Tests

In this section, we derive robust OPG tests for testing the null hypotheses $H_0^\rho : \rho_0 = 0$ and $H_0^\lambda : \lambda_0 = 0$ under both homekadastic and heteroskedastic cases. We also determine the asymptotic distribution of tests under the local alternative hypotheses defined as $H_a^\rho : \rho_0 = \delta_\rho / \sqrt{n}$ and $H_a^\lambda : \lambda_0 = \delta_\lambda / \sqrt{n}$, where δ_ρ and δ_λ are non-stochastic bounded constants. In testing $H_0^\rho : \rho_0 = 0$, the local alternative hypothesis $H_a^\lambda : \lambda_0 = \delta_\lambda / \sqrt{n}$ also serves as the local parametric misspecification. Similarly, in testing $H_0^\lambda : \lambda_0 = 0$, the local alternative hypothesis $H_a^\rho : \rho_0 = \delta_\rho / \sqrt{n}$ represents the local parametric misspecification in our setting.

3.1. Tests Under Homoskedasticity

Let $S_a(\theta) = \frac{1}{n} \frac{\partial \ln L(\theta_0)}{\partial a}$, where $a \in \{\beta, \sigma^2, \lambda, \rho\}$. Our robust OPG tests are based on the score functions of $\ln L(\theta)/n$ derived as

$$\begin{aligned} S_\beta(\theta_0) &= \frac{1}{n\sigma_0^2} X' R' V, \quad S_{\sigma^2}(\theta_0) = \frac{1}{2n\sigma_0^4} V' V - \frac{1}{2\sigma_0^2}, \\ S_\lambda(\theta_0) &= \frac{1}{n\sigma_0^2} V' \bar{G} V - \frac{1}{n} \text{tr}(G) + \frac{1}{n\sigma_0^2} (RGX\beta_0)' V, \quad S_\rho(\theta_0) = \frac{1}{n\sigma_0^2} V' H V - \frac{1}{n} \text{tr}(H), \end{aligned} \quad (3.1)$$

where $G = WS^{-1}$, $\bar{G} = RGR^{-1}$ and $H = MR^{-1}$. Let $\gamma = (\beta', \sigma^2)'$. Then, the score functions can be expressed as a vector of linear quadratic forms in the following way

$$S(\theta_0) = (S'_\gamma(\theta_0), S_\lambda(\theta_0), S_\rho(\theta_0))' = \frac{1}{n} \begin{pmatrix} V' A_1 V - \sigma_0^2 \text{tr}(A_1) + b'_1 V \\ \vdots \\ V' A_{k_x} V - \sigma_0^2 \text{tr}(A_{k_x}) + b'_{k_x} V \\ V' A_{k_x+1} V - \sigma_0^2 \text{tr}(A_{k_x+1}) + b'_{k_x+1} V \\ V' A_{k_x+2} V - \sigma_0^2 \text{tr}(A_{k_x+2}) + b'_{k_x+2} V \\ V' A_{k_x+3} V - \sigma_0^2 \text{tr}(A_{k_x+3}) + b'_{k_x+3} V \end{pmatrix}, \quad (3.2)$$

where $\text{tr}(\cdot)$ is the trace operator, $A_1 = \dots = A_{k_x} = 0$, $b_j = RX_j / \sigma_0^2$ for $j = 1, 2, \dots, k_x$, X_j is the j th column of X , $A_{k_x+1} = I_n / 2\sigma_0^4$, $b_{k_x+1} = 0$, $A_{k_x+2} = \bar{G} / \sigma_0^2$, $b_{k_x+2} = (RGX\beta) / \sigma_0^2$, $A_{k_x+3} = H / \sigma_0^2$ and $b_{k_x+3} = 0$. We denote the (i, j) th element of A_k by $a_{k,ij}$, and the i th element of b_k by b_{ki} , where $k \in \{1, 2, \dots, k_x + 3\}$. Then, we can express $S(\theta_0)$ as a sum of martingale differences in the following way

$$S(\theta_0) = \sum_{i=1}^n S_i(\theta_0), \quad (3.3)$$

where $S_i(\theta_0) = \left(S'_{i\gamma}(\theta_0), S_{i\lambda}(\theta_0), S_{i\rho}(\theta_0) \right)'$ with the k th element ($k \in \{1, 2, \dots, k_x + 3\}$)

$$S_{ik}(\theta_0) = \frac{1}{n} \left(a_{k,ii}(v_i^2 - \sigma_0^2) + v_i \sum_{j=1}^{i-1} (a_{k,ij} + a_{k,ji})v_j + b_{ki}v_i \right).$$

To show that $S_i(\theta_0)$ is martingale difference sequence, consider the following σ -fields $\mathfrak{F}_0 = \{\emptyset, \Omega\}$ and $\mathfrak{F}_i = \sigma(v_1, v_2, \dots, v_i)$ for $1 \leq i \leq n$. Then, it follows that $\mathfrak{F}_{i-1} \subset \mathfrak{F}_i$ and $\mathbb{E}(S_i(\theta_0) | \mathfrak{F}_{i-1}) = 0$. Thus $\{S_i(\theta_0), \mathfrak{F}_i, 1 \leq i \leq n\}$ forms a martingale difference array. We can thus express the variance of $\sqrt{n}S(\theta_0)$ in the following way

$$\text{Var}(\sqrt{n}S(\theta_0)) = K(\theta_0) = \sum_{i=1}^n \mathbb{E} \left(nS_i(\theta_0)S_i'(\theta_0) \right) \quad (3.4)$$

Under some regularity conditions, $K(\theta_0)$ can be estimated by $K(\hat{\theta}) = n \sum_{i=1}^n S_i(\hat{\theta})S_i'(\hat{\theta})$, where $\hat{\theta}$ is a consistent estimator of θ_0 .

To define our suggested robust OPG test, we assume the following assumptions.

ASSUMPTION 1. *The innovation terms v_i s are iid with mean zero, variance σ_0^2 , and $\mathbb{E}|v_i|^{4+n} < \infty$ for some $\eta > 0$ for all i and n .*

ASSUMPTION 2. (i) *Let $\tilde{\theta}$ be a constrained estimator under under the joint null hypothesis $H_0 : \lambda_0 = \rho_0 = 0$. Then, $\sqrt{n}(\tilde{\theta} - \theta_0) = O_p(1)$ holds. (ii) $\sqrt{n}S(\theta_0) \xrightarrow{d} N(0, K(\theta_0))$, where $K(\theta_0)$ is a non-singular matrix. (iii) $-\frac{\partial S(\tilde{\theta})}{\partial \theta'} = J(\theta_0) + o_p(1)$, where $\tilde{\theta} = \theta_0 + o_p(1)$ and $J(\theta_0) = \mathbb{E} \left(-\frac{\partial S(\theta_0)}{\partial \theta'} \right)$ is a non-singular matrix.*

Assumption 1 requires that the disturbance terms are homoskedastic, but allows for a non-normal distribution. Assumption 2 is a high level assumption as our focus is on testing problem.² The first part of the assumption requires that the constrained estimator under the joint null hypothesis $H_0 : \lambda_0 = \rho_0 = 0$ is a \sqrt{n} -consistent estimator of θ_0 . Note that under $H_0 : \lambda_0 = \rho_0 = 0$, our model reduces to the linear regression model, which can be consistently estimated by the OLS estimator. Assumption 2 (ii) is the CLT condition for the score functions, which can be ensured by the CLT for linear-quadratic form in [13, 14]. Finally, Assumption 2 (iii) shows that the negative hessian evaluated at a consistent estimator converges to the information matrix.

We assume that $K(\theta_0)$ and $J(\theta_0)$ are partitioned into sub-matrices $K_{ab}(\theta_0)$ and $J_{ab}(\theta_0)$ according to the dimensions of a and b , where $a, b \in \{\lambda, \rho, \gamma\}$. Then, it can be shown that

$$\begin{aligned} J_{\lambda\lambda}(\theta_0) &= \frac{1}{n\sigma_0^2} (RGX\beta_0)' RGX\beta_0 + \frac{1}{n} \text{tr}(\bar{G}^{(s)}\bar{G}), & J_{\lambda\rho}(\theta_0) &= \frac{1}{n} \text{tr}(H^{(s)}\bar{G}), \\ J_{\lambda\gamma}(\theta_0) &= \left(\frac{1}{n\sigma_0^2} X' R' RGX\beta_0, \frac{1}{n\sigma_0^2} \text{tr}(G) \right), & J_{\rho\rho}(\theta_0) &= \frac{1}{n} \text{tr}(H^{(s)}H), \\ J_{\rho\gamma}(\theta_0) &= \left(0, \frac{1}{n\sigma_0^2} \text{tr}(H) \right), & J_{\gamma\gamma}(\theta_0) &= \text{Diag} \left(\frac{1}{n\sigma^2} X' R' RX, \frac{1}{2\sigma_0^4} \right), \end{aligned} \quad (3.5)$$

where $A^{(s)} = A + A'$ for any $n \times n$ matrix A .

Before we introduce our suggested tests, we introduce the following notations. We define $J_{\lambda,\gamma}(\theta) = J_{\lambda\lambda}(\theta) - J_{\lambda\gamma}(\theta)J_{\gamma\gamma}^{-1}(\theta)J_{\gamma\lambda}(\theta)$ and $J_{\lambda\rho,\gamma}(\theta) = J_{\lambda\rho}(\theta) - J_{\lambda\gamma}(\theta)J_{\gamma\gamma}^{-1}(\theta)J_{\gamma\rho}(\theta)$. Similarly, $J_{\rho,\gamma}(\theta) = J_{\rho\rho}(\theta) - J_{\rho\gamma}(\theta)J_{\gamma\gamma}^{-1}(\theta)J_{\gamma\rho}(\theta)$ and $J_{\rho\lambda,\gamma}(\theta) = J_{\rho\lambda}(\theta) - J_{\rho\gamma}(\theta)J_{\gamma\gamma}^{-1}(\theta)J_{\gamma\lambda}(\theta)$. Under Assumption 2, it

² The primitive conditions ensuring this assumption are provided in [16]. For the sake of brevity, we do not provide these primitive conditions.

can be shown that the mean value expansions of $\sqrt{n}S_\lambda(\tilde{\theta})$ and $\sqrt{n}S_\gamma(\tilde{\theta})$ around θ_0 , when both H_a^λ and H_a^ρ hold, yield the following result (see proof of Theorem 1)

$$\sqrt{n}S_\lambda(\tilde{\theta}) \xrightarrow{d} N [J_{\lambda\cdot\gamma}(\theta_0)\delta_\lambda + J_{\lambda\rho\cdot\gamma}(\theta_0)\delta_\rho, B_{\lambda\cdot\gamma}(\theta_0)], \quad (3.6)$$

where

$$\begin{aligned} B_{\lambda\cdot\gamma}(\theta_0) &= K_{\lambda\lambda}(\theta_0) + J_{\lambda\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)K_{\gamma\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)J'_{\lambda\gamma}(\theta_0) \\ &\quad - K_{\lambda\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)J_{\lambda\gamma}(\theta_0) - J_{\lambda\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)K_{\gamma\lambda}(\theta_0). \end{aligned} \quad (3.7)$$

Hence, under H_0^λ and H_0^ρ , it follows that $\sqrt{n}S_\lambda(\tilde{\theta}) - J_{\lambda\rho\cdot\gamma}(\theta_0)\delta_\rho \xrightarrow{d} N(0, B_{\lambda\cdot\gamma}(\theta_0))$. The non-zero asymptotic mean term $J_{\lambda\rho\cdot\gamma}(\theta_0)\delta_\rho$ indicates that the null asymptotic distribution of the OPG test based on the $\sqrt{n}S_\lambda(\tilde{\theta})$ will have a non-central chi-squared distribution in the local presence of ρ_0 . We show that δ_ρ is the asymptotic mean of $J_{\rho\cdot\gamma}^{-1}\sqrt{n}S_\rho(\tilde{\theta})$ (see the proof of Theorem 1). This result suggests that we can formulate a OPG test that is valid in the local presence of ρ_0 based on the following adjusted score function.

$$\sqrt{n}S_\lambda^*(\tilde{\theta}) = \sqrt{n} \left(S_\lambda(\tilde{\theta}) - J_{\lambda\rho\cdot\gamma}(\tilde{\theta})J_{\rho\cdot\gamma}^{-1}(\tilde{\theta})S_\rho(\tilde{\theta}) \right). \quad (3.8)$$

Then, our suggested robust OPG test is given by

$$LM_\lambda = n S_\lambda^{*2}(\tilde{\theta}) / D_{\lambda\cdot\gamma}(\tilde{\theta}), \quad (3.9)$$

where

$$\begin{aligned} D_{\lambda\cdot\gamma}(\tilde{\theta}) &= B_{\lambda\cdot\gamma}(\tilde{\theta}) + J_{\lambda\rho\cdot\gamma}(\tilde{\theta})J_{\rho\cdot\gamma}^{-1}(\tilde{\theta})B_{\rho\cdot\gamma}(\tilde{\theta})J_{\rho\cdot\gamma}^{-1}(\tilde{\theta})J_{\rho\lambda\cdot\gamma}(\tilde{\theta}) \\ &\quad - J_{\lambda\rho\cdot\gamma}(\tilde{\theta})J_{\rho\cdot\gamma}^{-1}(\tilde{\theta})B_{\rho\lambda\cdot\gamma}(\tilde{\theta}) - B_{\lambda\rho\cdot\gamma}(\tilde{\theta})J_{\rho\cdot\gamma}^{-1}(\tilde{\theta})J_{\rho\lambda\cdot\gamma}(\tilde{\theta}), \end{aligned} \quad (3.10)$$

with

$$\begin{aligned} B_{\lambda\cdot\gamma}(\tilde{\theta}) &= K_{\lambda\lambda}(\tilde{\theta}) + J_{\lambda\gamma}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})K_{\gamma\gamma}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})J_{\gamma\lambda}(\tilde{\theta}) \\ &\quad - K_{\lambda\gamma}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})J_{\gamma\lambda}(\tilde{\theta}) - J_{\lambda\gamma}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})K_{\gamma\lambda}(\tilde{\theta}), \end{aligned} \quad (3.11)$$

$$\begin{aligned} B_{\lambda\rho\cdot\gamma}(\tilde{\theta}) &= K_{\lambda\rho}(\tilde{\theta}) - J_{\lambda\gamma}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})K_{\gamma\rho}(\tilde{\theta}) - K_{\lambda\gamma}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})J_{\gamma\rho}(\tilde{\theta}) \\ &\quad + J_{\lambda\gamma}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})K_{\gamma\rho}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})J_{\gamma\rho}(\tilde{\theta}), \end{aligned} \quad (3.12)$$

$$\begin{aligned} B_{\rho\cdot\gamma}(\tilde{\theta}) &= K_{\rho\rho}(\tilde{\theta}) + J_{\rho\gamma}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})K_{\gamma\gamma}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})J_{\gamma\rho}(\tilde{\theta}) \\ &\quad - K_{\rho\gamma}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})J_{\gamma\rho}(\tilde{\theta}) - J_{\rho\gamma}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})K_{\gamma\rho}(\tilde{\theta}), \end{aligned} \quad (3.13)$$

$$\begin{aligned} B_{\rho\lambda\cdot\gamma}(\tilde{\theta}) &= K_{\rho\lambda}(\tilde{\theta}) - J_{\rho\gamma}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})K_{\gamma\lambda}(\tilde{\theta}) - K_{\rho\gamma}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})J_{\gamma\lambda}(\tilde{\theta}) \\ &\quad + J_{\rho\gamma}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})K_{\gamma\lambda}(\tilde{\theta})J_{\gamma\gamma}^{-1}(\tilde{\theta})J_{\gamma\lambda}(\tilde{\theta}). \end{aligned} \quad (3.14)$$

THEOREM 1. Assume that Assumptions 1-2 hold. Then, under H_a^λ , it follows that

$$LM_\lambda \xrightarrow{d} \chi_1^2(\vartheta_1), \quad (3.15)$$

where

$$\vartheta_1 = \delta_\lambda^2 (J_{\lambda\cdot\gamma}(\theta_0) - J_{\lambda\rho\cdot\gamma}(\theta_0)J_{\rho\cdot\gamma}^{-1}(\theta_0)J_{\rho\lambda\cdot\gamma}(\theta_0))^2 / D_{\lambda\cdot\gamma}(\theta_0). \quad (3.16)$$

PROOF. See Appendix 6.

Theorem 1 indicates that $LM_\lambda \xrightarrow{d} \chi_1^2$ under H_0^λ in the local presence of ρ_0 , i.e., LM_λ is a valid test statistic irrespective of whether H_0^ρ or H_a^ρ holds. Though our suggested test statistic has a lengthy expression, its calculation only requires $\tilde{\theta}$. Test statistic simplify significantly when $J_{\lambda\rho\cdot\gamma}(\tilde{\theta}) = 0$ holds. In that case, we have $LM_\lambda = nS_\lambda^2(\tilde{\theta})/D_{\lambda\cdot\gamma}(\tilde{\theta})$, where $D_{\lambda\cdot\gamma}(\tilde{\theta}) = B_{\lambda\cdot\gamma}(\tilde{\theta}) - B_{\lambda\rho\cdot\gamma}(\tilde{\theta})J_{\rho\cdot\gamma}^{-1}(\tilde{\theta})J_{\rho\lambda\cdot\gamma}(\tilde{\theta})$.

We can similarly determine the test statistic for testing H_a^ρ in the local presence of λ_0 . The robust OGP test statistic is given by

$$LM_\rho = nS_\rho^{*2}(\tilde{\theta})/D_{\rho\cdot\gamma}(\tilde{\theta}), \quad (3.17)$$

where $S_\rho^*(\tilde{\theta}) = \left(S_\rho(\tilde{\theta}) - J_{\rho\lambda\cdot\gamma}(\tilde{\theta})J_{\lambda\cdot\gamma}^{-1}(\tilde{\theta})S_\lambda(\tilde{\theta}) \right)$ is the adjusted score function and

$$D_{\rho\cdot\gamma}(\tilde{\theta}) = B_{\rho\cdot\gamma}(\tilde{\theta}) + J_{\rho\lambda\cdot\gamma}(\tilde{\theta})J_{\lambda\cdot\gamma}^{-1}(\tilde{\theta})B_{\lambda\cdot\gamma}(\tilde{\theta})J_{\lambda\cdot\gamma}^{-1}(\tilde{\theta})J_{\lambda\rho\cdot\gamma}(\tilde{\theta}) - J_{\rho\lambda\cdot\gamma}(\tilde{\theta})J_{\lambda\cdot\gamma}^{-1}(\tilde{\theta})B_{\lambda\rho\cdot\gamma}(\tilde{\theta}) - B_{\rho\lambda\cdot\gamma}(\tilde{\theta})J_{\lambda\cdot\gamma}^{-1}(\tilde{\theta})J_{\lambda\rho\cdot\gamma}(\tilde{\theta}). \quad (3.18)$$

The asymptotic null distribution of LM_ρ is a central chi-squared distribution in the local presence of λ_0 as shown in the following theorem.

THEOREM 2. Assume that Assumptions 1-2 hold. Then, under H_a^ρ , it follows that

$$LM_\rho \xrightarrow{d} \chi_1^2(\vartheta_2), \quad (3.19)$$

where

$$\vartheta_2 = \delta_\rho^2 \left(J_{\rho\cdot\gamma}(\theta_0) - J_{\rho\lambda\cdot\gamma}(\theta_0)J_{\lambda\cdot\gamma}^{-1}(\theta_0)J_{\lambda\rho\cdot\gamma}(\theta_0) \right)^2 / D_{\rho\cdot\gamma}(\theta_0). \quad (3.20)$$

PROOF. See Appendix 6.

Theorem 2 indicates that the asymptotic null distribution of LM_ρ is χ_1^2 , i.e., $LM_\rho \xrightarrow{d} \chi_1^2$ under H_0^ρ .

REMARK 1. In terms of our notation, the OPG variants of LM test suggested in [7] for H_0^ρ can be expressed as

$$LM_\rho^s = nS_\rho^2(\tilde{\theta})/K_{\rho\rho}(\tilde{\theta}) \quad (3.21)$$

Under the null hypothesis, this test statistic is asymptotically equivalent to the LM test statistic suggested in [8] and the standardized OPG-based LM test suggested in [5, Theorem 2.2]. Here, we show that this test statistic is invalid under the local parametric misspecification of the form $\lambda_0 = \delta_\lambda/\sqrt{n}$. Using the asymptotic argument given in the proof of Theorem 1, it can be shown that the asymptotic distribution of $\sqrt{n}S_\rho(\tilde{\theta})$ under H_a^λ and H_a^ρ is given by

$$\sqrt{n}S_\rho(\tilde{\theta}) \xrightarrow{d} N [J_{\rho\cdot\gamma}(\theta_0)\delta_\rho + J_{\rho\lambda\cdot\gamma}(\theta_0)\delta_\lambda, B_{\rho\cdot\gamma}(\theta_0)], \quad (3.22)$$

where

$$B_{\rho\cdot\gamma}(\theta_0) = K_{\rho\rho}(\theta_0) + J_{\rho\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)K_{\gamma\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)J'_{\lambda\gamma}(\theta_0) - K_{\rho\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)J_{\rho\gamma}(\theta_0) - J_{\rho\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)K_{\gamma\rho}(\theta_0). \quad (3.23)$$

Our results in (3.5) indicate that $J_{\rho\gamma}(\theta_0) = 0$ under the joint null hypothesis $H_0 : \lambda_0 = \rho_0 = 0$. Moreover, our Assumption 2 ensures that a consistent estimator of $B_{\rho\cdot\gamma}(\theta_0)$ is $B_{\rho\rho}(\tilde{\theta}) = K_{\rho\rho}(\tilde{\theta})$. Then, under H_a^λ and H_a^ρ , using Theorem 8.6 of White [24] on the asymptotic distribution of quadratic forms, we obtain

$$LM_\rho^s \xrightarrow{d} \chi_1^2(\varphi_1), \quad (3.24)$$

where $\varphi_1 = (J_{\rho,\gamma}(\theta_0)\delta_\rho + J_{\rho\lambda,\gamma}(\theta_0)\delta_\lambda)^2 / K_{\rho\rho}(\theta_0)$. In the local presence of λ_0 , the result in (3.24) indicates that the asymptotic null distribution of LM_ρ^s is $\chi_1^2(\varphi_2)$, where $\varphi_2 = (J_{\rho\lambda,\gamma}(\theta_0)\delta_\lambda)^2 / K_{\rho\rho}(\theta_0)$. That is, in the presence of local parametric misspecification of the form $\lambda_0 = \delta_\lambda / \sqrt{n}$, LM_ρ^a will over reject H_0^o due to the non-centrality parameter φ_2 .

REMARK 2. For testing H_0^λ , our result in (3.6) suggests the following non-robust test statistic

$$LM_\lambda^s = n S_\lambda^2(\tilde{\theta}) / B_{\lambda,\gamma}(\tilde{\theta}), \tag{3.25}$$

In contrast to LM_ρ^s , our result on the elements of 3.5 indicates that $B_{\lambda,\gamma}(\tilde{\theta})$ is not equal to $K_{\lambda\lambda}(\tilde{\theta})$ since $J_{\lambda\rho}(\tilde{\theta}) = \frac{1}{n} \text{tr}(M^{(s)}W) = O(1)$. Then, under H_a^λ and H_a^ρ , it follows that

$$LM_\lambda^s \xrightarrow{d} \chi_1^2(\varphi_3), \tag{3.26}$$

where $\varphi_3 = (J_{\lambda,\gamma}(\theta_0)\delta_\lambda + J_{\lambda\rho,\gamma}(\theta_0)\delta_\rho)^2 / B_{\lambda,\gamma}(\theta_0)$ is the non-centrality parameter. Thus, in the local presence of ρ_0 , the asymptotic null distribution of LM_λ is a non-central chi-squared distribution with the non-centrality parameter of $\varphi_4 = (J_{\lambda\rho,\gamma}(\theta_0)\delta_\rho)^2 / B_{\lambda,\gamma}(\theta_0)$. The test statistic suggested in [7] for testing H_0^λ is different from the one in (3.26). Their test statistic is equivalent to the LM test statistics derived in [1] for the SAR model $Y = \lambda_0 WY + X\beta_0 + U$ [7, Proposition 4.1]. Since the LM statistic in [1] for the SAR model is invalid in the local presence of ρ_0 [2], it follows that the OPG variant in [7] for testing H_0^λ is also invalid in the local presence of ρ_0 .

3.2. Tests Under Heteroskedasticity

In this section, we assume that the disturbance terms v_i 's are independent, but heteroskedastic with variances σ_i^2 . In this case, it is known that the score function derived from the quasi likelihood function in (2.2) do not have zero expected values [18, 10]. This result indicates that the QMLE defined by $\hat{\theta} = \arg \max_{\theta \in \Theta} \ln L(\theta)$ may not be a consistent estimator of θ_0 . Liu and Yang [19] and Yang [25] modify the score functions such that they have zero means, and define a consistent estimator as a root of adjusted score functions. We follow the same approach to formulate score-based OPG tests under heteroskedastic case. Since the modified scores can be considered as proper moment functions, our suggested tests can also be called the m-tests [24].

Let $\theta_0 = (\beta_0', \lambda_0, \rho_0)'$. We use $\text{Diag}(A_k)$ to denote the diagonal matrix whose diagonal elements are the (i, i) th elements $a_{k,ii}$'s, i.e., $\text{Diag}(A_k) = \text{Diag}(a_{k,11}, a_{k,22}, \dots, a_{k,nn})$. Following [19], we consider the following modified score functions that have zero means.

$$C_\beta(\theta_0) = \frac{1}{n} X' R' V, \quad C_\lambda(\theta_0) = \frac{1}{n} V' (\bar{G} - \text{Diag}(\bar{G})) V + \frac{1}{n} (RGX\beta_0)' V, \tag{3.27}$$

$$C_\rho(\theta_0) = \frac{1}{n} V' (H - \text{Diag}(H)) V, \tag{3.28}$$

where $V = R(SY - X\beta_0)$. In terms of linear quadratic forms, these score functions can be expressed as

$$C(\theta_0) = (C'_\beta(\theta_0), C'_\lambda(\theta_0), C'_\rho(\theta_0))' = \frac{1}{n} \begin{pmatrix} V' A_1 V + b'_1 V \\ \vdots \\ V' A_{k_x} V + b'_{k_x} V \\ V' A_{k_x+1} V + b'_{k_x+1} V \\ V' A_{k_x+2} V + b'_{k_x+2} V \end{pmatrix}, \tag{3.29}$$

where $A_1 = \dots = A_{k_x} = 0$, $b_j = RX_j$ for $j = 1, 2, \dots, k_x$, $A_{k_x+1} = \bar{G} - \text{Diag}(\bar{G})$, $b_{k_x+1} = RGX\beta$, $A_{k_x+2} = H - \text{Diag}(H)$ and $b_{k_x+2} = 0$. As before, we can express $S(\theta_0)$ as a sum of martingale differences in the following way

$$C(\theta_0) = \sum_{i=1}^n C_i(\theta_0), \tag{3.30}$$

where $C_i(\theta_0) = \left(C'_{i\beta}(\theta_0), C_{i\lambda}(\theta_0), C_{i\rho}(\theta_0) \right)'$ with the k th element ($k \in \{1, 2, \dots, k_x + 2\}$), $C_{ik}(\theta_0) = v_i \sum_{j=1}^{i-1} (a_{k,ij} + a_{k,ji})v_j + b_{ki}v_i$. As before, the variance of $\sqrt{n}C(\theta_0)$ takes the following form

$$\text{Var}(\sqrt{n}C(\theta_0)) = K(\theta_0) = n \sum_{i=1}^n \mathbb{E} \left(C_i(\theta_0) C'_i(\theta_0) \right). \quad (3.31)$$

Under some regularity condition, $K(\theta_0)$ can be estimated by $K(\hat{\theta}) = n \sum_{i=1}^n C_i(\hat{\theta}) C'_i(\hat{\theta})$, where $\hat{\theta}$ is a consistent estimator of θ_0 .

To define our suggested robust OGP test, we assume the following assumptions.

ASSUMPTION 3. *The innovation terms v_i s are independent with mean zero, variance σ_i^2 and $\mathbb{E}|v_i|^{4+\eta} < \infty$ for some $\eta > 0$ for all i and n .*

ASSUMPTION 4. (i) *Let $\tilde{\theta}$ be a constrained estimator under under the joint null hypothesis $H_0 : \lambda_0 = \rho_0 = 0$. Then, $\sqrt{n}(\tilde{\theta} - \theta_0) = O_p(1)$ holds. (ii) $\sqrt{n}C(\theta_0) \xrightarrow{d} N(0, K(\theta_0))$, where $K(\theta_0)$ is a non-singular matrix. (iii) $-\frac{\partial C(\tilde{\theta})}{\partial \theta'} = J(\theta_0) + o_p(1)$, where $\tilde{\theta} = \theta_0 + o_p(1)$ and $J(\theta_0) = \mathbb{E} \left(-\frac{\partial C(\theta_0)}{\partial \theta'} \right)$ is a non-singular matrix.*

Assumption 3 specifies heteroskedastic disturbance terms [14, 18]. As in the homoskedastic case, Assumption 4 is a high level assumption and provides conditions that are counterparts of those stated in Assumption 2.³

Under Assumption 3, the elements of $J(\theta_0)$ can be derived as

$$\begin{aligned} J_{\lambda\lambda}(\theta_0) &= \frac{1}{n} \text{tr} \left((\bar{G} - \text{Diag}(\bar{G}))^{(s)} \bar{G} \Sigma \right) + \frac{1}{n} (RGX\beta_0)' RGX\beta_0 \\ J_{\lambda\rho}(\theta_0) &= J_{\rho\lambda}(\theta_0) = \frac{1}{n} \text{tr} \left((H - \text{Diag}(H))^{(s)} \bar{G} \Sigma \right), \\ J_{\lambda\beta}(\theta_0) &= J'_{\beta\lambda}(\theta_0) = \frac{1}{n} \beta_0' X' G' R' R X, \\ J_{\rho\rho}(\theta_0) &= \frac{1}{n} \text{tr} \left((H - \text{Diag}(H))^{(s)} H \Sigma \right), \quad J_{\rho\beta}(\theta_0) = 0, \quad J_{\beta\rho}(\theta_0) = 0, \\ J_{\beta\beta}(\theta_0) &= \frac{1}{n} X' R' R X, \end{aligned}$$

where $\Sigma = \text{Diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$. These expressions show that we can use a plug-in estimator to estimate $J(\theta_0)$ (here, we can use $\tilde{\Sigma} = \text{Diag}(\hat{v}_1^2, \hat{v}_2^2, \dots, \hat{v}_n^2)$ for Σ).⁴

Under heteroskedastic disturbances, our suggested test statistics for H_0^λ and H_0^ρ are respectively given by

$$LM_\lambda^h = n C_\lambda^{*2}(\tilde{\theta}) / D_{\lambda,\beta}(\tilde{\theta}), \quad (3.32)$$

$$LM_\rho^h = n C_\rho^{*2}(\tilde{\theta}) / D_{\rho,\beta}(\tilde{\theta}) \quad (3.33)$$

where $C_\lambda^*(\tilde{\theta}) = \left(C_\lambda(\tilde{\theta}) - J_{\lambda\rho,\beta}(\tilde{\theta}) J_{\rho,\beta}^{-1}(\tilde{\theta}) C_\rho(\tilde{\theta}) \right)$ and $C_\rho^*(\tilde{\theta}) = \left(C_\rho(\tilde{\theta}) - J_{\rho\lambda,\beta}(\tilde{\theta}) J_{\lambda,\beta}^{-1}(\tilde{\theta}) C_\lambda(\tilde{\theta}) \right)$ are the adjusted score functions and

$$\begin{aligned} D_{\lambda,\beta}(\tilde{\theta}) &= B_{\lambda,\beta}(\tilde{\theta}) + J_{\lambda\rho,\beta}(\tilde{\theta}) J_{\rho,\beta}^{-1}(\tilde{\theta}) B_{\rho,\beta}(\tilde{\theta}) J_{\rho,\beta}^{-1}(\tilde{\theta}) J_{\rho\lambda,\beta}(\tilde{\theta}) \\ &\quad - J_{\lambda\rho,\beta}(\tilde{\theta}) J_{\rho,\beta}^{-1}(\tilde{\theta}) B_{\rho\lambda,\beta}(\tilde{\theta}) - B_{\lambda\rho,\beta}(\tilde{\theta}) J_{\rho,\beta}^{-1}(\tilde{\theta}) J_{\rho\lambda,\beta}(\tilde{\theta}), \end{aligned} \quad (3.34)$$

³ The primitive conditions ensuring this assumption are provided in [14, 18]. For the sake of brevity, we do not repeat these primitive conditions.

⁴ The the asymptotic argument is provided in [18].

$$D_{\rho,\beta}(\tilde{\theta}) = B_{\rho,\beta}(\tilde{\theta}) + J_{\rho\lambda,\beta}(\tilde{\theta})J_{\lambda,\beta}^{-1}(\tilde{\theta})B_{\lambda,\beta}(\tilde{\theta})J_{\lambda,\beta}^{-1}(\tilde{\theta})J_{\lambda\rho,\beta}(\tilde{\theta}) \\ - J_{\rho\lambda,\beta}(\tilde{\theta})J_{\lambda,\beta}^{-1}(\tilde{\theta})B_{\lambda\rho,\beta}(\tilde{\theta}) - B_{\rho\lambda,\beta}(\tilde{\theta})J_{\lambda,\beta}^{-1}(\tilde{\theta})J_{\lambda\rho,\beta}(\tilde{\theta}), \quad (3.35)$$

with

$$B_{\lambda,\beta}(\tilde{\theta}) = K_{\lambda\lambda}(\tilde{\theta}) + J_{\lambda\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})K_{\beta\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})J_{\beta\lambda}(\tilde{\theta}) \\ - K_{\lambda\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})J_{\beta\lambda}(\tilde{\theta}) - J_{\lambda\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})K_{\beta\lambda}(\tilde{\theta}), \quad (3.36)$$

$$B_{\lambda\rho,\beta}(\tilde{\theta}) = K_{\lambda\rho}(\tilde{\theta}) - J_{\lambda\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})K_{\beta\rho}(\tilde{\theta}) - K_{\lambda\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})J_{\beta\rho}(\tilde{\theta}) \\ + J_{\lambda\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})K_{\beta\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})J_{\beta\rho}(\tilde{\theta}), \quad (3.37)$$

$$B_{\rho,\beta}(\tilde{\theta}) = K_{\rho\rho}(\tilde{\theta}) + J_{\rho\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})K_{\beta\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})J_{\beta\rho}(\tilde{\theta}) \\ - K_{\rho\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})J_{\beta\rho}(\tilde{\theta}) - J_{\rho\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})K_{\beta\beta}(\tilde{\theta}), \quad (3.38)$$

$$B_{\rho\lambda,\beta}(\tilde{\theta}) = K_{\rho\lambda}(\tilde{\theta}) - J_{\rho\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})K_{\beta\lambda}(\tilde{\theta}) - K_{\rho\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})J_{\beta\lambda}(\tilde{\theta}) \\ + J_{\rho\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})K_{\beta\beta}(\tilde{\theta})J_{\beta\beta}^{-1}(\tilde{\theta})J_{\beta\lambda}(\tilde{\theta}). \quad (3.39)$$

The next theorem provides the asymptotic distributions of these tests.

THEOREM 3. Assume that Assumptions 3-4 hold.

1. Under H_a^λ , it follows that

$$LM_\lambda^h \xrightarrow{d} \chi_1^2(\vartheta_3), \quad (3.40)$$

where

$$\vartheta_3 = \delta_2^2 (J_{\lambda,\beta}(\theta_0) - J_{\lambda\rho,\beta}(\theta_0)J_{\rho,\beta}^{-1}(\theta_0)J_{\rho\lambda,\beta}(\theta_0))^2 / D_{\lambda,\beta}(\theta_0). \quad (3.41)$$

2. Under H_a^ρ , it follows that

$$LM_\rho^h \xrightarrow{d} \chi_1^2(\vartheta_4), \quad (3.42)$$

where

$$\vartheta_4 = \delta_\rho^2 (J_{\rho,\beta}(\theta_0) - J_{\rho\lambda,\beta}(\theta_0)J_{\lambda,\beta}^{-1}(\theta_0)J_{\lambda\rho,\beta}(\theta_0))^2 / D_{\rho,\beta}(\theta_0). \quad (3.43)$$

PROOF. See Appendix 6.

Theorem 3 shows that LM_λ and LM_λ are valid test statistics in the presence of parametric misspecification. That is, $LM_\lambda^h \xrightarrow{d} \chi_1^2$ under H_0^λ , and $LM_\rho^h \xrightarrow{d} \chi_1^2$ under H_0^ρ .

REMARK 3. The non-robust versions stated in Remarks 1 and 2 take the following forms

$$LM_\rho^c = n C_\rho^2(\tilde{\theta}) / B_{\rho,\gamma}(\tilde{\theta}), \quad (3.44)$$

$$LM_\lambda^c = n C_\lambda^2(\tilde{\theta}) / B_{\lambda,\gamma}(\tilde{\theta}), \quad (3.45)$$

These variants are invalid under local parametric misspecification. Under the heteroskedastic case, we also have $J_{\rho\beta}(\tilde{\theta}) = 0$, implying that $B_{\rho,\gamma}(\tilde{\theta}) = K_{\rho\rho}(\tilde{\theta})$. Thus, LM_ρ^c can also be formulated with $K_{\rho\rho}(\tilde{\theta})$.

4. Monte Carlo Simulations

4.1. Design

In order to study the finite sample properties of the suggested tests, we design a Monte Carlo study. For the specification in (2.1), we consider two regressors $X = (X_1, X_2)$ with the parameter vector $(\beta_{01}, \beta_{02})' = (1, 1)'$. For X_1 and X_2 , we use the U.S. county-level data set of [21] on the 1980 presidential election: X_1 is the standardized value of log income per-capita and X_2 is the standardized value of log homeownership. This data set describes 3107 U.S. counties, and we use the first n observations in our Monte Carlo study. We consider the three specifications for the weights matrix considered by [22]: (i) the 49×49 contiguity based weights matrix generated for 48 US states and the District of Columbia, (ii) the 98×98 contiguity based weights matrix corresponding to five nearest neighbors of each of the 98 census tracts in Toledo, Ohio, and (iii) the 361×361 contiguity based weights matrix corresponding to whether the school districts are in the same county in Iowa in 2009. In our simulation, we set $W = M$. The spatial autoregressive parameters, λ_0 and ρ_0 , take values from $\{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$.

In generating disturbance terms, we consider two cases: (i) a homoskedastic case and (ii) a heteroskedastic case based on a skedastic function. In the homoskedastic case, we generate disturbance terms that have standard normal and chi-squared distributions: (i) $v_i \sim N(0, 1)$ and (ii) $v_i \sim (\chi_2^2 - 2)/2$. In the skedastic case, we generate the disturbance terms according to $v_i = \sigma_i \xi_i$, where $\sigma_i^2 = \exp(0.1 + 0.35 \cdot X_{i,1})$, and the innovation term ξ_i is generated according to (i) $\xi_i \sim N(0, 1)$, and (ii) $v_i \sim (\chi_2^2 - 2)/2$. We set the nominal size set to 0.05 and the number of repetitions is 1000.

In our simulation, we will also consider the test statistics suggested in [7] and [4]. More specifically, we consider the tests in the equations (3.2) and (3.4) of [7], and the tests in the equations (16) and (17) of [4]. Moreover, we also consider the robust test statistics that are derived in the LM framework in [2] under the assumption that the disturbance terms are homoskedastic.

4.2. Simulation Results

The simulation results are presented in Tables 1 through 6 for the normal distribution case.⁵ In these tables, (i) LM_λ is the test statistic in Theorem 1, (ii) LM_ρ is the test statistic in Theorem 2, (iii) LM_λ^h and LM_ρ^h are the test statistics given in Theorem 3, (iv) LM_ρ^B and LM_λ^B are the test statistics suggested in [7], (v) LM_λ^Z and LM_ρ^Z are the tests statistics suggested in [4], and (vi) LM_λ^A and LM_ρ^A are the test statistics suggested in [2]. The salient features of the results in Tables 1-6 are as follows.

1. The first row in each table shows the empirical size of the tests when there is no parametric misspecification. Our suggested tests, i.e., LM_ρ , LM_λ , LM_ρ^h and LM_λ^h , and those suggested in [2], i.e., LM_ρ^A and LM_λ^A , report size values that are close to the nominal value of 0.05 under both homoskedasticity and heteroskedasticity. The test statistics LM_ρ^B , LM_ρ^Z , LM_λ^B and LM_λ^Z are over-sized, especially LM_ρ^B and LM_ρ^Z .

2. The empirical size properties of tests for testing H_0^λ under parametric misspecification can be examined through the first panel of each table. In the local presence of ρ_0 , our suggested tests and those suggested in [2] perform better than other tests. As expected, LM_λ^Z and LM_λ^B are not robust to the local presence of ρ_0 .

3. The empirical size properties of tests for testing H_0^ρ under parametric misspecification can be examined under the cases where $\rho_0 = 0$ and $\lambda_0 \neq 0$. Both LM_ρ^h and LM_ρ^A perform relatively better than other tests in all cases. LM_ρ^Z and LM_ρ^B are over-sized in all cases as these tests are not robust to local parametric misspecification.

⁵ We do not present the simulation results for the chi-squared distribution case, since they are similar to the normal distribution case. These results are available from the authors upon request.

4. The simulation results in Tables 1-6 indicates that heteroskedasticity specified in the form of a skedastic function may not affect the performance of the robust tests suggested in [2].

5. The empirical power properties of test statistics for testing H_0^λ can be examined when λ_0 deviates from zero. When λ_0 is near zero, LM_λ^B and LM_λ^Z report relatively large empirical powers under both homoskedastic and heteroskedastic cases. As λ_0 gets larger, our suggested tests and those in [2] get larger empirical powers, and perform similar to LM_λ^B and LM_λ^Z . Similarly, the empirical power properties of test statistics for testing H_0^ρ can be examined when ρ_0 deviates from zero. Here, we also observe a similar pattern. That is, LM_λ^B and LM_λ^Z perform better than other tests under both homoskedastic and heteroskedastic cases only when ρ_0 is near to zero.

TABLE 1. Empirical size and power of tests: Homoskedastic disturbances and $n = 49$

λ_0	ρ_0	LM_p	LM_p^A	LM_p^B	LM_p^Z	LM_p^h	LM_λ	LM_λ^A	LM_λ^B	LM_λ^Z	LM_λ^h
0.0	0.0	0.049	0.044	0.168	0.171	0.051	0.050	0.052	0.075	0.081	0.050
0.0	0.1	0.048	0.061	0.190	0.211	0.051	0.057	0.060	0.092	0.096	0.057
0.0	0.2	0.056	0.113	0.312	0.342	0.091	0.060	0.059	0.137	0.149	0.060
0.0	0.3	0.055	0.197	0.460	0.486	0.150	0.058	0.057	0.205	0.215	0.058
0.0	0.4	0.063	0.335	0.656	0.688	0.258	0.082	0.086	0.337	0.349	0.082
0.0	0.5	0.079	0.496	0.805	0.835	0.397	0.100	0.108	0.525	0.541	0.100
0.0	0.6	0.133	0.655	0.915	0.928	0.531	0.124	0.137	0.705	0.717	0.124
0.1	0.0	0.075	0.051	0.185	0.207	0.052	0.098	0.104	0.157	0.162	0.098
0.1	0.1	0.061	0.068	0.261	0.302	0.054	0.095	0.105	0.206	0.220	0.095
0.1	0.2	0.057	0.126	0.477	0.513	0.086	0.093	0.102	0.340	0.354	0.093
0.1	0.3	0.063	0.211	0.638	0.670	0.152	0.116	0.121	0.457	0.478	0.116
0.1	0.4	0.064	0.342	0.780	0.806	0.242	0.134	0.146	0.607	0.630	0.134
0.1	0.5	0.105	0.483	0.891	0.909	0.356	0.174	0.184	0.772	0.783	0.174
0.1	0.6	0.192	0.644	0.957	0.966	0.493	0.175	0.185	0.858	0.872	0.175
0.2	0.0	0.119	0.048	0.295	0.331	0.042	0.238	0.248	0.389	0.405	0.238
0.2	0.1	0.108	0.079	0.430	0.488	0.056	0.251	0.278	0.502	0.517	0.251
0.2	0.2	0.084	0.123	0.611	0.661	0.083	0.262	0.277	0.628	0.645	0.262
0.2	0.3	0.090	0.209	0.760	0.796	0.139	0.264	0.278	0.706	0.718	0.264
0.2	0.4	0.100	0.328	0.872	0.893	0.226	0.270	0.287	0.815	0.822	0.270
0.2	0.5	0.167	0.485	0.937	0.949	0.330	0.264	0.281	0.893	0.898	0.264
0.2	0.6	0.304	0.622	0.973	0.983	0.444	0.276	0.297	0.937	0.942	0.276
0.3	0.0	0.230	0.053	0.451	0.523	0.048	0.505	0.533	0.733	0.745	0.505
0.3	0.1	0.191	0.067	0.597	0.657	0.045	0.488	0.509	0.776	0.787	0.488
0.3	0.2	0.148	0.124	0.760	0.798	0.075	0.471	0.498	0.852	0.858	0.471
0.3	0.3	0.167	0.192	0.852	0.886	0.124	0.478	0.492	0.904	0.908	0.478
0.3	0.4	0.220	0.302	0.918	0.946	0.185	0.451	0.475	0.943	0.949	0.451
0.3	0.5	0.306	0.456	0.972	0.983	0.300	0.425	0.447	0.961	0.964	0.425
0.3	0.6	0.469	0.570	0.986	0.991	0.386	0.401	0.421	0.983	0.984	0.401
0.4	0.0	0.343	0.049	0.632	0.695	0.041	0.781	0.801	0.928	0.932	0.781
0.4	0.1	0.324	0.061	0.758	0.826	0.043	0.782	0.801	0.957	0.959	0.782
0.4	0.2	0.297	0.117	0.855	0.898	0.060	0.716	0.749	0.965	0.970	0.716
0.4	0.3	0.308	0.189	0.922	0.950	0.112	0.673	0.698	0.981	0.982	0.673
0.4	0.4	0.388	0.282	0.961	0.986	0.160	0.644	0.658	0.987	0.988	0.644
0.4	0.5	0.487	0.409	0.985	0.992	0.234	0.604	0.627	0.988	0.989	0.604
0.4	0.6	0.657	0.509	0.994	0.998	0.329	0.534	0.553	0.996	0.997	0.534
0.5	0.0	0.509	0.043	0.756	0.843	0.038	0.932	0.945	0.992	0.993	0.932
0.5	0.1	0.495	0.050	0.853	0.904	0.029	0.921	0.931	0.996	0.996	0.921
0.5	0.2	0.475	0.088	0.905	0.946	0.049	0.892	0.903	0.994	0.994	0.892
0.5	0.3	0.510	0.138	0.951	0.977	0.074	0.851	0.866	0.997	0.997	0.851
0.5	0.4	0.600	0.217	0.972	0.989	0.110	0.808	0.822	0.996	0.997	0.808
0.5	0.5	0.680	0.333	0.989	0.994	0.185	0.737	0.760	0.997	0.998	0.737
0.5	0.6	0.830	0.423	0.992	0.997	0.243	0.663	0.682	0.998	0.999	0.663
0.6	0.0	0.712	0.034	0.873	0.943	0.029	0.989	0.992	1.000	1.000	0.989
0.6	0.1	0.723	0.038	0.929	0.974	0.022	0.980	0.985	1.000	1.000	0.980
0.6	0.2	0.741	0.060	0.949	0.982	0.033	0.963	0.965	1.000	1.000	0.963
0.6	0.3	0.753	0.104	0.975	0.993	0.045	0.939	0.952	1.000	1.000	0.939
0.6	0.4	0.803	0.172	0.989	0.995	0.076	0.912	0.922	1.000	1.000	0.912
0.6	0.5	0.872	0.248	0.989	0.997	0.113	0.849	0.869	1.000	1.000	0.849
0.6	0.6	0.934	0.312	0.995	0.999	0.155	0.781	0.801	1.000	1.000	0.781

TABLE 2. Empirical size and power of tests: Heteroskedastic disturbances and $n = 49$

λ_0	ρ_0	LM_p	LM_p^A	LM_p^B	LM_p^Z	LM_p^h	LM_λ	LM_λ^A	LM_λ^B	LM_λ^Z	LM_λ^h
0.0	0.0	0.045	0.046	0.176	0.188	0.046	0.051	0.061	0.074	0.079	0.051
0.0	0.1	0.056	0.066	0.211	0.223	0.052	0.053	0.071	0.089	0.093	0.053
0.0	0.2	0.058	0.121	0.338	0.349	0.090	0.059	0.072	0.129	0.134	0.059
0.0	0.3	0.058	0.213	0.485	0.496	0.150	0.054	0.065	0.193	0.196	0.054
0.0	0.4	0.082	0.350	0.673	0.698	0.257	0.076	0.092	0.320	0.328	0.076
0.0	0.5	0.090	0.519	0.828	0.835	0.411	0.093	0.119	0.500	0.506	0.093
0.0	0.6	0.116	0.671	0.926	0.934	0.531	0.114	0.137	0.673	0.685	0.114
0.1	0.0	0.067	0.052	0.206	0.219	0.044	0.102	0.117	0.150	0.153	0.102
0.1	0.1	0.052	0.074	0.296	0.313	0.053	0.096	0.120	0.202	0.207	0.096
0.1	0.2	0.053	0.134	0.491	0.510	0.087	0.095	0.111	0.325	0.334	0.095
0.1	0.3	0.058	0.225	0.666	0.688	0.161	0.113	0.131	0.448	0.453	0.113
0.1	0.4	0.064	0.356	0.797	0.817	0.247	0.128	0.154	0.589	0.599	0.128
0.1	0.5	0.089	0.502	0.903	0.916	0.356	0.168	0.194	0.757	0.761	0.168
0.1	0.6	0.147	0.653	0.964	0.969	0.500	0.162	0.184	0.847	0.856	0.162
0.2	0.0	0.108	0.045	0.324	0.341	0.043	0.229	0.276	0.386	0.392	0.229
0.2	0.1	0.098	0.091	0.476	0.502	0.057	0.258	0.297	0.508	0.514	0.258
0.2	0.2	0.075	0.137	0.639	0.664	0.083	0.260	0.303	0.612	0.624	0.260
0.2	0.3	0.075	0.215	0.789	0.807	0.137	0.255	0.291	0.701	0.705	0.255
0.2	0.4	0.087	0.341	0.883	0.898	0.229	0.271	0.303	0.802	0.807	0.271
0.2	0.5	0.135	0.501	0.950	0.960	0.322	0.267	0.299	0.887	0.891	0.267
0.2	0.6	0.247	0.646	0.979	0.984	0.441	0.271	0.302	0.935	0.940	0.271
0.3	0.0	0.210	0.053	0.489	0.523	0.045	0.509	0.568	0.730	0.736	0.509
0.3	0.1	0.168	0.070	0.635	0.670	0.044	0.494	0.545	0.778	0.783	0.494
0.3	0.2	0.113	0.128	0.787	0.807	0.067	0.482	0.531	0.851	0.851	0.482
0.3	0.3	0.128	0.203	0.864	0.890	0.126	0.476	0.526	0.904	0.905	0.476
0.3	0.4	0.167	0.315	0.936	0.951	0.187	0.462	0.503	0.942	0.943	0.462
0.3	0.5	0.232	0.465	0.975	0.981	0.303	0.426	0.463	0.962	0.963	0.426
0.3	0.6	0.396	0.585	0.986	0.988	0.384	0.406	0.439	0.982	0.983	0.406
0.4	0.0	0.313	0.050	0.661	0.700	0.041	0.779	0.821	0.933	0.935	0.779
0.4	0.1	0.284	0.060	0.789	0.828	0.039	0.791	0.824	0.964	0.964	0.791
0.4	0.2	0.259	0.114	0.865	0.899	0.059	0.728	0.773	0.963	0.967	0.728
0.4	0.3	0.254	0.191	0.928	0.947	0.108	0.688	0.730	0.980	0.980	0.688
0.4	0.4	0.339	0.293	0.973	0.985	0.158	0.662	0.700	0.987	0.988	0.662
0.4	0.5	0.399	0.420	0.987	0.991	0.234	0.617	0.657	0.989	0.990	0.617
0.4	0.6	0.594	0.520	0.996	0.998	0.323	0.545	0.582	0.996	0.998	0.545
0.5	0.0	0.475	0.038	0.790	0.845	0.036	0.943	0.958	0.995	0.995	0.943
0.5	0.1	0.446	0.048	0.879	0.914	0.028	0.927	0.939	0.996	0.997	0.927
0.5	0.2	0.417	0.089	0.919	0.949	0.041	0.906	0.924	0.994	0.995	0.906
0.5	0.3	0.444	0.142	0.962	0.981	0.067	0.864	0.888	0.997	0.997	0.864
0.5	0.4	0.525	0.213	0.981	0.991	0.110	0.830	0.854	0.998	0.998	0.830
0.5	0.5	0.613	0.340	0.991	0.995	0.178	0.762	0.785	0.998	0.999	0.762
0.5	0.6	0.769	0.430	0.996	0.998	0.234	0.680	0.711	0.999	0.999	0.680
0.6	0.0	0.667	0.031	0.899	0.948	0.025	0.992	0.994	1.000	1.000	0.992
0.6	0.1	0.668	0.035	0.941	0.979	0.021	0.984	0.991	1.000	1.000	0.984
0.6	0.2	0.686	0.059	0.960	0.984	0.028	0.971	0.976	1.000	1.000	0.971
0.6	0.3	0.700	0.100	0.979	0.992	0.041	0.951	0.963	1.000	1.000	0.951
0.6	0.4	0.729	0.166	0.993	0.996	0.070	0.922	0.935	1.000	1.000	0.922
0.6	0.5	0.828	0.251	0.990	0.997	0.104	0.868	0.884	1.000	1.000	0.868
0.6	0.6	0.904	0.312	0.997	1.000	0.160	0.791	0.822	1.000	1.000	0.791

TABLE 3. Empirical size and power of tests: Homoskedastic disturbances and $n = 98$

λ_0	ρ_0	LM_p	LM_p^A	LM_p^B	LM_p^Z	LM_p^h	LM_λ	LM_λ^A	LM_λ^B	LM_λ^Z	LM_λ^h
0.0	0.0	0.052	0.046	0.158	0.153	0.049	0.058	0.053	0.077	0.074	0.058
0.0	0.1	0.061	0.086	0.208	0.226	0.064	0.060	0.062	0.099	0.104	0.060
0.0	0.2	0.071	0.172	0.379	0.412	0.137	0.061	0.063	0.172	0.179	0.061
0.0	0.3	0.115	0.325	0.622	0.649	0.285	0.099	0.100	0.310	0.321	0.099
0.0	0.4	0.129	0.561	0.844	0.862	0.500	0.101	0.102	0.506	0.523	0.101
0.0	0.5	0.141	0.745	0.957	0.961	0.694	0.143	0.149	0.733	0.743	0.143
0.0	0.6	0.184	0.908	0.990	0.992	0.877	0.208	0.214	0.890	0.896	0.208
0.1	0.0	0.073	0.046	0.198	0.223	0.050	0.119	0.130	0.194	0.205	0.119
0.1	0.1	0.075	0.099	0.380	0.410	0.082	0.142	0.146	0.316	0.331	0.142
0.1	0.2	0.068	0.205	0.614	0.647	0.173	0.139	0.148	0.459	0.476	0.139
0.1	0.3	0.085	0.363	0.814	0.836	0.300	0.183	0.196	0.623	0.637	0.183
0.1	0.4	0.106	0.611	0.925	0.936	0.539	0.183	0.188	0.778	0.786	0.183
0.1	0.5	0.129	0.789	0.985	0.986	0.716	0.225	0.238	0.896	0.903	0.225
0.1	0.6	0.213	0.919	0.998	0.998	0.883	0.270	0.273	0.973	0.975	0.270
0.2	0.0	0.164	0.065	0.406	0.450	0.056	0.374	0.389	0.569	0.586	0.374
0.2	0.1	0.119	0.118	0.630	0.665	0.095	0.378	0.400	0.700	0.709	0.378
0.2	0.2	0.103	0.240	0.821	0.842	0.201	0.365	0.388	0.804	0.813	0.365
0.2	0.3	0.092	0.435	0.927	0.940	0.371	0.378	0.384	0.870	0.875	0.378
0.2	0.4	0.118	0.642	0.981	0.984	0.567	0.365	0.388	0.935	0.942	0.365
0.2	0.5	0.167	0.822	0.996	0.997	0.767	0.385	0.395	0.979	0.982	0.385
0.2	0.6	0.300	0.932	1.000	1.000	0.897	0.383	0.390	0.989	0.991	0.383
0.3	0.0	0.258	0.070	0.696	0.739	0.062	0.718	0.732	0.893	0.898	0.718
0.3	0.1	0.197	0.131	0.858	0.876	0.102	0.716	0.734	0.944	0.946	0.716
0.3	0.2	0.149	0.281	0.931	0.943	0.212	0.670	0.688	0.958	0.963	0.670
0.3	0.3	0.147	0.492	0.973	0.981	0.412	0.635	0.655	0.978	0.980	0.635
0.3	0.4	0.174	0.693	0.995	0.998	0.606	0.611	0.620	0.988	0.990	0.611
0.3	0.5	0.257	0.848	1.000	1.000	0.786	0.583	0.588	0.997	0.998	0.583
0.3	0.6	0.449	0.956	1.000	1.000	0.917	0.544	0.556	0.998	0.998	0.544
0.4	0.0	0.339	0.095	0.925	0.945	0.066	0.933	0.943	0.993	0.994	0.933
0.4	0.1	0.302	0.185	0.971	0.982	0.141	0.917	0.928	0.995	0.996	0.917
0.4	0.2	0.245	0.371	0.992	0.996	0.284	0.892	0.904	0.996	0.996	0.892
0.4	0.3	0.246	0.524	0.996	0.998	0.428	0.863	0.871	0.998	0.998	0.863
0.4	0.4	0.305	0.735	1.000	1.000	0.655	0.803	0.807	0.998	0.998	0.803
0.4	0.5	0.426	0.874	1.000	1.000	0.799	0.763	0.770	1.000	1.000	0.763
0.4	0.6	0.620	0.955	1.000	1.000	0.914	0.679	0.691	1.000	1.000	0.679
0.5	0.0	0.480	0.135	0.993	0.996	0.102	0.993	0.994	1.000	1.000	0.993
0.5	0.1	0.450	0.274	0.996	0.997	0.188	0.990	0.993	1.000	1.000	0.990
0.5	0.2	0.427	0.438	0.999	1.000	0.341	0.980	0.981	1.000	1.000	0.980
0.5	0.3	0.443	0.613	0.999	0.999	0.516	0.966	0.969	1.000	1.000	0.966
0.5	0.4	0.495	0.773	1.000	1.000	0.665	0.936	0.938	1.000	1.000	0.936
0.5	0.5	0.623	0.881	1.000	1.000	0.802	0.886	0.890	1.000	1.000	0.886
0.5	0.6	0.814	0.951	1.000	1.000	0.896	0.804	0.809	1.000	1.000	0.804
0.6	0.0	0.690	0.216	1.000	1.000	0.150	1.000	1.000	1.000	1.000	1.000
0.6	0.1	0.670	0.319	1.000	1.000	0.225	1.000	1.000	1.000	1.000	1.000
0.6	0.2	0.674	0.482	1.000	1.000	0.349	0.997	0.998	1.000	1.000	0.997
0.6	0.3	0.724	0.623	1.000	1.000	0.498	0.994	0.995	1.000	1.000	0.994
0.6	0.4	0.762	0.785	1.000	1.000	0.666	0.970	0.973	1.000	1.000	0.970
0.6	0.5	0.852	0.869	1.000	1.000	0.772	0.945	0.951	1.000	1.000	0.945
0.6	0.6	0.935	0.934	1.000	1.000	0.852	0.886	0.887	1.000	1.000	0.886

TABLE 4. Empirical size and power of tests: Heteroskedastic disturbances and $n = 98$

λ_0	ρ_0	LM_ρ	LM_ρ^A	LM_ρ^B	LM_ρ^Z	LM_ρ^h	LM_λ	LM_λ^A	LM_λ^B	LM_λ^Z	LM_λ^h
0.0	0.0	0.054	0.057	0.158	0.161	0.052	0.058	0.086	0.073	0.071	0.058
0.0	0.1	0.063	0.090	0.220	0.230	0.062	0.063	0.080	0.092	0.092	0.063
0.0	0.2	0.073	0.182	0.391	0.403	0.131	0.061	0.081	0.163	0.168	0.061
0.0	0.3	0.120	0.331	0.630	0.645	0.283	0.096	0.122	0.284	0.288	0.096
0.0	0.4	0.148	0.566	0.847	0.858	0.487	0.093	0.127	0.460	0.469	0.093
0.0	0.5	0.160	0.758	0.960	0.964	0.686	0.134	0.169	0.691	0.701	0.134
0.0	0.6	0.183	0.911	0.990	0.990	0.872	0.186	0.223	0.867	0.871	0.186
0.1	0.0	0.071	0.054	0.207	0.215	0.050	0.117	0.154	0.190	0.193	0.117
0.1	0.1	0.072	0.102	0.382	0.402	0.079	0.129	0.173	0.301	0.306	0.129
0.1	0.2	0.072	0.220	0.624	0.635	0.172	0.132	0.169	0.419	0.425	0.132
0.1	0.3	0.093	0.362	0.818	0.831	0.293	0.183	0.227	0.591	0.601	0.183
0.1	0.4	0.113	0.612	0.930	0.934	0.532	0.164	0.209	0.749	0.753	0.164
0.1	0.5	0.128	0.778	0.982	0.983	0.709	0.222	0.255	0.882	0.885	0.222
0.1	0.6	0.171	0.923	0.998	0.998	0.879	0.254	0.287	0.965	0.966	0.254
0.2	0.0	0.143	0.066	0.421	0.442	0.057	0.350	0.413	0.539	0.546	0.350
0.2	0.1	0.110	0.124	0.642	0.659	0.098	0.357	0.433	0.670	0.675	0.357
0.2	0.2	0.089	0.244	0.823	0.836	0.198	0.344	0.401	0.770	0.777	0.344
0.2	0.3	0.087	0.430	0.927	0.936	0.364	0.362	0.412	0.849	0.852	0.362
0.2	0.4	0.107	0.640	0.981	0.983	0.557	0.352	0.405	0.919	0.920	0.352
0.2	0.5	0.136	0.814	0.997	0.997	0.758	0.370	0.414	0.971	0.974	0.370
0.2	0.6	0.233	0.930	1.000	1.000	0.895	0.368	0.404	0.988	0.988	0.368
0.3	0.0	0.224	0.071	0.707	0.735	0.060	0.683	0.741	0.875	0.879	0.683
0.3	0.1	0.163	0.134	0.855	0.869	0.102	0.693	0.754	0.934	0.936	0.693
0.3	0.2	0.125	0.283	0.939	0.948	0.220	0.652	0.707	0.946	0.949	0.652
0.3	0.3	0.129	0.486	0.973	0.978	0.410	0.611	0.669	0.975	0.975	0.611
0.3	0.4	0.130	0.691	0.995	0.996	0.597	0.600	0.644	0.984	0.985	0.600
0.3	0.5	0.204	0.847	1.000	1.000	0.785	0.565	0.598	0.994	0.993	0.565
0.3	0.6	0.369	0.950	1.000	1.000	0.915	0.534	0.564	0.998	0.998	0.534
0.4	0.0	0.302	0.101	0.937	0.947	0.074	0.925	0.943	0.991	0.992	0.925
0.4	0.1	0.240	0.184	0.977	0.982	0.150	0.907	0.931	0.994	0.994	0.907
0.4	0.2	0.197	0.381	0.995	0.996	0.306	0.886	0.914	0.995	0.995	0.886
0.4	0.3	0.200	0.521	0.997	0.998	0.441	0.857	0.881	0.998	0.998	0.857
0.4	0.4	0.239	0.732	1.000	1.000	0.663	0.799	0.828	0.998	0.998	0.799
0.4	0.5	0.342	0.865	1.000	1.000	0.803	0.765	0.782	0.999	1.000	0.765
0.4	0.6	0.526	0.951	1.000	1.000	0.912	0.685	0.711	1.000	1.000	0.685
0.5	0.0	0.417	0.137	0.994	0.996	0.108	0.991	0.995	1.000	1.000	0.991
0.5	0.1	0.385	0.267	0.996	0.997	0.198	0.990	0.995	1.000	1.000	0.990
0.5	0.2	0.353	0.429	0.999	0.999	0.356	0.977	0.985	1.000	1.000	0.977
0.5	0.3	0.364	0.608	1.000	1.000	0.523	0.968	0.974	1.000	1.000	0.968
0.5	0.4	0.420	0.769	1.000	1.000	0.669	0.935	0.944	1.000	1.000	0.935
0.5	0.5	0.540	0.877	1.000	1.000	0.803	0.890	0.904	1.000	1.000	0.890
0.5	0.6	0.745	0.948	1.000	1.000	0.896	0.807	0.826	1.000	1.000	0.807
0.6	0.0	0.646	0.212	1.000	1.000	0.178	1.000	1.000	1.000	1.000	1.000
0.6	0.1	0.606	0.314	1.000	1.000	0.247	1.000	1.000	1.000	1.000	1.000
0.6	0.2	0.599	0.479	1.000	1.000	0.372	0.998	0.999	1.000	1.000	0.998
0.6	0.3	0.643	0.620	1.000	1.000	0.526	0.995	0.998	1.000	1.000	0.995
0.6	0.4	0.688	0.773	1.000	1.000	0.672	0.975	0.979	1.000	1.000	0.975
0.6	0.5	0.793	0.861	1.000	1.000	0.769	0.955	0.960	1.000	1.000	0.955
0.6	0.6	0.895	0.927	1.000	1.000	0.844	0.897	0.904	1.000	1.000	0.897

TABLE 5. Empirical size and power of tests: Homoskedastic disturbances and $n = 361$

λ_0	ρ_0	LM_p	LM_p^A	LM_p^B	LM_p^Z	LM_p^h	LM_λ	LM_λ^A	LM_λ^B	LM_λ^Z	LM_λ^h
0.0	0.0	0.059	0.043	0.173	0.172	0.043	0.061	0.061	0.092	0.091	0.061
0.0	0.1	0.060	0.148	0.417	0.416	0.135	0.061	0.059	0.211	0.211	0.061
0.0	0.2	0.071	0.456	0.847	0.848	0.424	0.069	0.072	0.518	0.518	0.069
0.0	0.3	0.098	0.780	0.985	0.985	0.753	0.096	0.102	0.846	0.847	0.096
0.0	0.4	0.144	0.957	1.000	1.000	0.944	0.110	0.108	0.979	0.979	0.110
0.0	0.5	0.311	0.996	1.000	1.000	0.995	0.145	0.144	0.999	0.999	0.145
0.0	0.6	0.707	1.000	1.000	1.000	1.000	0.189	0.188	1.000	1.000	0.189
0.1	0.0	0.240	0.055	0.431	0.431	0.054	0.252	0.259	0.495	0.491	0.252
0.1	0.1	0.238	0.171	0.860	0.860	0.148	0.249	0.257	0.794	0.794	0.249
0.1	0.2	0.278	0.497	0.984	0.984	0.472	0.254	0.253	0.951	0.951	0.254
0.1	0.3	0.390	0.815	0.999	0.999	0.788	0.288	0.291	0.994	0.994	0.288
0.1	0.4	0.533	0.970	1.000	1.000	0.960	0.279	0.283	1.000	1.000	0.279
0.1	0.5	0.791	0.997	1.000	1.000	0.995	0.294	0.293	1.000	1.000	0.294
0.1	0.6	0.963	1.000	1.000	1.000	1.000	0.311	0.311	1.000	1.000	0.311
0.2	0.0	0.693	0.061	0.873	0.874	0.054	0.702	0.708	0.958	0.959	0.702
0.2	0.1	0.699	0.206	0.991	0.991	0.182	0.697	0.704	0.997	0.997	0.697
0.2	0.2	0.731	0.543	1.000	1.000	0.509	0.653	0.658	0.999	0.999	0.653
0.2	0.3	0.829	0.851	1.000	1.000	0.819	0.656	0.664	1.000	1.000	0.656
0.2	0.4	0.908	0.976	1.000	1.000	0.966	0.597	0.608	1.000	1.000	0.597
0.2	0.5	0.982	0.999	1.000	1.000	0.996	0.585	0.594	1.000	1.000	0.585
0.2	0.6	0.997	1.000	1.000	1.000	1.000	0.522	0.523	1.000	1.000	0.522
0.3	0.0	0.961	0.073	0.997	0.997	0.059	0.960	0.960	1.000	1.000	0.960
0.3	0.1	0.962	0.259	1.000	1.000	0.231	0.943	0.946	1.000	1.000	0.943
0.3	0.2	0.977	0.607	1.000	1.000	0.550	0.932	0.941	1.000	1.000	0.932
0.3	0.3	0.989	0.868	1.000	1.000	0.839	0.916	0.919	1.000	1.000	0.916
0.3	0.4	0.996	0.986	1.000	1.000	0.967	0.852	0.855	1.000	1.000	0.852
0.3	0.5	1.000	1.000	1.000	1.000	0.998	0.818	0.825	1.000	1.000	0.818
0.3	0.6	1.000	1.000	1.000	1.000	1.000	0.760	0.765	1.000	1.000	0.760
0.4	0.0	1.000	0.093	1.000	1.000	0.076	0.998	0.999	1.000	1.000	0.998
0.4	0.1	1.000	0.337	1.000	1.000	0.281	0.999	1.000	1.000	1.000	0.999
0.4	0.2	1.000	0.673	1.000	1.000	0.590	0.997	0.998	1.000	1.000	0.997
0.4	0.3	1.000	0.898	1.000	1.000	0.842	0.986	0.988	1.000	1.000	0.986
0.4	0.4	1.000	0.988	1.000	1.000	0.968	0.968	0.971	1.000	1.000	0.968
0.4	0.5	1.000	0.999	1.000	1.000	0.992	0.949	0.954	1.000	1.000	0.949
0.4	0.6	1.000	1.000	1.000	1.000	0.999	0.875	0.883	1.000	1.000	0.875
0.5	0.0	1.000	0.167	1.000	1.000	0.120	1.000	1.000	1.000	1.000	1.000
0.5	0.1	1.000	0.419	1.000	1.000	0.315	1.000	1.000	1.000	1.000	1.000
0.5	0.2	1.000	0.725	1.000	1.000	0.613	0.999	1.000	1.000	1.000	0.999
0.5	0.3	1.000	0.927	1.000	1.000	0.852	0.999	0.999	1.000	1.000	0.999
0.5	0.4	1.000	0.983	1.000	1.000	0.957	0.996	0.995	1.000	1.000	0.996
0.5	0.5	1.000	0.997	1.000	1.000	0.984	0.983	0.984	1.000	1.000	0.983
0.5	0.6	1.000	1.000	1.000	1.000	0.995	0.962	0.966	1.000	1.000	0.962
0.6	0.0	1.000	0.232	1.000	1.000	0.150	1.000	1.000	1.000	1.000	1.000
0.6	0.1	1.000	0.484	1.000	1.000	0.339	1.000	1.000	1.000	1.000	1.000
0.6	0.2	1.000	0.727	1.000	1.000	0.566	1.000	1.000	1.000	1.000	1.000
0.6	0.3	1.000	0.895	1.000	1.000	0.783	1.000	1.000	1.000	1.000	1.000
0.6	0.4	1.000	0.971	1.000	1.000	0.917	0.999	0.999	1.000	1.000	0.999
0.6	0.5	1.000	0.991	1.000	1.000	0.952	0.998	0.998	1.000	1.000	0.998
0.6	0.6	1.000	0.997	1.000	1.000	0.976	0.985	0.985	1.000	1.000	0.985

TABLE 6. Empirical size and power of tests: Heteroskedastic disturbances and $n = 361$

λ_0	ρ_0	LM_ρ	LM_ρ^A	LM_ρ^B	LM_ρ^Z	LM_ρ^h	LM_λ	LM_λ^A	LM_λ^B	LM_λ^Z	LM_λ^h
0.0	0.0	0.059	0.044	0.172	0.174	0.049	0.059	0.055	0.099	0.098	0.059
0.0	0.1	0.059	0.139	0.418	0.415	0.131	0.058	0.058	0.224	0.222	0.058
0.0	0.2	0.091	0.419	0.853	0.851	0.398	0.073	0.073	0.558	0.555	0.073
0.0	0.3	0.165	0.742	0.988	0.988	0.732	0.099	0.090	0.873	0.871	0.099
0.0	0.4	0.313	0.939	1.000	1.000	0.929	0.108	0.102	0.986	0.986	0.108
0.0	0.5	0.611	0.993	1.000	1.000	0.991	0.148	0.141	1.000	1.000	0.148
0.0	0.6	0.915	1.000	1.000	1.000	1.000	0.187	0.186	1.000	1.000	0.187
0.1	0.0	0.278	0.051	0.432	0.432	0.054	0.231	0.228	0.483	0.482	0.231
0.1	0.1	0.332	0.163	0.859	0.860	0.151	0.225	0.223	0.804	0.803	0.225
0.1	0.2	0.420	0.462	0.984	0.984	0.450	0.237	0.230	0.958	0.957	0.237
0.1	0.3	0.609	0.782	0.998	0.998	0.771	0.268	0.259	0.995	0.995	0.268
0.1	0.4	0.790	0.953	1.000	1.000	0.950	0.261	0.263	1.000	1.000	0.261
0.1	0.5	0.933	0.994	1.000	1.000	0.992	0.273	0.271	1.000	1.000	0.273
0.1	0.6	0.997	1.000	1.000	1.000	1.000	0.304	0.294	1.000	1.000	0.304
0.2	0.0	0.751	0.054	0.876	0.876	0.053	0.645	0.652	0.956	0.955	0.645
0.2	0.1	0.820	0.192	0.990	0.991	0.185	0.642	0.639	0.994	0.993	0.642
0.2	0.2	0.855	0.500	1.000	1.000	0.483	0.611	0.601	0.999	0.999	0.611
0.2	0.3	0.943	0.812	1.000	1.000	0.790	0.596	0.604	1.000	1.000	0.596
0.2	0.4	0.978	0.964	1.000	1.000	0.961	0.558	0.558	1.000	1.000	0.558
0.2	0.5	0.999	0.998	1.000	1.000	0.994	0.548	0.554	1.000	1.000	0.548
0.2	0.6	1.000	1.000	1.000	1.000	1.000	0.491	0.494	1.000	1.000	0.491
0.3	0.0	0.979	0.061	0.997	0.997	0.057	0.942	0.937	1.000	0.999	0.942
0.3	0.1	0.987	0.234	1.000	1.000	0.218	0.916	0.918	1.000	1.000	0.916
0.3	0.2	0.992	0.548	1.000	1.000	0.520	0.904	0.906	1.000	1.000	0.904
0.3	0.3	0.997	0.839	1.000	1.000	0.815	0.875	0.877	1.000	1.000	0.875
0.3	0.4	0.999	0.972	1.000	1.000	0.958	0.828	0.821	1.000	1.000	0.828
0.3	0.5	1.000	0.999	1.000	1.000	0.997	0.781	0.787	1.000	1.000	0.781
0.3	0.6	1.000	1.000	1.000	1.000	1.000	0.721	0.719	1.000	1.000	0.721
0.4	0.0	1.000	0.086	1.000	1.000	0.075	0.995	0.995	1.000	1.000	0.995
0.4	0.1	1.000	0.291	1.000	1.000	0.260	0.997	0.997	1.000	1.000	0.997
0.4	0.2	1.000	0.619	1.000	1.000	0.560	0.991	0.991	1.000	1.000	0.991
0.4	0.3	1.000	0.856	1.000	1.000	0.823	0.976	0.976	1.000	1.000	0.976
0.4	0.4	1.000	0.975	1.000	1.000	0.956	0.951	0.956	1.000	1.000	0.951
0.4	0.5	1.000	0.994	1.000	1.000	0.987	0.923	0.922	1.000	1.000	0.923
0.4	0.6	1.000	0.999	1.000	1.000	0.999	0.837	0.843	1.000	1.000	0.837
0.5	0.0	1.000	0.136	1.000	1.000	0.112	1.000	1.000	1.000	1.000	1.000
0.5	0.1	1.000	0.358	1.000	1.000	0.288	1.000	1.000	1.000	1.000	1.000
0.5	0.2	1.000	0.659	1.000	1.000	0.582	0.999	0.999	1.000	1.000	0.999
0.5	0.3	1.000	0.884	1.000	1.000	0.829	0.998	0.998	1.000	1.000	0.998
0.5	0.4	1.000	0.975	1.000	1.000	0.945	0.991	0.990	1.000	1.000	0.991
0.5	0.5	1.000	0.993	1.000	1.000	0.978	0.971	0.972	1.000	1.000	0.971
0.5	0.6	1.000	0.999	1.000	1.000	0.994	0.934	0.940	1.000	1.000	0.934
0.6	0.0	1.000	0.172	1.000	1.000	0.132	1.000	1.000	1.000	1.000	1.000
0.6	0.1	1.000	0.407	1.000	1.000	0.318	1.000	1.000	1.000	1.000	1.000
0.6	0.2	1.000	0.657	1.000	1.000	0.534	1.000	1.000	1.000	1.000	1.000
0.6	0.3	1.000	0.846	1.000	1.000	0.753	1.000	1.000	1.000	1.000	1.000
0.6	0.4	1.000	0.953	1.000	1.000	0.898	0.998	0.998	1.000	1.000	0.998
0.6	0.5	1.000	0.985	1.000	1.000	0.945	0.992	0.992	1.000	1.000	0.992
0.6	0.6	1.000	0.992	1.000	1.000	0.968	0.972	0.973	1.000	1.000	0.972

5. An Empirical Illustration

In this section, we illustrate the use of our proposed test statistic. To this end, we use an example from [21] on the US presidential election in 1980. The dataset contains variables on the election results and county characteristics for 3107 US counties. We consider the following regression model

$$\ln(\text{PR_VOTES}) = \beta_0 + \lambda W \ln(\text{PR_VOTES}) + \beta_1 \ln(\text{POP}) + \beta_2 \ln(\text{EDUC}) + \beta_3 \ln(\text{HOUSE}) + \beta_4 \ln(\text{INC}) + U, \quad U = \rho WU + V. \tag{5.1}$$

The outcome variable is the natural log of the proportion of votes cast for both candidates in the 1980 presidential election (PR_VOTES). The explanatory variables are the natural log of the population in each county of eighteen years of age or older (POP), the natural log of the population in each county with a 12th grade or higher education (EDUC), number of owner-occupied housing units (HOUSE), and the aggregate income (INC). The spatial weights matrix W is the delaunay contiguity based weights matrix constructed using the latitudes and longitudes of the counties (see [17] for the details).

TABLE 7. The tests results

LM_ρ	LM_ρ^h	LM_ρ^B	LM_ρ^Z	LM_ρ^A	LM_λ	LM_λ^h	LM_λ^B	LM_λ^Z	LM_λ^A
198.766	5.598	27.740	27.767	530.488	19.036	0.168	25.301	25.279	26.831

Our goal is to test the presence of λ and ρ in (5.1). The test statistics' values are presented in Table 7. In this table, (i) LM_λ is the test statistic in Theorem 1, (ii) LM_ρ is the test statistic in Theorem 2, (iii) LM_λ^h and LM_ρ^h are the test statistics given in Theorem 3, (iv) LM_ρ^B and LM_λ^B are the test statistics suggested in [7], (v) LM_λ^Z and LM_ρ^Z are the tests statistics suggested in [4], and (vi) LM_λ^A and LM_ρ^A are the test statistics suggested in [2].

We observe that all tests show strong statistical evidence for the existence of spatial dependence in the error term.⁶ The lowest value is observed for LM_ρ^h which still rejects the null hypothesis of no spatial dependence at the 0.05 significance level. Similarly, all tests, except LM_λ^h , show strong statistical evidence for the existence of spatial dependence in the outcome variable at the 0.05 significance level. The lowest value is observed for LM_λ^h which fails to reject the null hypothesis of no spatial dependence at the conventional significance level of 0.05. Given that an unknown form of heteroskedasticity is likely to be present in the observational cross-sectional dataset, the dichotomy between LM_λ^h and the rest of the test statistics is important, and the empirical modeling accounting for spatial dependence needs to consider estimating the nested null specification as a robustness check.

6. Conclusion

In this paper, we proposed the OPG variants of the LM test statistic for testing spatial dependence in spatial models with homoskedastic disturbance terms and in spatial models with heteroskedastic disturbance terms. Our OPG tests for testing one type of spatial dependence (the spatial lag in the dependent variable or the spatial lag in the disturbance term) are valid whether or not the other type of spatial dependence is present. We showed how such robust OPG tests can be systematically constructed in the quasi maximum likelihood (QML) framework. We derived the asymptotic distributions of the suggested tests under the null and local alternative hypotheses.

⁶ Note that our suggested test statistics and those suggested by [2] have asymptotic χ_1^2 distribution. The critical value based on χ_1^2 at the 0.05 significance level is 3.841. The other tests, those suggested by [7] and [4], have asymptotic $N(0, 1)$ distribution. The critical value based on $N(0, 1)$ at the 0.05 significance level is 1.96.

Our suggested tests are simple to compute, since they only require the OLS estimates from a linear regression model.

In a Monte Carlo study, we investigated the finite sample properties of our tests along with some alternative tests suggested in the literature. For testing the presence of spatial lag term, our simulation results showed that the suggested test statistic (LM_λ and LM_λ^h) has good size and power properties under both homoskedasticity and heteroskedasticity. Our results showed that LM_ρ^h has a satisfactory performance in finite samples. The results also indicated that the robust test statistics suggested in [2] may perform well under local parametric misspecification. Moreover, heteroskedasticity specified in the form of a skedastic function seems to be not affecting the performance of these tests. The simulation results also showed that the test statistics suggested in [7] and [4] can be over-sized under local parametric misspecification.

In future studies, our testing approach can be extended to other variants of spatial models. First, our approach can be easily extended to the cross-sectional spatial models that have higher order spatial lags in the dependent and the disturbance terms. Second, our approach can be used to develop similar tests for testing the presence of spatial dependence in the static and dynamic spatial panel data models. Finally, our testing approach can be considered for the matrix exponential spatial models suggested in the literature. All of these extensions can be explored in future studies.

Appendix

In this section, we provide only the proof of Theorem 1. Other theorems can be proved similarly, so we omit their proofs. Consider the mean value expansions of $\sqrt{n}S_\lambda(\tilde{\theta})$, $\sqrt{n}S_\rho(\tilde{\theta})$ and $\sqrt{n}S_\gamma(\tilde{\theta})$ around θ_0 when both H_a^λ and H_a^ρ hold:

$$\sqrt{n}S_\lambda(\tilde{\theta}) = \sqrt{n}S_\lambda(\theta_0) - \frac{\partial S_\lambda(\tilde{\theta})}{\partial \lambda} \delta_\lambda - \frac{\partial S_\lambda(\tilde{\theta})}{\partial \rho} \delta_\rho + \frac{\partial S_\lambda(\tilde{\theta})}{\partial \gamma'} \sqrt{n}(\tilde{\gamma} - \gamma_0), \quad (6.1)$$

$$\sqrt{n}S_\rho(\tilde{\theta}) = \sqrt{n}S_\rho(\theta_0) - \frac{\partial S_\rho(\tilde{\theta})}{\partial \lambda} \delta_\lambda - \frac{\partial S_\rho(\tilde{\theta})}{\partial \rho} \delta_\rho + \frac{\partial S_\rho(\tilde{\theta})}{\partial \gamma'} \sqrt{n}(\tilde{\gamma} - \gamma_0), \quad (6.2)$$

$$\sqrt{n}S_\gamma(\tilde{\theta}) = \sqrt{n}S_\gamma(\theta_0) - \frac{\partial S_\gamma(\tilde{\theta})}{\partial \lambda} \delta_\lambda - \frac{\partial S_\gamma(\tilde{\theta})}{\partial \rho} \delta_\rho + \frac{\partial S_\gamma(\tilde{\theta})}{\partial \gamma'} \sqrt{n}(\tilde{\gamma} - \gamma_0). \quad (6.3)$$

Our Assumption 2 ensures that

$$\sqrt{n}S_\lambda(\tilde{\theta}) = \sqrt{n}S_\lambda(\theta_0) + J_{\lambda\lambda}(\theta_0)\delta_\lambda + J_{\lambda\rho}(\theta_0)\delta_\rho - J_{\lambda\gamma}(\theta_0)\sqrt{n}(\tilde{\gamma} - \gamma_0) + o_p(1), \quad (6.4)$$

$$\sqrt{n}S_\rho(\tilde{\theta}) = \sqrt{n}S_\rho(\theta_0) + J_{\rho\lambda}(\theta_0)\delta_\lambda + J_{\rho\rho}(\theta_0)\delta_\rho - J_{\rho\gamma}(\theta_0)\sqrt{n}(\tilde{\gamma} - \gamma_0) + o_p(1), \quad (6.5)$$

$$\sqrt{n}S_\gamma(\tilde{\theta}) = \sqrt{n}S_\gamma(\theta_0) + J_{\gamma\lambda}(\theta_0)\delta_\lambda + J_{\gamma\rho}(\theta_0)\delta_\rho - J_{\gamma\gamma}(\theta_0)\sqrt{n}(\tilde{\gamma} - \gamma_0) + o_p(1), \quad (6.6)$$

Note that $\sqrt{n}S_\gamma(\tilde{\theta}) = 0$ holds in (6.6) by definition. Then, solving (6.6) for $\sqrt{n}(\tilde{\gamma} - \gamma_0)$ and substituting the resulting equation into (6.4) and (6.5), we obtain

$$\sqrt{n}S_\lambda(\tilde{\theta}) = (1, -J_{\lambda\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)) \begin{pmatrix} \sqrt{n}S_\lambda(\theta_0) \\ \sqrt{n}S_\gamma(\theta_0) \end{pmatrix} + J_{\lambda\gamma}(\theta_0)\delta_\lambda + J_{\lambda\rho\gamma}(\theta_0)\delta_\rho + o_p(1), \quad (6.7)$$

$$\sqrt{n}S_\rho(\tilde{\theta}) = (1, -J_{\rho\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)) \begin{pmatrix} \sqrt{n}S_\rho(\theta_0) \\ \sqrt{n}S_\gamma(\theta_0) \end{pmatrix} + J_{\rho\gamma}(\theta_0)\delta_\rho + J_{\rho\lambda\gamma}(\theta_0)\delta_\lambda + o_p(1). \quad (6.8)$$

We first show how the adjusted score function that has a zero asymptotic mean can be derived. We then determine the asymptotic distribution of the adjusted score function. The asymptotic distribution of $\sqrt{n}S_\lambda(\tilde{\theta})$ can be determined from (6.7) by using the asymptotic normality of score functions given in Assumption 2. Thus, it follows that

$$\sqrt{n}S_\lambda(\tilde{\theta}) \xrightarrow{d} N[J_{\lambda\gamma}(\theta_0)\delta_\lambda + J_{\lambda\rho\gamma}(\theta_0)\delta_\rho, B_{\lambda\gamma}(\theta_0)], \quad (6.9)$$

where

$$B_{\lambda\cdot\gamma}(\theta_0) = K_{\lambda\lambda}(\theta_0) + J_{\lambda\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)K_{\gamma\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)J'_{\lambda\gamma}(\theta_0) - K_{\lambda\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)J_{\lambda\gamma}(\theta_0) - J_{\lambda\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)K_{\gamma\lambda}(\theta_0). \quad (6.10)$$

Similarly, from (6.8), we obtain

$$\sqrt{n}S_\rho(\tilde{\theta}) \xrightarrow{d} N [J_{\rho\cdot\gamma}(\theta_0)\delta_\rho + J_{\rho\lambda\cdot\gamma}(\theta_0)\delta_\lambda, B_{\rho\cdot\gamma}(\theta_0)], \quad (6.11)$$

where

$$B_{\rho\cdot\gamma}(\theta_0) = K_{\rho\rho}(\theta_0) + J_{\rho\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)K_{\gamma\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)J'_{\rho\gamma}(\theta_0) - K_{\rho\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)J_{\rho\gamma}(\theta_0) - J_{\rho\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)K_{\gamma\rho}(\theta_0). \quad (6.12)$$

Under H_0^λ , the result in (6.11) shows that $J_{\rho\cdot\gamma}^{-1}(\theta_0)\sqrt{n}S_\rho(\tilde{\theta}) \xrightarrow{d} N [\delta_\rho, B_{\rho\cdot\gamma}(\theta_0)]$. Then, using (6.9) and this last result, an adjusted score function that has zero asymptotic mean in the local presence of ρ_0 can be derived as

$$\sqrt{n}S_\lambda^*(\tilde{\theta}) = \sqrt{n} \left(S_\lambda(\tilde{\theta}) - J_{\lambda\rho\cdot\gamma}(\tilde{\theta})J_{\rho\cdot\gamma}^{-1}(\tilde{\theta})S_\rho(\tilde{\theta}) \right). \quad (6.13)$$

Next we show how to determine the asymptotic distribution of $\sqrt{n}S_\lambda^*(\tilde{\theta})$. For this purpose, we consider (6.7) and (6.8) as a combined system

$$\begin{pmatrix} \sqrt{n}S_\lambda(\tilde{\theta}) \\ \sqrt{n}S_\rho(\tilde{\theta}) \end{pmatrix} = \begin{pmatrix} -J_{\lambda\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0) & 1 & 0 \\ -J_{\rho\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0) & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{n}S_\gamma(\theta_0) \\ \sqrt{n}S_\lambda(\theta_0) \\ \sqrt{n}S_\rho(\theta_0) \end{pmatrix} + \begin{pmatrix} J_{\lambda\cdot\gamma}(\theta_0)\delta_\lambda + J_{\lambda\rho\cdot\gamma}(\theta_0)\delta_\rho \\ J_{\rho\cdot\gamma}(\theta_0)\delta_\rho + J_{\rho\lambda\cdot\gamma}(\theta_0)\delta_\lambda \end{pmatrix} + o_p(1). \quad (6.14)$$

The joint asymptotic distribution of $\sqrt{n}S_\lambda(\tilde{\theta})$ and $\sqrt{n}S_\rho(\tilde{\theta})$ can now be determined from (6.14) by using the asymptotic normality of score functions in Assumption 2. Thus, we have

$$\begin{pmatrix} \sqrt{n}S_\lambda(\tilde{\theta}) \\ \sqrt{n}S_\rho(\tilde{\theta}) \end{pmatrix} \xrightarrow{d} N \left[\begin{pmatrix} J_{\lambda\cdot\gamma}(\theta_0)\delta_\lambda + J_{\lambda\rho\cdot\gamma}(\theta_0)\delta_\rho \\ J_{\rho\cdot\gamma}(\theta_0)\delta_\rho + J_{\rho\lambda\cdot\gamma}(\theta_0)\delta_\lambda \end{pmatrix}, \begin{pmatrix} B_{\lambda\cdot\gamma}(\theta_0) & B_{\lambda\rho\cdot\gamma}(\theta_0) \\ B_{\rho\lambda\cdot\gamma}(\theta_0) & B_{\rho\cdot\gamma}(\theta_0) \end{pmatrix} \right], \quad (6.15)$$

where

$$B_{\lambda\cdot\gamma}(\theta_0) = K_{\lambda\lambda}(\theta_0) + J_{\lambda\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)K_{\gamma\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)J'_{\lambda\gamma}(\theta_0) - K_{\lambda\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)J_{\lambda\gamma}(\theta_0) - J_{\lambda\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)K_{\gamma\lambda}(\theta_0). \quad (6.16)$$

$$B_{\lambda\rho\cdot\gamma}(\theta_0) = K_{\lambda\rho}(\theta_0) - J_{\lambda\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)K_{\gamma\rho}(\theta_0) - K_{\lambda\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)J_{\gamma\rho}(\theta_0) + J_{\lambda\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)K_{\gamma\gamma}(\theta_0)J_{\gamma\gamma}^{-1}(\theta_0)J_{\gamma\rho}(\theta_0), \quad (6.17)$$

$B_{\rho\cdot\gamma}(\theta_0)$ and $B_{\rho\lambda\cdot\gamma}(\theta_0)$ are defined similarly. Under our assumptions, we have

$$\sqrt{n}S_\lambda^*(\tilde{\theta}) = (1, -J_{\lambda\rho\cdot\gamma}(\theta_0)J_{\rho\cdot\gamma}^{-1}(\theta_0)) \begin{pmatrix} \sqrt{n}S_\lambda(\tilde{\theta}) \\ \sqrt{n}S_\rho(\tilde{\theta}) \end{pmatrix} + o_p(1). \quad (6.18)$$

Then, using (6.15), under H_0^λ and H_a^ρ , we obtain

$$\sqrt{n}S_\lambda^*(\tilde{\theta}) \xrightarrow{d} N [0, D_{\lambda\cdot\gamma}(\theta_0)], \quad (6.19)$$

where

$$D_{\lambda,\gamma}(\theta_0) = B_{\lambda,\gamma}(\theta_0) + J_{\lambda\rho,\gamma}(\theta_0)J_{\rho,\gamma}^{-1}(\theta_0)B_{\rho,\gamma}(\theta_0)J_{\rho,\gamma}^{-1}(\theta_0)J_{\rho\lambda,\gamma}(\theta_0) \\ - J_{\lambda\rho,\gamma}(\theta_0)J_{\rho,\gamma}^{-1}(\theta_0)B_{\rho\lambda,\gamma}(\theta_0) - B_{\lambda\rho,\gamma}(\theta_0)J_{\rho,\gamma}^{-1}(\theta_0)J_{\rho\lambda,\gamma}(\theta_0). \quad (6.20)$$

Now, consider the asymptotic distribution of $\sqrt{n}S_{\lambda}^*(\tilde{\theta})$ under H_a^{λ} and H_0^{ρ} . Using (6.15) and (6.18), we can derive that

$$\sqrt{n}S_{\lambda}^*(\tilde{\theta}) \xrightarrow{d} N \left[(J_{\lambda,\gamma}(\theta_0) - J_{\lambda\rho,\gamma}(\theta_0)J_{\rho,\gamma}^{-1}(\theta_0)J_{\rho\lambda,\gamma}(\theta_0)) \delta_{\lambda}, D_{\lambda,\gamma}(\theta_0) \right], \quad (6.21)$$

Thus, $LM_{\lambda} \overset{A}{\sim} \chi_1^2(\vartheta_1)$ by Theorem 8.6 of White [24] on the asymptotic distribution of quadratic forms, where $\vartheta_1 = \delta_{\lambda}^2 (J_{\lambda,\gamma}(\theta_0) - J_{\lambda\rho,\gamma}(\theta_0)J_{\rho,\gamma}^{-1}(\theta_0)J_{\rho\lambda,\gamma}(\theta_0))^2 / D_{\lambda,\gamma}(\theta_0)$ is the non-centrality parameter.

References

- [1] Anselin, L. (1998). *Spatial Econometrics: Methods and Models*. Springer, New York.
- [2] Anselin, L., Bera, A.K., Florax, R. and Yoon, M.J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26(1), 77-104.
- [3] Baltagi, B.H., Chang, Y.J. and Li, Q. (1992). Monte Carlo results on several new and existing tests for the error component model. *Journal of Econometrics*, 54(1), 95-120.
- [4] Baltagi, B.H. and Yang, Z. (2013a). Heteroskedasticity and non-normality robust LM tests for spatial dependence. *Regional Science and Urban Economics*, 43(5), 725-739.
- [5] Baltagi, B.H. and Yang, Z. (2013b). Standardized LM tests for spatial error dependence in linear or panel regressions. *The Econometrics Journal*, 16(1), 103-134.
- [6] Bera, A.K. and Yoon, M.J. (1993). Specification testing with locally misspecified alternatives. *Econometric Theory*, 9(4), 649-658.
- [7] Born, B. and Breitung, J. (2011). Simple regression-based tests for spatial dependence. *The Econometrics Journal*, 14(2), 330-342.
- [8] Burridge, P. (1980). On the Cliff-Ord test for spatial correlation. *Journal of the Royal Statistical Society, Series B: Methodological*, 42, 107-108.
- [9] Davidson, R. and MacKinnon, J.G. (1987). Implicit alternatives and the local power of test statistics. *Econometrica*, 55(6), 1305-1329.
- [10] Doğan, O. (2015). Heteroskedasticity of unknown form in spatial autoregressive models with a moving average disturbance term. *Econometrics*, 3(1), 101-127.
- [11] Elhorst, J.P. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer, Berlin.
- [12] Jin, F. and Lee, L. (2018). Outer-product-of-gradients tests for spatial autoregressive models. *Regional Science and Urban Economics*, 72, 35-57.
- [13] Kelejian, H.H. and Prucha, I.R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*, 104(2), 219-257.
- [14] Kelejian, H.H. and Prucha, I.R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157, 53-67.
- [15] Koenker, R. (1981). A note on studentizing a test for heteroscedasticity. *Journal of Econometrics*, 17(1), 107-112.
- [16] Lee, L. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72(6), 1899-1925.
- [17] LeSage, L. and Pace, R.K. (2009). *Introduction to Spatial Econometrics*. Chapman and Hall/CRC, London.

- [18] Lin, X. and Lee, L. (2010). GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics*, 157(1), 34-52.
- [19] Liu, S.F. and Yang, Z. (2014). Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and nonnormality. *Regional Science and Urban Economics*, 52, 50-70.
- [20] Moulton, B.R. and Randolph, W.C. (1989). Alternative tests of the error components model. *Econometrica*, 57(3), 685-693.
- [21] Pace, R.K. and Barry, R. (1997). Quick computation of spatial autoregressive estimators. *Geographical Analysis*, 29(3), 232-247.
- [22] Qu, X. and Lee, L. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, 184(2), 209-232.
- [23] Saikkonen, P. (1989). Asymptotic relative efficiency of the classical test statistics under misspecification. *Journal of Econometrics*, 42(3), 351-369.
- [24] White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge University Press, Cambridge.
- [25] Yang, Z. (2018). Unified M-estimation of fixed-effects spatial dynamic models with short panels. *Journal of Econometrics*, 205(2), 423-447.

VOLUME/CİLT 13 NUMBER/SAYI 3 DECEMBER/ARALIK 2021

İSTATİSTİK

JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION
TÜRK İSTATİSTİK DERNEĞİ DERGİSİ

ISSN: 1300-4077