THREE ESSAYS IN SPATIAL ECONOMETRICS AND LABOR ECONOMICS

by

CANH QUANG LE

B.A., National Economics University, Vietnam, 1997
M.A., National Economics University, Vietnam, 2001

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Economics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2009

Abstract

This dissertation is a combination of three essays on spatial econometrics and labor economics. Essays 1 and 2 developed double length regression (DLR) tests for testing functional form and spatial dependence, which includes spatial error dependence and spatial lag dependence. More specifically, these essays derive the DLR joint, DLR one-direction, and DLR conditional tests for testing functional forms and spatial dependence. The essays also provide empirical examples and Monte Carlo simulations to examine how the DLR tests perform in the empirical work and how the power of the DLR test depends on changes in functional form and spatial dependence. The results suggested that DLR tests work similarly to its Lagrangian Multiplier (LM) counterpart for testing functional form and spatial dependence in the empirical example and simulations. The DLR tests do not require the second-order derivatives of the log-likelihood function, so they provide practitioners an easy-to-use method to test for functional forms and spatial dependence.

Essay 3 investigates the effects of fertility on parental labor force participation and labor supply in Vietnam. The essay uses instrumental variable (IV) probit models to estimate the effects of fertility on parental labor force participation and the IV models to estimate the effects of fertility on parental labor supply. Using the gender of the first child and the same gender of the first two children as two instrumental variables, this essay found negative effects of fertility on maternal labor force participation and labor supply. It also found positive effects of fertility on paternal labor force participation and labor supply. The results suggest that fertility had the specialization effect on parental labor force participation and labor supply in Vietnam. The homogeneity test results indicate that the magnitude of the effects of fertility on parental labor force participation and labor supply is different among parents and locations.

THREE ESSAYS IN SPATIAL ECONOMETRICS AND LABOR ECONOMICS

by

CANH QUANG LE

B.A., National Economics University, Vietnam, 1997
M.A., National Economics University, Vietnam, 2001

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Economics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2009

Approved by:

Major Professor
Dr. Dong Li

Abstract

This dissertation is a combination of three essays on spatial econometrics and labor economics. Essays 1 and 2 developed double length regression (DLR) tests for testing functional form and spatial dependence, which includes spatial error dependence and spatial lag dependence. More specifically, these essays derive the DLR joint, DLR one-direction, and DLR conditional tests for testing functional forms and spatial dependence. The essays also provide empirical examples and Monte Carlo simulations to examine how the DLR tests perform in the empirical work and how the power of the DLR test depends on changes in functional form and spatial dependence. The results suggested that DLR tests work similarly to its Lagrangian Multiplier (LM) counterpart for testing functional form and spatial dependence in the empirical example and simulations. The DLR tests do not require the second-order derivatives of the log-likelihood function, so they provide practitioners an easy-to-use method to test for functional forms and spatial dependence.

Essay 3 investigates the effects of fertility on parental labor force participation and labor supply in Vietnam. The essay uses instrumental variable (IV) probit models to estimate the effects of fertility on parental labor force participation and the IV models to estimate the effects of fertility on parental labor supply. Using the gender of the first child and the same gender of the first two children as two instrumental variables, this essay found negative effects of fertility on maternal labor force participation and labor supply. It also found positive effects of fertility on paternal labor force participation and labor supply. The results suggest that fertility had the specialization effect on parental labor force participation and labor supply in Vietnam. The homogeneity test results indicate that the magnitude of the effects of fertility on parental labor force participation and labor supply is different among parents and locations.

Table of Contents

List of Figures

List of Tables

List of Appendix

Acknowledgements

I am grateful to my major professor Dr. Dong Li for his great support, guidance, and encouragement in every aspect of writing this dissertation, my study, and daily life. He read and edited every version of this dissertation from its infancy. I am also grateful to him for the original idea and his substantial inputs needed to finish this dissertation.

I would also like to thank the remaining members of my committee, Drs. Dennis L. Weisman, Lance Bachmeier, and Tian Xia and outside Chairperson Dr. James Neill for their constructive and valuable comments and suggestions regarding to this work. Their comment and suggestion make the dissertation more valuable.

My special thanks are sent to Drs. Lloyd Thomas, William Blankenau for their support and encouragement over the years I have been at K-State. I have benefited not only from their classes in terms of academic and teaching experience but also from their management skills.

I am also thankful for the opportunity to pursue a Ph.D. in Economics at Kansas State University. My experience with the department and university has significantly enhanced my research ability. I owe many thanks to the faculty and students for making my time as a graduate student such a rewarding experience. Additionally, I extend my appreciation to Melanie Graber, Steve Graber, and Roger Fridemann for their editorial efforts.

I thank my wife Thoan Nguyen, my sons Tung Duong and Viet Anh for their enthusiasm, encouragement, patience, and support over the years at K-State. I am also grateful to my parents Tuc and Vui, Bat and Thoa for their understanding and support during my hardship. I love you all.

Dedication

This work is dedicated to my parents Le Xuan Tuc and Nguyen Thi Vui who scarified their own interests to give their sons and daughters brighter lives.

Essay 1: Double Length Regression Tests for Testing Functional Form and Spatial Error Dependence

1. Introduction

Spatial dependence and its regression models have attracted much attention from econometricians and economists. Spatial dependence is a commonly used concept in regional science, urban economics, transport economics, and economic geography. It occurs among regions or spatial units via trade, investments, capital flows, and immigration (Anselin, 1988a). Many studies have proposed theoretical econometric models dealing with spatial dependence. For example, Anselin (1981, 1988b) specified a linear model with a spatial autoregressive structure in the disturbance. Those papers developed estimation methods for spatial autoregressive structures and studied the properties of estimators for the linear model with spatial autoregressive structures. Cliff and Ord (1981) reviewed some spatial processes and introduced several models and applications. Anselin (1988a) performed a comprehensive investigation of methods and models in spatial econometrics. In this work, he laid the foundations for the econometric analysis of spatial processes, developed estimation methods and hypotheses testing for spatial processes, and discussed model validation in spatial econometric models. LeSage (1998) also contributed to the literature on spatial dependence, spatial models, estimations, and applications, while Upton and Fingleton (1985) added to the spatial econometrics literature concerning spatial data analysis, methods, and models. McMillen (1992) developed two methods - the expectation-maximization (EM) algorithm and the Maximum Likelihood estimation - for probit models with spatial autocorrelation. Baltagi and Li (2004b) proposed a model that allowed for prediction in the panel models with spatial correlation. Rey and Anselin (2006) provided methods, applications, and software for spatial analysis in social sciences; Anselin et al. (2007) added to the literature about spatial panel econometrics with research involving spatial effects, spatial panel model specifications, estimations, and spatial dependence testing. The theoretical spatial models have enriched the spatial econometrics literature in general and shed light on the empirical study of spatial dependence.

Many empirical studies have focused on estimating spatial dependence and its effects. For example, Anselin (1995) investigated the spatial patterns in Appalachian economic growth and development; Anselin and Can (1998) studied the spatial effects in the mortgage market; Can and Megbolugbe (1997) considered spatial correlation in the construction of a housing price index; Boxall et al. (2005) used a spatial hedonic analysis for property value assessment; and Lundberg (2006) used a spatial interaction model to estimate spillovers of public services among municipalities in Sweden. Some other studies have focused on the relationship between spatial dependence and economic growth. For example, Lundberg (2004) used spatial econometrics to investigate municipal economic growth in Sweden; Koch (2006) considered the relationship between economic growth and spatial dependence in European countries; Ertur et al. (2006) attempted to answer whether spatial dependence affected beta convergence and how spatial dependence affected the beta-convergence process in European countries. In other studies, Mobley et al. (2006) used the spatial approach for analyzing the probability of the elderly accessing primary care services; Longhi and Nijkamp (2007), taking into account of spatial error and spatial lag dependence, proposed models to estimate and forecast employment in regional labor markets in West Germany; Lacombe and Shaughnessy (2007) used spatial error models to examine whether spatial error dependence affected the 2004 presidential county vote outcome.

Testing for spatial dependence is an important issue not only in spatial econometrics, but also in regional science, urban economics, transport economics, and economic geography. In traditional econometrics, spatial dependence is often treated exogenously even though this assumption can violate the traditional Gauss-Markov assumptions when regression models are run and inferences made (LeSage, 1998). Many studies have paid attention to the issue of how to test for spatial dependence. For example, Cliff and Ord (1972) developed a model that tests for spatial autocorrelation among regression residuals; Anselin (1985) proposed specification tests for choosing appropriate models for spatial interactions and the structure of spatial dependence. Anselin and Rey (1991) investigated the performance of tests for spatial dependence in linear regression models; Anselin et al. (1996) proposed a simple diagnostic test for spatial dependence. Baltagi et al. (2002) derived Lagrange Multiplier (LM) tests for panel data regression models with spatial error correlation. In general, however, the specification tests for spatial dependence can be categorized into two categories: The first is the Morian's I test, proposed by Moran (1948, 1950) and further extended by Cliffs and Ord (1972). The second includes the maximum likelihood-based tests, including the LM and the Rao Score (RS) tests. The extensions of the LM and RS tests can be found in Anselin (1988b) for panel data, in Anselin (2001) for spatial error components and a direct representation model, and in Pinkse (1998) and Pinkse and Slade (1998) for the probit models.

Testing for functional form and spatial dependence can be more complicated than testing for spatial dependence because of unknown functional forms and/or suspicious non-linearity. In reality, non-linearity is present in many fields and data sets that relate to population, immigration, and housing prices. Many studies have suggested that non-linearity has to be involved when one models issues that relate to population, immigration, and housing prices. For example, Ledent (1986) attempted to find an appropriate model that described migration exchange between rural and urban areas. He suggested that the model should be specified for the non-linearity of migration flows. Some other papers also found a non-linear structure in housing price models (Cassel and Mendelsohn, 1985 and Craig et al., 1991) for housing prices in Houston, and Shimizu et al. (2007) for housing prices in Tokyo. Elad et al. (1994) found non-linearity when they specified and estimated a model of farmland prices in Georgia. Non-linearity was also found in many other studies (Copper et al., 1988; Mok et al., 1995; Huh and Kwak, 1997; Griffith et al., 1998; Fik and Mulligan, 1998; Pace et al., 1999; Graaf et al., 2001; Crushing, 2005; and Fattouh et al., 2005). The existence of non-linearity and spatial dependence, thus, makes testing for functional form and spatial dependence important. It may require using the Box-Cox model (Graff et al., 2001; Miyazaki, 2005), the maximum likelihood methods to account for the spatial dependence (Upton and Fingleton, 1985), or the linearized Box-Cox transformation (Griffith et al., 1998).

In the literature, however, few studies have simultaneously tested for functional form and spatial dependence. Fik and Mulligan (1998) dealt with functional form and functional misspecification in regression-based spatial interaction models for United States labor migration. In their research, they used three models: production-constrained gravity, competing and intervening destinations (CID), and an extension of the CID models for testing functional form and spatial interactions. Baltagi and Li (2001, 2004a) derived the LM tests based on the maximum likelihood approach for functional form and spatial dependence. They proposed LM joint, conditional tests and local misspecification RS tests for testing functional form and spatial error dependence and testing functional form and spatial lag dependence. Yang et al. (2006), using the maximum likelihood-based method, proposed a generalized dynamic error component model that simultaneously accounted for functional form and spatial dependence.

This essay derives double length regression (DLR) tests for functional form and spatial error dependence. It uses the transformation procedure developed by Box and Cox (1964) and the double length regression proposed by Davidson and Mackinnon (1984, 1985). In particular, the essay derives the joint, one-direction, and conditional tests for functional form and spatial error dependence. The DLR joint test is used for testing linear and log-linear models with no spatial error dependence against a general Box-Cox model with spatial error dependence. The DLR one-direction test is used either for assuming log-linear and linear functional form to test for spatial error dependence or assuming no spatial error dependence to test for functional form. The DLR conditional test is developed for testing spatial error dependence conditional on a general Box-Cox model or linearity/log-linearity conditional on a general spatial error dependence. These tests are illustrated and simulated by using the crime data set in Anselin (1988a).

The results of using these tests have shown that the DLR tests perform similarly to the LM counterpart reported in Baltagi and Li (2001). Furthermore, using the DLR test does not require the second-order derivatives of the maximum likelihood function. The DLR tests are computationally cheaper and provide good alternatives to their LM counterparts. Based on our limited knowledge, this is the first work using DLR test statistics for testing functional form and spatial error dependence jointly. In this essay, the DLR test for functional form and spatial lag dependence are not derived even though the spatial lag dependence is an important alternative to spatial models (Anselin, 2001). The DLR tests for testing functional form and spatial lag dependence are derived in the next essay.

The remainder of this essay is organized as follows. Section 2 shows how to set up the DLR models to test hypotheses. It also derives the DLR joint, one-direction, and conditional tests for functional form and spatial error dependence. Section 3 briefly describes the sample data, formation of spatial weight matrix, and the results of the DLR tests based on the crime data set. Section 4 presents the Monte Carlo simulations. Section 5 concludes the essay.

2. Spatial error dependence and DLR tests

The conventional Box-Cox model is expressed as follows:

$$y^{(r)} = X^{(r)}\beta + Z\gamma + u \tag{1.1}$$

where both $y^{(r)}$ and $X^{(r)}$ are required to take positive values and follow the Box-Cox transformation:

$$X^{(r)} = \begin{cases} \dfrac{X^{r}-1}{r} & \text{with } r \neq 0 \\ \log(X) & \text{with } r = 0, \end{cases}$$

Variable Z is not subject to the Box-Cox transformation, and it could include dummy variables, a time trend, and the intercept. When $r = 1$, the model in equation (1.1) is linear, and if $r = 0$, it becomes a log-linear model. u is an $T \times 1$ vector of regression residuals, which are assumed to follow a spatial dependence process:

$$u = \lambda W u + \varepsilon \tag{1.2}$$

where λ is a spatial error dependence coefficient ($-1 < \lambda < 1$). W is an $T \times T$ matrix of known spatial weights in which T is the number of observations, and ε is an $T \times 1$ residual vector that follows the normal distribution N $(0, \delta_{\varepsilon}^{2}$ I).

Substituting (1.2) into (1.1) and rearranging, we have:

$$(I - \lambda W) y^{(r)} = (I - \lambda W) X^{(r)}\beta + (I - \lambda W) Z\gamma + \varepsilon \tag{1.3}$$

where $y^{(r)}$ is $T \times 1$; $X^{(r)}$ is $T \times k$; Z is $T \times s$; β is $k \times 1$; and γ is $s \times 1$. According to Davidson and MacKinnon (1984, 1985), applying the DLR requires error terms, ε, in (1.3) following the normal standard distribution, $N(0, 1)$. To obtain the normal standard errors, we divide both sides of equation (1.3) by the standard deviation.

$$\frac{1}{\sigma}(I - \lambda W)y^{(r)} = \frac{1}{\sigma}\left[(I - \lambda W)X^{(r)}\beta + (I - \lambda W)Z\gamma\right] + \varepsilon^{*} \tag{1.4}$$

where $\varepsilon^{*} = \dfrac{\varepsilon}{\sigma} \sim N(0, I)$

One question is, how can one obtain consistent estimates of parameters β and γ from the spatial error dependence model? Because $y^{(r)}$ and $X^{(r)}$ follow the Box-Cox transformation, and spatial autoregressive parameter λ appears in the regression model, the Ordinary Least Square (OLS) method does not work. In this case, the maximum likelihood method is often used. The log-likelihood function for a general Box-Cox with spatial error dependence model is given by:

$$Log(L) = -\frac{T}{2}\log(2\pi) - T\log(\sigma) + \log\left|I - \lambda W\right| + (r-1)\sum_{t=1}^{T}\log(y_t) -$$

$$\frac{1}{2\sigma^2}\left[(I - \lambda W)y^{(r)} - (I - \lambda W)X^{(r)}\beta - (I - \lambda W)Z\gamma\right]\left[(I - \lambda W)y^{(r)} - (I - \lambda W)X^{(r)}\beta - (I - \lambda W)Z\gamma\right]$$

Note that the log-likelihood function is a combination of the Jacobian term and total sum of squared residuals. If we define the residuals in the equation (1.3) and the Jacobian term as $f(y, \theta)$ and $k(y, \theta)$ respectively, we have:

$$f(y, \theta) = \frac{1}{\sigma}\left[(I - \lambda W)y^{(r)} - (I - \lambda W)X^{(r)}\beta - (I - \lambda W)Z\gamma\right], \text{ and} \tag{1.5}$$

$$k(y_t,\theta) = -T\log(\sigma) + \sum_{t=1}^{T}\log(1-\lambda\varpi_t) + (r-1)\sum_{t=1}^{T}\log(y_t)$$

where $\theta = (\sigma, \beta', \gamma', \lambda, r)'$

Ord (1975) and Anselin (1988b) showed that $\log\left|I-\lambda W\right| = \sum_{t=1}^{T}\log(1-\lambda\varpi_t)$, where ϖ_t's are eigenvalues of the spatial weight matrix W. Thus, a typical element of $k(y,\theta)$ is:

$$k_t(y_t,\theta) = -\log(\sigma) + \log(1-\lambda\varpi_t) + (r-1)\log(y_t) \tag{1.6}$$

If we define $F_t(y_t,\theta) = \dfrac{\partial f_t(y_t,\theta)}{\partial\theta}$ and $K_t(y_t,\theta) = \dfrac{\partial k_t(y_t,\theta)}{\partial\theta}$, the DLR model can be written as an artificial regression with double the number of observations:

$$\begin{bmatrix} f(y,\theta) \\ \iota_T \end{bmatrix} = \begin{bmatrix} -F(y,\theta) \\ K(y,\theta) \end{bmatrix} b + residuals \tag{1.7}$$

where b is an $k \times 1$ vector; ι is an $T \times 1$ vector of unity; $F_t(y_t,\theta)$ and $K(y,\theta)$ are an $T \times (k+s+3)$ matrix whose typical elements are $\dfrac{\partial f_t(y_t,\theta)}{\partial\theta}$ and $\dfrac{\partial k_t(y_t,\theta)}{\partial\theta}$ respectively for $t = 1, 2, \ldots. T$.

The model in (1.7) can be expressed in a more detailed formula as:

$$\begin{bmatrix} \dfrac{\hat{\varepsilon}_i}{\sigma} \\ \iota_T \end{bmatrix} = \begin{bmatrix} -\dfrac{\partial f(\cdot)}{\partial\sigma} & -\dfrac{\partial f(\cdot)}{\partial\beta} & -\dfrac{\partial f(\cdot)}{\partial\gamma} & -\dfrac{\partial f(\cdot)}{\partial\lambda} & -\dfrac{\partial f(\cdot)}{\partial r} \\ \dfrac{\partial k(\cdot)}{\partial\sigma} & \dfrac{\partial k(\cdot)}{\partial\beta} & \dfrac{\partial k(\cdot)}{\partial\gamma} & \dfrac{\partial k(\cdot)}{\partial\lambda} & \dfrac{\partial k(\cdot)}{\partial r} \end{bmatrix} b \;+\; residuals \tag{1.8}$$

In this expression, the dependent variable is the normalized residuals for observations from 1 to T and unity for observations from $T+1$ to $2T$. The regressors are the negative first-order derivatives of normalized standard residuals with respect to θ for observations from

1 to T and first-order derivatives of the Jacobian term with respect to θ for observations from $T+1$ to $2T$. The typical elements F_t and K_t are derived as:

For σ: $\frac{\partial f_t(y_t,\theta)}{\partial\sigma} = -\frac{1}{\sigma^2}(I-\lambda W)(y^{(r)} - X^{(r)}\beta - Z\gamma)$ $\qquad \frac{\partial k_t(y_t,\theta)}{\partial\sigma} = -\frac{1}{\sigma}$

For β: $\frac{\partial f(y,\theta)}{\partial\beta} = -\frac{1}{\sigma}(I-\lambda W)X^{(r)}$ $\qquad \frac{\partial k_t(y_t,\theta)}{\partial\beta} = 0$

For γ: $\frac{\partial f(y,\theta)}{\partial\gamma} = -\frac{1}{\sigma}(I-\lambda W)Z$ $\qquad \frac{\partial k_t(y_t,\theta)}{\partial\gamma} = 0$

For λ: $\frac{\partial f_t(y_t,\theta)}{\partial\lambda} = -\frac{1}{\sigma}W(y^{(r)} - X^{(r)}\beta - Z\gamma)$ $\qquad \frac{\partial k_t(y_t,\theta)}{\partial\lambda} = -\frac{1}{(1-\lambda\varpi_t)}\varpi_t$

For r: $\frac{\partial f_t(y_t,\theta)}{\partial r} = \frac{1}{\sigma}(I-\lambda W)[C(y,r) - C(x,r)\beta]$ $\qquad \frac{\partial k_t(y_t,\theta)}{\partial r} = \log y_t$

where $C(y,\, r) = \frac{1}{r^2}[y^r(r\log y - 1) + 1]$. All these elements of regressors will be evaluated under specific null hypotheses.

Davidson and MacKinnon (1993) showed that the total sum of squares of the left-hand side variable in (1.8) is $2T$, where T is the number of observations. The explained sum of squares, $2T - SSR$ [1], provides an asymptotically valid test statistic. They also showed that this test statistic asymptotically follows the Chi-square distribution with degrees of freedom equal to the number of restrictions under the null hypothesis. Taking advantage of this property, we can use $2T - SSR$ as a DLR test statistic to perform hypotheses testing.

[1] The explained sum of squares (SSE) equals the total sum of squares (SST) minus the residual sum of squares (SSR).

2.1. DLR joint test

In this subsection, the derivation of two DLR joint tests is presented. First, we derive a DLR joint test for the hypothesis $H_0^1 : r = 0 \; and \; \lambda = 0$. Under this hypothesis, the model is log-linear with no spatial error dependence. Since $y^{(0)} = \log(y)$ and $C(y,\theta) = \lim_{r \to 0} C(y,\theta) = \frac{1}{2}(\log y)^2$, typical elements of regressors are:

$$\frac{\partial f_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma^2}\left[\log y_t - (\log x_t)\beta - z_t\gamma\right] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \beta} = -\frac{1}{\sigma}\log x_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \beta} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\frac{1}{\sigma} z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \lambda} = -\frac{1}{\sigma}\varpi_t\left[\log y_t - (\log x_t)\beta - z_t\gamma\right] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \lambda} = -\varpi_t$$

$$\frac{\partial f_t(y_t,\theta)}{\partial r} = \frac{1}{2\sigma}[(\log y_t)^2 - (\log x_t)^2\beta] \qquad \frac{\partial k_t(y_t,\theta)}{\partial r} = \log y_t$$

where ϖ_t is eigenvalues of the spatial weight matrix.

Second, we derive a DLR joint test for the hypothesis $H_0^2 : r = 1 \; and \; \lambda = 0$. Under this hypothesis, the model becomes linear with no spatial error dependence. Since $y^{(1)} = y - 1$ and $C(y,\theta) = \lim_{r \to 1} C(y,\theta) = y\log y - y + 1$, elements of regressors are given:

$$\frac{\partial f_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma^2}[(y_t - 1) - (x_t - 1)\beta - z_t\gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \beta} = -\frac{1}{\sigma}(x_t - 1) \qquad \frac{\partial k_t(y_t,\theta)}{\partial \beta} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\frac{1}{\sigma} z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \lambda} = -\frac{1}{\sigma}\varpi_t[(y_t - 1) - (x_t - 1)\beta - z_t\gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \lambda} = -\varpi_t$$

$$\frac{\partial f_t(y_t,\theta)}{\partial r} = \frac{1}{\sigma}[(y_t(\log y_t - 1) + 1) - (x_t(\log x_t - 1) + 1)\beta] \qquad \frac{\partial k_t(y_t,\theta)}{\partial r} = \log y_t$$

2.2. DLR one-direction test

In this subsection, the derivation of the DLR one-direction test is presented for two cases. One case assumes the functional form is known, and the DLR one direction test is used to test for spatial error dependence; and the other case assumes no spatial error dependence, the DLR one-direction test is used to test for functional form. Like the DLR joint tests, the DLR one-direction tests are performed by using the OLS regression technique, and the DLR one-direction test statistics are asymptotically distributed as Chi-square with degrees of freedom equal to the number of restrictions under the null hypothesis. There are four hypotheses for the DLR one-direction tests.

Under the hypothesis $H_0^3 : \lambda = 0$ assuming $r = 0$, our model assumed log-linearity with no spatial error dependence. In other words, functional form was known to be log-linear, while the spatial error dependence needed to be tested. Recall that in the model in (1.4)

$$\frac{1}{\sigma}(I - \lambda W)\log(y) = \frac{1}{\sigma}\left[(I - \lambda W)\log(X)\beta + (I - \lambda W)Z\gamma\right] + \varepsilon^*$$

where $\varepsilon^* \sim N(0, I)$. Under this hypothesis, the parameter vector is $\theta = (\sigma, \beta', \gamma', \lambda)'$.

Typical elements of regressors of this model are:

$$\frac{\partial f_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma^2}[\log y_t - \log x_t\beta - z_t\gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \beta} = -\frac{1}{\sigma}\log x_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \beta} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\frac{1}{\sigma}z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \lambda} = -\frac{1}{\sigma}\varpi_t[\log y_t - \log x_t\beta - z_t\gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \lambda} = -\varpi_t$$

Under the hypothesis $H_0^4 : \lambda = 0$ assuming $r = 1$, the model assumed linearity with no spatial error dependence. The hypothesis meant that functional form was known to be linear, while the spatial error dependence needed to be tested.

Typical elements of regressors of the DLR model under this hypothesis are:

$$\frac{\partial f_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma^2}[(y_t - 1) - (x_t - 1)\beta - z_t\gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \beta} = -\frac{1}{\sigma}(x_t - 1) \qquad \frac{\partial k_t(y_t,\theta)}{\partial \beta} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\frac{1}{\sigma}z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \lambda} = -\frac{1}{\sigma}\varpi_t[(y_t - 1) - (x_t - 1)\beta - z_t\gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \lambda} = -\varpi_t$$

Under the hypothesis $H_0^5 : r = 0$ assuming $\lambda = 0$, the model became log-linear assuming no spatial error dependence. The model under the hypothesis had a form as follows:

$$\frac{1}{\sigma}\log(y)=\frac{1}{\sigma}\left[\log(X)\beta+Z\gamma\right]+\varepsilon^*, \text{ where } \varepsilon^*\sim N(0,I).$$

In this case, we need to test whether the model is log-linear when no spatial error dependence is assumed. The parameters of the DLR model are $\theta=(\sigma,\beta',\gamma',r)'$, and the typical elements of regressors are as follows:

$$\frac{\partial f_t(y_t,\theta)}{\partial\sigma}=-\frac{1}{\sigma^2}[\log y_t-\log x_t\beta-z_t\gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial\sigma}=-\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial\beta}=-\frac{1}{\sigma}\log x_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial\beta}=0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial\gamma}=-\frac{1}{\sigma}z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial\gamma}=0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial r}=-\frac{1}{2\sigma}[(\log y_t)^2-(\log x_t)^2\beta] \qquad \frac{\partial k_t(y_t,\theta)}{\partial r}=\log(y_t)$$

Under the hypothesis $H_0^6: r=1$ assuming $\lambda=0$, the model became linear assuming no spatial error dependence. The model under the hypothesis is:

$$\frac{1}{\sigma}(y-1)=\frac{1}{\sigma}\left[(X-1)\beta+Z\gamma\right]+\varepsilon^*$$

where $\varepsilon^*\sim N(0,\ I)$. Parameters of the DLR model are $\theta=(\sigma,\beta',\gamma',r)'$, and typical elements of regressors are derived as:

$$\frac{\partial f_t(y_t,\theta)}{\partial\sigma}=-\frac{1}{\sigma^2}[(y_t-1)-(x_t-1)\beta-z_t\gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial\sigma}=-\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial\beta}=-\frac{1}{\sigma}(x_t-1) \qquad \frac{\partial k_t(y_t,\theta)}{\partial\beta}=0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\frac{1}{\sigma} z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial r} = \frac{1}{\sigma}[y_t(\log y_t - 1) + 1 - (x_t(\log x_t - 1) - 1)\beta] \qquad \frac{\partial k_t(y_t,\theta)}{\partial r} = \log(y_t)$$

2.3. DLR conditional test

In the real world, researchers and econometricians frequently want to test more than one restriction, so the joint test should be used. Consequently, rejection of the joint test does not point to the right model. To deal with this limitation of joint tests, this section presents the derivation of the DLR conditional tests. The DLR conditional tests take into account either the possibility of spatial error dependence when testing for functional form or the possibility of functional misspecification when testing for spatial error dependence.

DLR test for spatial error dependence conditional on a general Box-Cox model

This conditional test considered the hypothesis $H_0^7 : \lambda = 0 | unkown\ r$. Under this hypothesis, the model becomes a general Box-Cox model with no spatial error dependence. Its elements of regressors are derived as follows:

$$\frac{\partial f_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma^2}(y_t^{(r)} - x_t^{(r)}\beta - z_t\gamma) \qquad \frac{\partial k_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \beta} = -\frac{1}{\sigma} x_t^{(r)} \qquad \frac{\partial k_t(y_t,\theta)}{\partial \beta} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\frac{1}{\sigma} z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \lambda} = -\frac{1}{\sigma}\varpi_t(y_t^{(r)} - x_t^{(r)}\beta - z_t\gamma)$$

$$\frac{\partial k_t(y_t,\theta)}{\partial \lambda} = -\varpi_t \qquad \frac{\partial f_t(y_t,\theta)}{\partial r} = \frac{1}{\sigma r^2}\left[[y^r(r\log y - 1) + 1] - [x^r(r\log x - 1) + 1]\beta]\right]$$

$$\frac{\partial k_t(y_t,\theta)}{\partial r} = \log y_t$$

DLR test for log-linear model conditional on spatial error dependence

Under the hypothesis $H_0^8 : r = 0 | unkown\ \lambda$, the model became log-linear conditional on spatial error dependence. It has typical elements of regressors as follows:

$$\frac{\partial f_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma^2}(i_t - \lambda\varpi_t)\left[\log y_t - (\log x_t)\beta - z_t\gamma\right] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \beta} = -\frac{1}{\sigma}(i_t - \lambda\varpi_t)\log x_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \beta} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\frac{1}{\sigma}(i_t - \lambda\varpi_t)z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \lambda} = -\frac{1}{\sigma}\varpi_t\left[\log y_t - (\log x_t)\beta - z_t\gamma\right] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \lambda} = -\frac{1}{1-\lambda\varpi_t}\varpi_t$$

$$\frac{\partial f_t(y_t,\theta)}{\partial r} = \frac{1}{2\sigma}(i_t - \lambda\varpi_t)[(\log y_t)^2 - (\log x_t)^2\beta] \qquad \frac{\partial k_t(y_t,\theta)}{\partial r} = \log y_t$$

where i_t is the row t^{th} of the unity matrix I.

DLR test for linear model conditional on spatial error dependence

The DLR test for a linear model conditional on spatial error dependence was performed under the hypothesis $H_0^9 : r = 1 | unkown\ \lambda$. The model was linear conditional on spatial error dependence, and its typical elements of regressors are:

$$\frac{\partial f_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma^2}(i_t - \lambda \varpi_t)[(y_t - 1) - (x_t - 1)\beta - z_t\gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \beta} = -\frac{1}{\sigma}(i_t - \lambda \varpi_t)(x_t - 1) \qquad \frac{\partial k_t(y_t,\theta)}{\partial \beta} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\frac{1}{\sigma}(i_t - \lambda \varpi_t)z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \lambda} = -\frac{1}{\sigma}\varpi_t[(y_t - 1) - (x_t - 1)\beta - z_t\gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \lambda} = -\frac{\varpi_t}{1 - \lambda \varpi_t}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial r} = \frac{1}{\sigma}(i_t - \lambda \varpi_t)[(y_t(\log y_t - 1) + 1) - (x_t(\log x_t - 1) + 1)\beta] \qquad \frac{\partial k_t(y_t,\theta)}{\partial r} = \log y_t$$

3. Illustrative crime data example

This section provides information about the spatial weight matrix and data used to illustrate how DLR tests work in empirical examples. It also presents the results of the DLR tests based on the crime data set in the neighborhood of Columbus, Ohio.

3.1. Spatial weight matrix

Calculating the spatial weight matrix is one of the key components of spatial dependence models. It provides a theoretical and pre-understanding about the nature of spatial interactions among geographical regions or among economic agents. If the spatial weight matrix is appropriately calculated, the spatial models are closer to the reality. If not, it will lead to biases in regression results and eventually misleading inferences.

The spatial weight matrix can be calculated in several ways. Cliff and Ord (1981) calculated the spatial weight matrix by combining the distance and length of common borders among spatial units. Dacey (1968) claimed that the weight matrix should also take into account the physical features of spatial agents including distances, common borders, and areas. Bodson and Peeter (1975) computed the spatial weight matrix by combining impacts of channels of communication among spatial units. Some other studies used a binary spatial weight matrix with a distance-based critical cut-off and the travel time (Gatrell, 1985 and Koch, 2006).

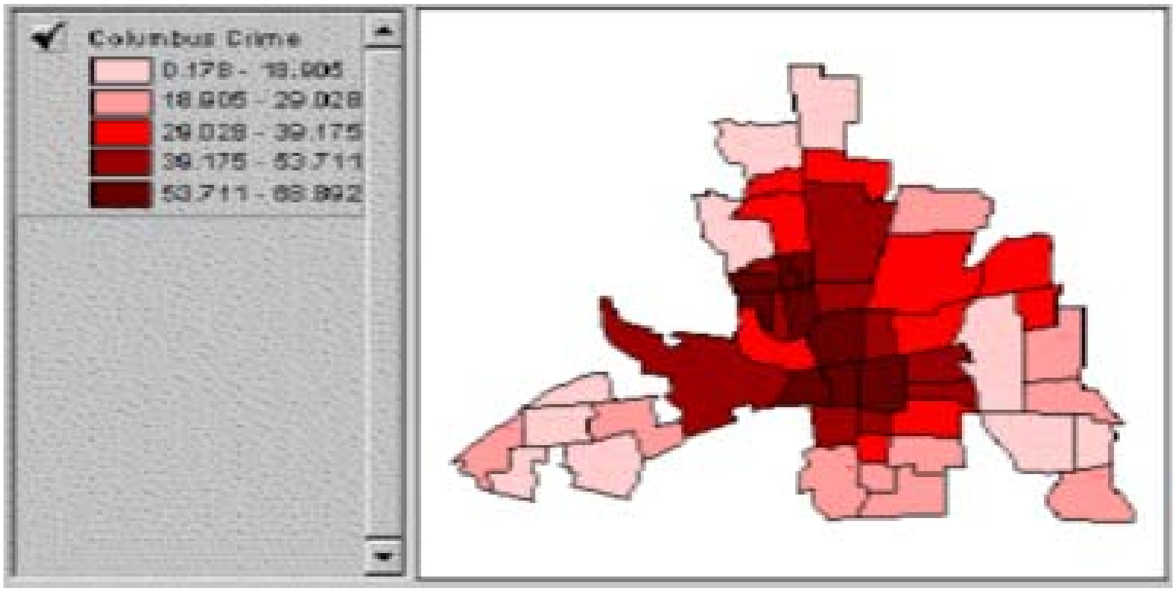
This essay follows Anselin (1988a) approach to computing the spatial weight matrix based on binary contiguity. If zone j is adjacent to zone i, the regional interaction receives a weight of one; otherwise, the interaction receives a weight of zero. The internal interaction of a region w_{ii}^{*} receives a weight of zero. The row standardized spatial weight matrix can be generated as:

$$w_{ij} = \frac{w_{ij}^{*}}{\sum_{j} w_{ij}^{*}},$$

where $\sum_{j} w_{ij} = 1$; $i = 1, 2, 3...T$; $j = 1, 2, 3...T$; w_{ij}^{*} is a typical element of the raw spatial weight matrix while w_{ij} is an element of row standardized spatial weight matrix.

It is important to note that typical spatial interaction terms, w_{ij}, are exogenous to the model, so they can avoid the identification problem (Manski, 1993). In this dissertation, using the crime data in Anselin (1988a), we illustrate how one can apply the DLR tests. The spatial layout and the crime distribution map are shown in Figure 1.1 on the next page.

Figure 1. 1: Crime distribution map of the 49 Neighborhoods in Columbus, Ohio



Source: Generated from GeoDa™ developed by Anselin

The spatial weight matrix is produced by using Anselin's (1988a) binary contiguity approach. As a part of a 49×49 spatial weight matrix, Table 1 gives a 10×10 matrix for the purpose of illustration.

Table 1. 1: A part of the spatial weight matrix

Zone ij	1	2	3	4	5	6	7	8	9	10
1	0.00	0.33	0.00	0.00	0.33	0.33	0.00	0.00	0.00	0.00
2	0.25	0.00	0.25	0.00	0.00	0.25	0.25	0.00	0.00	0.00
3	0.00	0.17	0.00	0.17	0.00	0.00	0.17	0.00	0.00	0.00
4	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.50	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00
6	0.25	0.25	0.00	0.00	0.25	0.00	0.25	0.00	0.00	0.00
7	0.00	0.13	0.13	0.00	0.00	0.13	0.00	0.13	0.13	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

3.2. Crime data

Data used in this example were taken from Table 12.1, Anselin (1988a, p.189). These data were used to investigate how the number of crimes is related to household incomes and/or housing values in the 49 neighborhoods in Columbus, Ohio, in 1980. The crime variable comprises the number of residential burglaries and vehicle thefts per thousand households in the neighborhood. Housing values and household income are measured in thousands of dollars. Spatial effects are visually represented in Figure 1.1 on the preceding page.

The OLS results are:

$$\text{CRIME} = \underset{(14.490)}{68.619} - \underset{(4.781)}{1.597}\ \text{H.INCOME} - \underset{(2.654)}{0.274}\ \text{H.VALUE}$$

$$N = 49 \qquad R^2 = 0.5524$$

where the numbers in the parentheses are t-statistics. The results show that the estimated coefficients are significant at 5% level.

3.3. Results of the DLR tests

In this subsection, we present the results of the DLR tests for functional form and spatial error dependence for the crime data set. The dependent variable is the number of crimes, and independent variables are the housing values and household incomes. All three variables are subject to the Box-Cox transformation, while the intercept is not. The DLR test statistics under the hypotheses are shown in Table 1.2 on the next page.

Table 1. 2: DLR test statistics for functional form and spatial error dependence

	DLR Test	LM Test*
DLR joint tests		
$H_0^1 : r = 0 \; and \; \lambda = 0$	57.050	54.058
	(0.000)	(0.000)
$H_0^2 : r = 1 \; and \; \lambda = 0$	7.403	13.528
	(0.007)	(0.001)
DLR One-Direction tests		
$H_0^3 : \lambda = 0$ assuming $r = 0$	1.959	2.063
	(0.162)	(0.151)
$H_0^4 : \lambda = 0$ assuming $r = 1$	5.697	11.442
	(0.017)	(0.001)
$H_0^5 : r = 0$ assuming $\lambda = 0$	56.235	53.754
	(0.000)	(0.000)
$H_0^6 : r = 1$ assuming $\lambda = 0$	1.465	0.024
	(0.226)	(0.878)
DLR conditional tests		
$H_0^7 : \lambda = 0 \| unkown \; r$	3.742	7.600
	(0.053)	(0.006)
$H_0^8 : r = 0 \| unkown \; \lambda$	54.392	75.534
	(0.000)	(0.000)
$H_0^9 : r = 1 \| unkown \; \lambda$	2.250	0.272
	(0.134)	(0.602)

Numbers in the parentheses are p-values.

* These numbers are reproduced from Table 1 in Baltagi and Li (2001) with permission from Sage Publications.

According to Table 1.2, joint test statistics for the null hypotheses $H_0^1 : r = 0 \; and \; \lambda = 0$ and $H_0^2 : r = 1 \; and \; \lambda = 0$ are 57.050 and 7.403 respectively. They are statistically significant at the 1% level. These results mean that a log-linear and a linear model with no spatial error correlation are strongly rejected in favor of a general Box-Cox model with spatial error dependence.

For the DLR one-direction tests, the null hypothesis H_0^3 was not rejected at the 5% level, while hypothesis H_0^4 was strongly rejected. In other words, when assuming a log-linear model, one cannot reject the absence of spatial error dependence, but can reject the presence of spatial error dependence when assuming linearity. When assuming no spatial error dependence, the log-linear model under the hypothesis H_0^5 is strongly rejected, but the linear model under hypothesis H_0^6 is not.

The DLR test statistic of hypothesis H_0^7 was 3.742, which rejects the null hypothesis marginally at the 5% level. The test statistic of 54.392 and the zero p-value for a linear model conditional on no spatial error dependence means that hypothesis H_0^8 is strongly rejected. Meanwhile test statistic for hypothesis H_0^9 is 2.250, so the hypothesis cannot be rejected at the 5% level.

The results presented in Table 1.2 show that outcomes of the DLR and LM tests are similar in the crime example except for H_0^7, under which the p-values were 0.053 for the DLR and 0.006 for the LM counterpart. The DLR tests do not need the second-order derivatives of the likelihood function (or the Hessian). The DLR test is a good alternative for testing functional form and spatial error dependence.

Among the models considered in this essay, a linear model with spatial error dependence should be specified when crime is modeled in its relationship with house values and household incomes in Columbus neighborhoods in Ohio. We will explore this data further in the next essay.

4. Monte Carlo results

This section describes the Monte Carlo experiment. It shows how the power of the DLR tests for functional form and spatial error dependence in finite samples. The model considered in the simulations is:

$$(I-\lambda W)y^{(r)} = (I-\lambda W)X^{(r)}\beta + (I-\lambda W)Z\gamma + \varepsilon\,,$$

where the number of crimes is represented by $y^{(r)}$ and household value and income by $X^{(r)}$. The only constant term is in Z. The Monte Carlo experiment follows the work done by Anselin and Ray (1991), Anselin et al. (1996), and Baltagi and Li (2001). The spatial weight matrix is computed by using 49 observations in the Anselin (1988a) data set. The number of replications is 1000. The regressors included in X are generated from a uniform (0, 10) distribution with its coefficient β set to be 1, while the constant term Z has a coefficient γ assumed to be 4. Error terms, ε, are randomly generated from a standard normal distribution. The DLR tests are evaluated at their asymptotic critical value at the 5% level. This essay follows the Monte Carlo setups in Baltagi and Li (2001). The power of the DLR tests is reported in Figures 1.2 through 1.10.

Figure 1. 2: Power of DLR joint test under hypothesis $H_0^1 : r = 0 \text{ and } \lambda = 0$

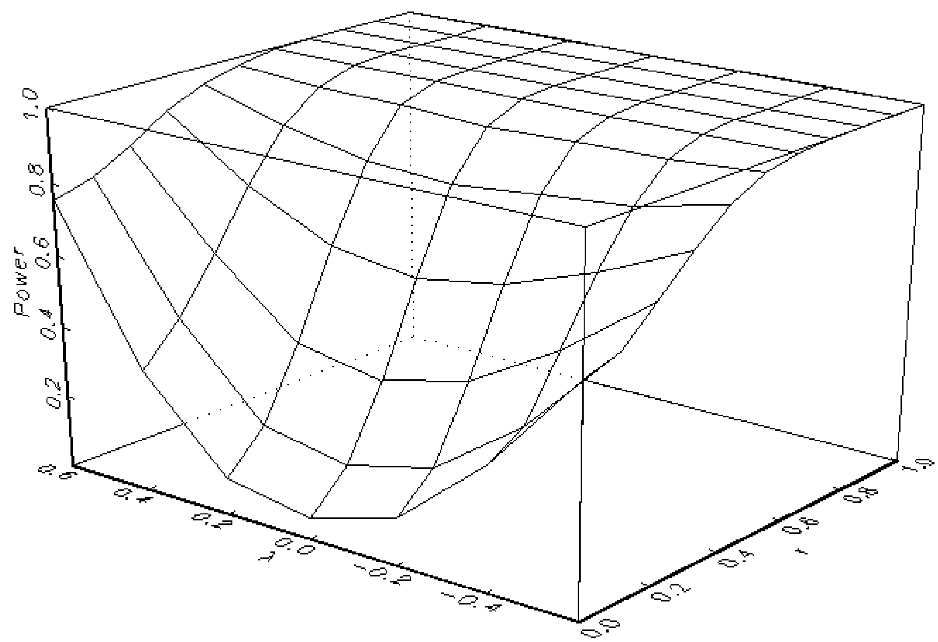
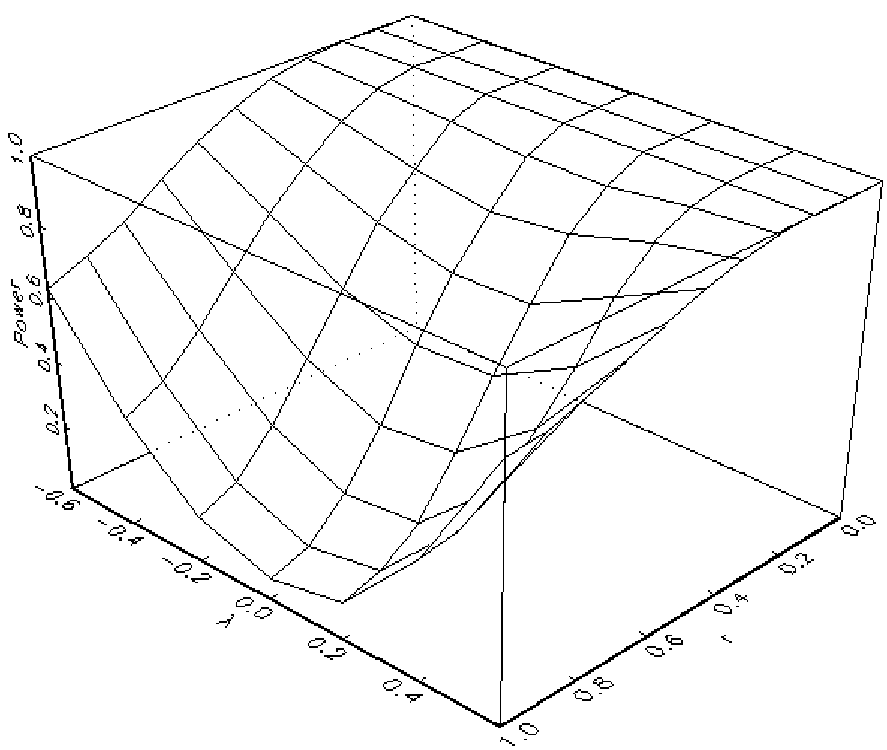


Figure 1. 3: Power of DLR joint test under hypothesis $H_0^2 : r = 1 \text{ and } \lambda = 0$



Figures 1.2 and 1.3 show the power of the DLR joint tests under the null hypotheses H_0^1 and H_0^2, respectively. The power increases when r and λ depart from their hypothesized values. More specifically, under the hypothesis H_0^1, the power dramatically increases when r moves toward 1 and λ moves toward either 1 or -1. Moreover, the power quickly converges to 100% of rejection when r moves away from 0. This result implies that the power of the DLR joint test under hypothesis H_0^1 would be 100% if the true model is linear. The power of the DLR joint test under the hypothesis H_0^2 also increases when r moves toward 0 and λ moves toward either 1 or -1. The power of this test quickly reaches 100% of rejection when r departs from 1, and it is 100% if the true model is log-linear.

Figure 1. 4: Power of DLR one-direction test under hypothesis $H_0^3 : \lambda = 0$ assuming $r = 0$

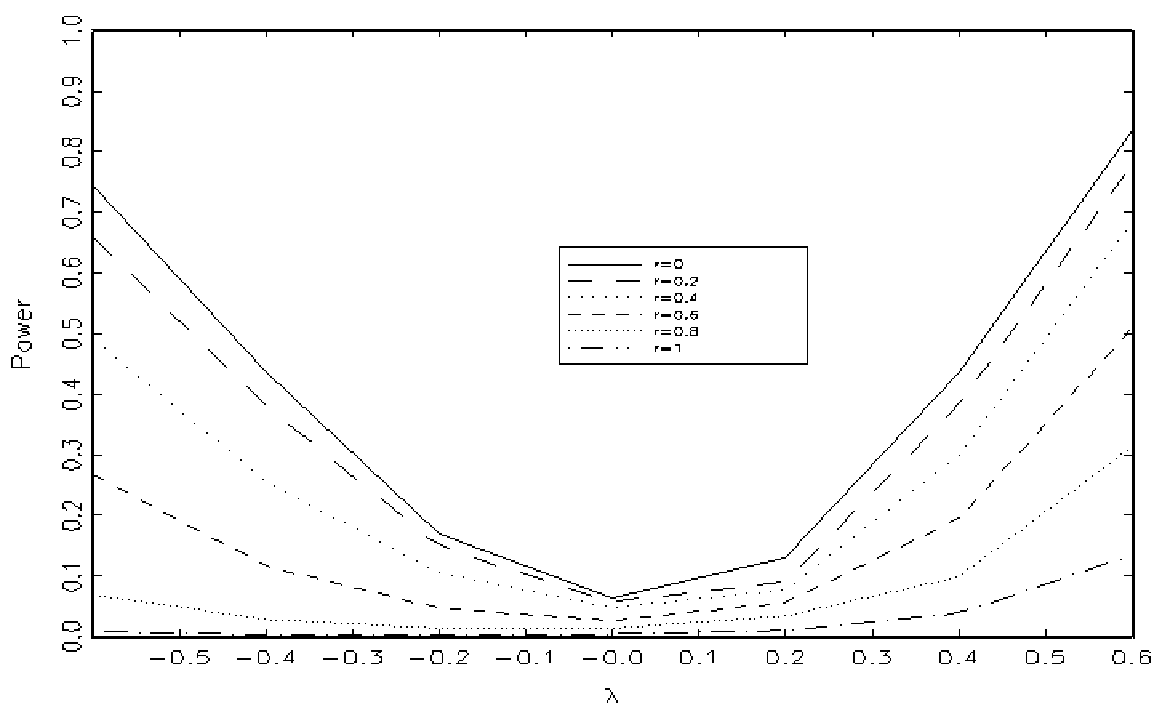


Figure 1.4 presents the power of the DLR one-direction test in 1000 replications under the hypothesis H_0^3 when the model is assumed to be a log-linear one. The power of the DLR

one-direction test under the hypothesis increases when the spatial coefficient moves away from its hypothesized value. The power of this test is also sensitive to a movement of the functional form coefficient.

Figure 1. 5: Power of DLR one-direction test under hypothesis $H_0^4 : \lambda = 0$ assuming $r = 1$

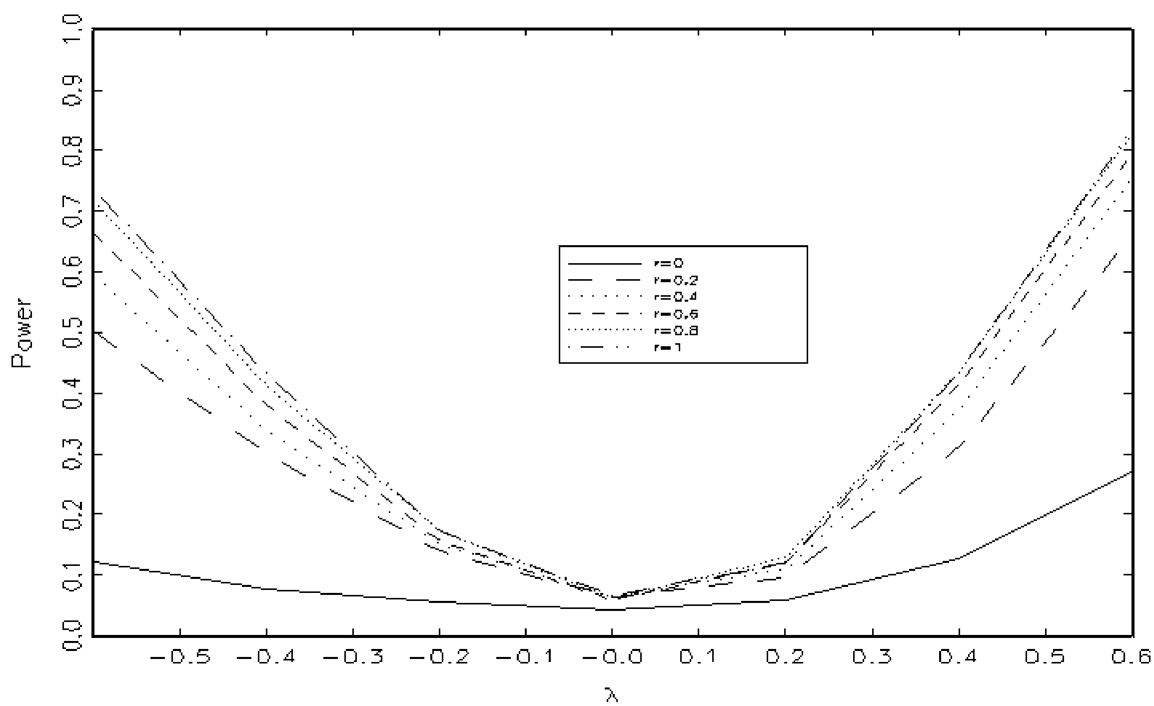


Figure 1.5 shows the power of the DLR one-direction test under the hypothesis H_0^4 : no spatial error dependence assuming a linear model. For $r = 1$, the power of this test increases when λ departs from its hypothesized value. In particular, it increases when the spatial error coefficient moves from 0 to either 1 or -1. However, the power of this test is also sensitive to departures of the functional form coefficient from 1 to 0.

Figure 1.6 shows the power of the DLR one-direction test under the hypothesis H_0^5. The power of this test increases when the functional form coefficient departs from 0 and moves toward 1, regardless of the true spatial error dependence. It quickly converges to 100% of rejection when the log-linear functional form moves toward linearity.

Figure 1. 6: Power of DLR one-direction test under hypothesis $H_0^5 : r = 0$ **assuming** $\lambda = 0$

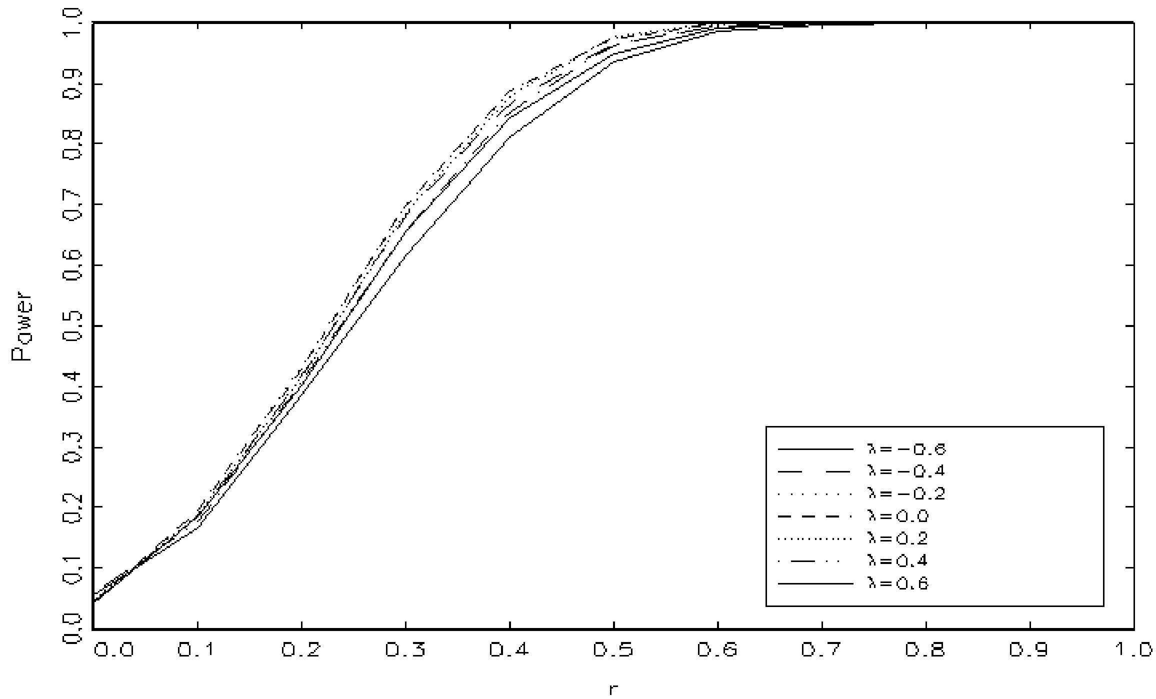


Figure 1.7 on the next page provides the power of the DLR one-direction test under the hypothesis H_0^6: linearity assuming no spatial error dependence. When $\lambda = 0$, the power of this DLR test increases as the functional coefficient moves far away from its null hypothesized value. Under the hypothesis, however, the frequency of rejections quickly converges to 100% regardless of the true spatial error dependence coefficients. It is likely that the power of the DLR test under the hypothesis H_0^6 is not sensitive to movements of the spatial error coefficient from 0 to either 1 or -1.

Figure 1. 7: Power of DLR one-direction test under hypothesis $H_0^6 : r = 1$ assuming $\lambda = 0$

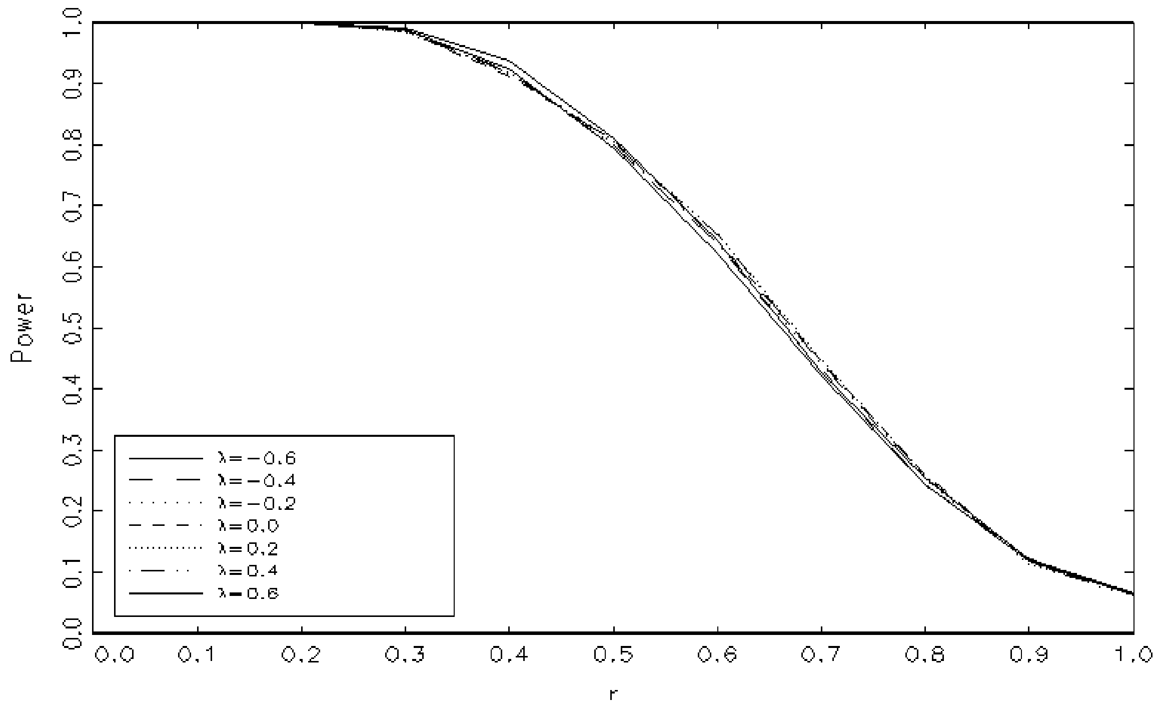


Figure 1. 8: Power of DLR conditional test under hypothesis $H_0^7 : \lambda = 0 | unkown \;\; r$

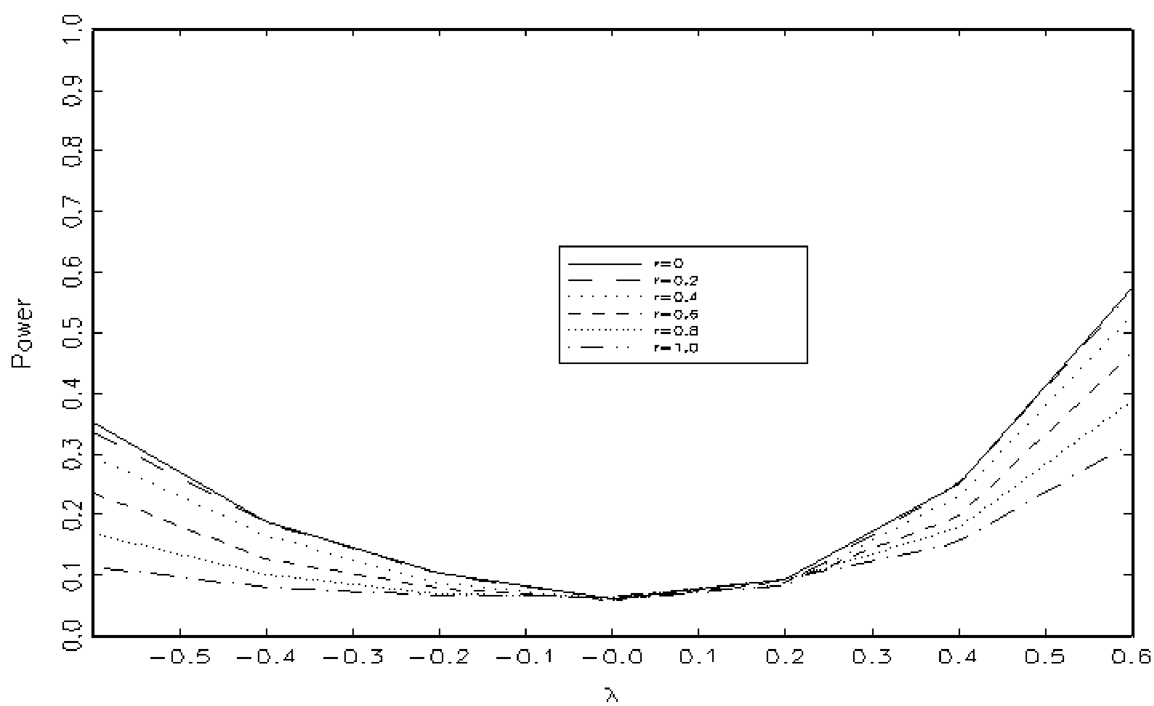


Figure 1.8 on the preceding page plots the power of the DLR conditional test for spatial error dependence conditional on a general Box-Cox model. The power of this test depends on both the functional form and the spatial error dependence. It increases when the spatial coefficient departs from its hypothesized value of zero.

Figure 1.9 shows the power of the DLR conditional test under the hypothesis H_0^8. The power of this test increases when the functional form coefficient moves away from its hypothesized value of zero. The power of this DLR conditional test converges to 100% of rejection as the functional form moves from log-linearity to linearity.

Figure 1. 9: Power of DLR conditional test under hypothesis $H_0^8 : r = 0 | unkown\ \lambda$

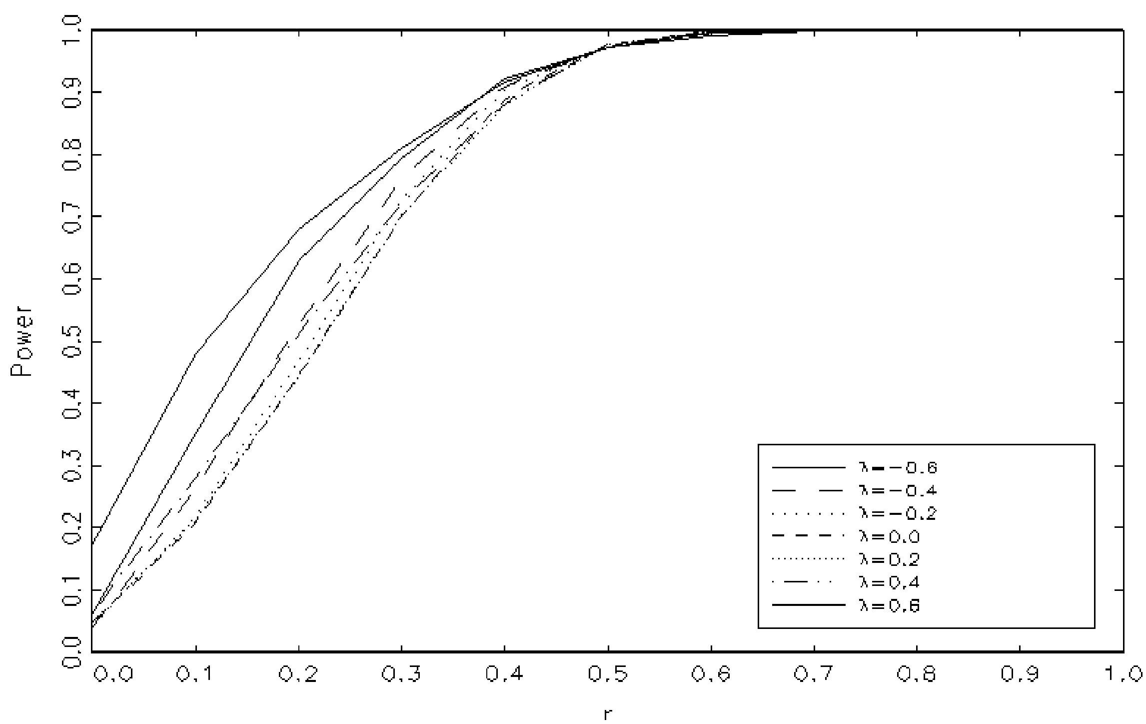


Figure 1. 10: Power of the DLR conditional test under hypothesis $H_0^9 : r = 1 | unkown\ \lambda$

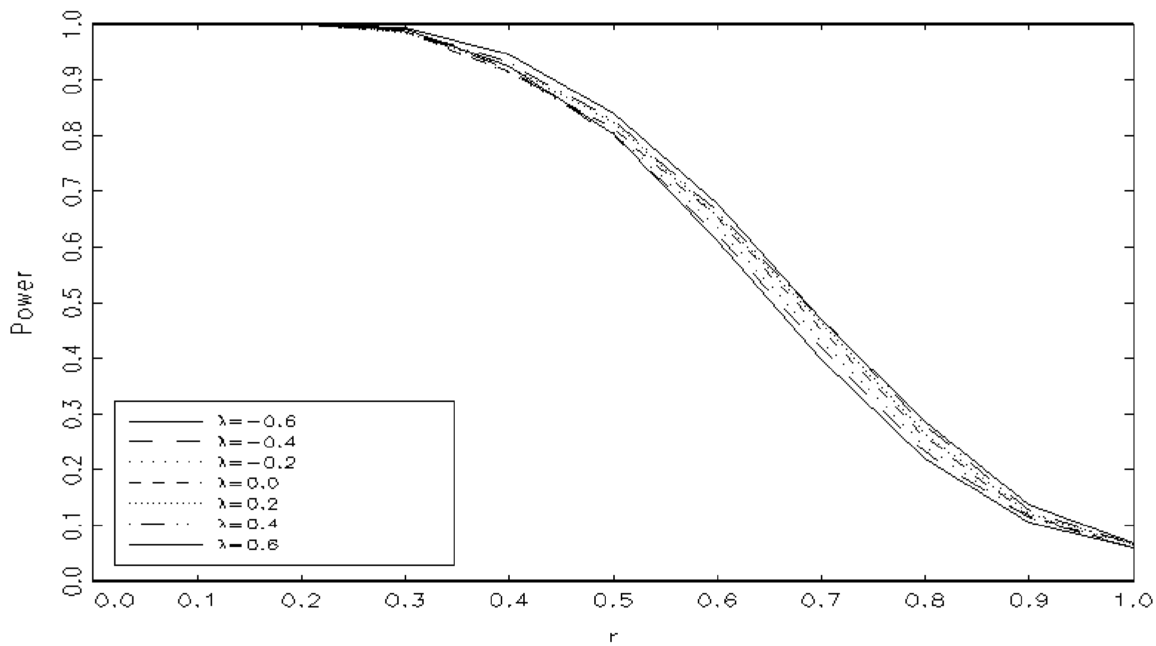


Figure 1.10 plots the power of the DLR conditional test under the hypothesis H_0^9. The figure shows that the power of the DLR test increases when the functional form coefficient departs from its null hypothesized value of 1. It also converges to 100% of rejection when the functional form closes to log-linearity.

The Monte Carlo experiment results have shown that the power of the DLR tests increases when the functional coefficient and/or the spatial coefficient deviates from their hypothesized value(s). The tests perform reasonably well. They are more sensitive to the functional coefficient than to the spatial error coefficient. The DLR tests perform similarly to their LM counterparts in Baltagi and Li (2001) in the Monte Carlo simulations. Our graphs show almost identical patterns as those shown in Baltagi and Li (2001).

5. Conclusions

This essay derived the DLR joint, one-direction, and conditional tests for testing functional form and spatial error dependence. These DLR tests were illustrated by using the Anselin (1988a) crime data set. The results showed that the DLR tests performed similarly to their LM counterparts in the empirical example.

The Monte Carlo experiments showed the power of the DLR tests proposed in this essay. The power of the DLR tests is more sensitive to functional misspecification than spatial error dependence. The power of the DLR tests also increases when functional and spatial coefficients depart from their hypothesized values. Thus, when testing for functional form and spatial error dependence, ignoring either functional form or spatial error dependence would lead to misspecification problems and eventually misleading inferences.

The DLR tests do not require the second-order derivatives of the likelihood function, but they provide similar test power as their LM counterparts. The DLR tests are easy to implement and provide researchers/practitioners an alternative method for testing functional form and spatial error dependence.

Essay 2: Double Length Regression Tests for Testing Functional Form and Spatial Lag Dependence

1. Introduction

Spatial econometric methods have been adopted increasingly in regional science, urban economics, and economic geography. Spatial econometric methods have also been used in empirical studies in international economics, labor economics, agricultural and environmental economics, public economics, and public finance (Anselin, 1999). Some methodological studies in spatial econometrics included Cliff and Ord (1981), Anselin (1988), LeSage (1998a), Anselin (2006), and Anselin et al. (2007). Although spatial dependence can violate the classical Gauss-Markov assumptions when regressions are run and performing hypotheses tested (LeSage, 1998), traditional econometrics has largely ignored spatial dependence. Testing for spatial effects, therefore, becomes important for estimating the spatial effects of economic activities in empirical research. Based on theoretical spatial models, many studies have tested for spatial effects in empirical work (Cliff and Ord, 1972; Anselin, 1985; Anselin, 1988b; Anselin and Rey, 1991; Anselin et al., 1996; Baltagi et al., 2002; Pinkse, 1998; Pinkse and Slade, 1998; Anselin, 1999; and Anselin, 2006). These spatial tests can be categorized into two types. The first is the Moran's I test, developed by Moran (1948, 1950) and further extended by Cliff and Ord (1972). The other category includes the maximum likelihood-based tests including the Lagrangian Multiplier (LM) and the Rao Scores (RS) tests. The LM and RS tests were further extended by Anselin (1988b), Pinkse (1998), Pinkse and Slade (1998), and Anselin (2001). Those tests, however, assume that the functional form is known, and that no control for functional form is needed. In practice, it is hard to know the true functional form of the models used in empirical studies. Researchers have to develop appropriate models that describe the relationship of variables and spatial effects as well. In the literature, however, very few studies have attempted to develop methods for testing functional form and spatial dependence (Baltagi and Li, 2001; Baltagi and Li, 2004; and Yang et al., 2006).

Spatial dependence usually includes two types of dependence: spatial error dependence and spatial lag dependence (Anselin, 1988a). In the previous essay, we derived double length regression (DLR) joint, one-direction, and conditional tests for functional form and spatial error dependence. This essay develops DLR tests for functional form and spatial lag dependence. Three different classes of DLR tests will be proposed. The first is the DLR joint test. This test is for linearity or log-linearity with no spatial lag dependence against a Box-Cox alternative with spatial lag dependence. The second is the DLR one-direction test. This test is used for testing linear/log-linear models that assume no spatial lag dependence and for testing spatial lag dependence that assume linearity/log-linearity. The third is the DLR conditional test. It is used for determining the possible presence of spatial lag dependence when testing for functional form and for a possible nonlinear functional form when testing for spatial lag dependence. Additionally, Monte Carlo experiments were performed to investigate the finite sample properties of the DLR tests.

The results of this essay suggested that DLR-based tests provide computationally cheaper alternatives for testing functional form and spatial lag dependence compared to their LM counterparts, since the DLR tests do not require the second-order derivatives of the likelihood function or a Hessian matrix. Both the LM and DLR tests perform similarly in Monte Carlo simulations.

The remainder of this essay is organized as follows. Section 2 develops the use of DLR tests for testing functional form and spatial lag dependence. In this section, three classes of DLR tests and nine different null hypotheses are considered. Section 3 presents an empirical example. Section 4 describes the Monte Carlo experiments and presents their results. Section 5 concludes the essay.

2. Spatial lag dependence and DLR tests

In this section, I develop the DLR model and its tests for functional form and spatial lag dependence. In particular, I derive the three DLR tests with nine null hypotheses. This section starts with the conventional Box-Cox model with spatial lag dependence, which has the following form:

$$y^{(r)} = \lambda W y^{(r)} + X^{(r)}\beta + Z\gamma + \varepsilon \qquad (2.1)$$

where λ is the spatial autoregressive/lag coefficient that satisfies $-1 < \lambda < 1$; and β and γ are coefficients that need to be estimated. Both $y^{(r)}$ and $X^{(r)}$ are required to take positive values and follow the Box-Cox transformation. When $r = 1$, the model in equation (2.1) is linear, $y = \lambda W y + X\beta + Z\gamma + \varepsilon$, and if $r = 0$, the model becomes log-linear, $\log(y) = \lambda W \log(y) + \log(X)\beta + Z\gamma + \varepsilon$. Variable Z is not subject to the Box-Cox transformation, and it could include a time trend, dummy variables, and the intercept. W is an $T \times T$ matrix of known spatial weights, while r is a scalar that varies from zero to one. ε is an $T \times 1$ residual vector that is assumed to follow the normal distribution N $(0, \delta_{\varepsilon}^{2}$ I).

The model in (2.1) can be rewritten as:

$$(I - \lambda W) y^{(r)} = X^{(r)}\beta + Z\gamma + \varepsilon \qquad (2.2)$$

where I is an identity matrix; $y^{(r)}$ is $T \times 1$; $X^{(r)}$ is $T \times k$; Z is $T \times s$; W is $T \times T$; β is $k \times 1$; γ is $s \times 1$; and ε is $T \times 1$. Applying the DLR requires error terms, ε, in (2.2) following the normal standard distribution N (0, I) that can be obtained by dividing both sides of equation (2.2) by standard deviation, σ.

$$\frac{1}{\sigma}(I - \lambda W) y^{(r)} = \frac{1}{\sigma}(X^{(r)}\beta + Z\gamma + \varepsilon)$$

where $\frac{\varepsilon}{\sigma} \sim N(0, I)$.

Because $y^{(r)}$ and $X^{(r)}$ follow the Box-Cox transformation, the OLS estimation technique does not work. To estimate the parameters, the maximum likelihood method is often used. The log likelihood function for a general Box-Cox and spatial lag dependence model is given by:

$$Log(L) = -\frac{T}{2}\log(2\pi) - T\log(\sigma) + \log\left|I - \lambda W\right| + (r-1)\sum_{t=1}^{T}\log(y_t) -$$

$$\frac{1}{2\sigma^2}\left[(I - \lambda W) y^{(r)} - X^{(r)}\beta - Z\gamma\right]'\left[(I - \lambda W) y^{(r)} - X^{(r)}\beta - Z\gamma\right]$$

This log-likelihood function consists of two components: the Jacobian terms and the sum of squared residuals. By defining the normalized standard residuals to be $f(y, \theta)$ and the Jacobian term to be $k(y, \theta)$, we have:

$$f(y, \theta) = \frac{1}{\sigma}\left[(I - \lambda W) y^{(r)} - X^{(r)}\beta - Z\gamma\right], \text{ and} \qquad (2.3)$$

$$k(y_t, \theta) = -T\log(\sigma) + \sum_{t=1}^{T}\log(1 - \lambda\varpi_t) + (r-1)\sum_{t=1}^{T}\log(y_t),$$

where $\theta = (\sigma, \beta', \gamma', \lambda, r)'$

Ord (1975) and Anselin (1988b) showed that $\log|I - \lambda W| = \sum_{t=1}^{T} \log(1 - \lambda \varpi_t)$, where ϖ_t's are eigenvalues of the matrix W. So a typical element of $k(y, \theta)$ is:

$$k_t(y_t, \theta) = -\log(\sigma) + \log(1 - \lambda \varpi_t) + (r - 1)\log(y_t) \qquad (2.4)$$

Also, we define $F_t(y_t, \theta) = \dfrac{\partial f_t(y_t, \theta)}{\partial \theta}$ and $K_t(y_t, \theta) = \dfrac{\partial k_t(y_t, \theta)}{\partial \theta}$

The DLR can be described as an artificial regression with double number of observations:

$$\begin{bmatrix} f(y,\theta) \\ \iota_T \end{bmatrix} = \begin{bmatrix} -F(y,\theta) \\ K(y,\theta) \end{bmatrix} b + residuals \qquad (2.5)$$

where b is a vector of $k \times 1$. $f(y, \theta)$ is as defined in (2.3) and its typical element is $f_t(y_t, \theta)$. ι is a unity vector of $T \times 1$. $K(y, \theta)$ is an $T \times (k + s + 3)$ matrix whose typical element is $K_t(y_t, \theta)$ for $t = 1,\ 2, ... T$.

The model in (2.5) can be expressed as:

$$\begin{bmatrix} \dfrac{\hat{\varepsilon}_i}{\sigma} \\ \iota_T \end{bmatrix} = \begin{bmatrix} -\dfrac{\partial f(\cdot)}{\partial \sigma} & -\dfrac{\partial f(\cdot)}{\partial \beta} & -\dfrac{\partial f(\cdot)}{\partial \gamma} & -\dfrac{\partial f(\cdot)}{\partial \lambda} & -\dfrac{\partial f(\cdot)}{\partial r} \\ \dfrac{\partial k(\cdot)}{\partial \sigma} & \dfrac{\partial k(\cdot)}{\partial \beta} & \dfrac{\partial k(\cdot)}{\partial \gamma} & \dfrac{\partial k(\cdot)}{\partial \lambda} & \dfrac{\partial k(\cdot)}{\partial r} \end{bmatrix} b \ + \ residuals \qquad (2.6)$$

In this model, the dependent variable is the normalized standard residuals for observations from 1 to T and unity for observations from $T + 1$ to $2T$, while the regressor is the negative first-order derivatives of the normalized standard residuals with respect to θ for observations from 1 to T and the first-order derivatives of the Jacobian terms with respect to θ for observations from $T+1$ to $2T$, where $\theta = (\sigma, \beta', \gamma', \lambda, r)'$.

Typical elements of regressors F_t and K_t are derived as:

For σ: $\dfrac{\partial f_t(y_t,\theta)}{\partial \sigma} = -\dfrac{1}{\sigma^2}(I-\lambda W)y^{(r)} - X^{(r)}\beta - Z\gamma$ $\qquad \dfrac{\partial k_t(y_t,\theta)}{\partial \sigma} = -\dfrac{1}{\sigma}$

For β: $\dfrac{\partial f_t(y_t,\theta)}{\partial \beta} = -\dfrac{1}{\sigma}X^{(r)}$ $\qquad \dfrac{\partial k_t(y_t,\theta)}{\partial \beta} = 0$

For γ: $\dfrac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\dfrac{1}{\sigma}Z$ $\qquad \dfrac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$

For λ: $\dfrac{\partial f_t(y_t,\theta)}{\partial \lambda} = -\dfrac{1}{\sigma}Wy^{(r)}$ $\qquad \dfrac{\partial k_t(y_t,\theta)}{\partial \lambda} = -\dfrac{\varpi_t}{(1-\lambda\varpi_t)}$

For r: $\dfrac{\partial f_t(y_t,\theta)}{\partial r} = \dfrac{1}{\sigma}(I-\lambda W)C(y,r) - C(x,r)\beta$ $\qquad \dfrac{\partial k_t(y_t,\theta)}{\partial r} = \log y_t$

where $C(y, r) = \dfrac{1}{r^2}[y^r(r\log y - 1) + 1]$.

Davidson and MacKinnon (1993) showed that total sum of squares (SST) of the dependent variables in (2.6) is $2T$. The value of $2T - SSR$[2] always equals the explained sum of squares, and it provides an asymptotically valid test statistic. In this case, $2n - SSR$ asymptotically follows the Chi-square distribution with the degrees of freedom being equal to the number of restrictions under the null hypotheses.

[2] SSR stands for the sum of squares of residuals.

2.1. DLR joint tests

Two DLR joint tests were derived to test for log-linear or linear models with no spatial lag dependence. First, we derived the DLR joint test for the null hypothesis $H_0^{10}: r = 0 \textit{ and } \lambda = 0$. Under this hypothesis, the model is log-linear with no spatial lag dependence. Since $y^{(0)} = \log(y)$ and $C(y,\theta) = \lim_{r\to 0} C(y,\theta) = \frac{1}{2}(\log y)^2$, its typical regressor elements were derived as:

$$\frac{\partial f_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma^2}\left[\log y_t - (\log x_t)\beta - z_t\gamma\right] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \beta} = -\frac{1}{\sigma}\log x_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \beta} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\frac{1}{\sigma} z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \lambda} = -\frac{1}{\sigma}\varpi_t \log y_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \lambda} = \varpi_t$$

$$\frac{\partial f_t(y_t,\theta)}{\partial r} = \frac{1}{2\sigma}[(\log y_t)^2 - (\log x_t)^2\beta] \qquad \frac{\partial k_t(y_t,\theta)}{\partial r} = \log y_t$$

where ϖ_t is eigenvalues of the spatial weight matrix.

Second, we developed the DLR joint test for the hypothesis $H_0^{11}: r = 1 \textit{ and } \lambda = 0$. Under this hypothesis, our model becomes linear with no spatial lag dependence. Because $y^{(1)} = y - 1$ and $C(y,\theta) = \lim_{r\to 1} C(y,\theta) = y\log y - y + 1$, the typical regressor elements of the model are given as:

$$\frac{\partial f_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma^2}[(y_t - 1) - (x_t - 1)\beta - z_t\gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \beta} = -\frac{1}{\sigma}(x_t - 1) \qquad \frac{\partial k_t(y_t,\theta)}{\partial \beta} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\frac{1}{\sigma} z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \lambda} = -\frac{1}{\sigma}\varpi_t(y_t - 1) \qquad \frac{\partial k_t(y_t,\theta)}{\partial \lambda} = -\varpi_t$$

$$\frac{\partial f_t(y_t,\theta)}{\partial r} = \frac{1}{\sigma}[(y_t(\log y_t - 1) + 1) - (x_t(\log x_t - 1) + 1)\beta] \qquad \frac{\partial k_t(y_t,\theta)}{\partial r} = \log y_t$$

2.2. DLR one-direction tests

This subsection describes my derivation of the DLR one-direction tests that are used for testing these two cases: when the functional form is known and the spatial lag dependence is tested and when no spatial lag dependence is known and the functional form is tested. These tests ignore the possibility of unknown functional form when spatial lag dependence is tested and unknown spatial lag dependence when functional form is tested. As with the DLR joint tests, the DLR one-direction tests are also estimated by using the OLS technique, and the DLR test statistic is asymptotically distributed as Chi-square with the degrees of freedom being equal to the number of restrictions under the null hypotheses. In this subsection, I propose the DLR one-direction test for four hypotheses.

Under the hypothesis $H_0^{12} : \lambda = 0$ assuming $r = 0$, our model is given as log-linearity with no spatial lag dependence. The hypothesis means that log-linearity is known, while the spatial lag dependence needs to be tested as in.

$$\frac{1}{\sigma}(I-\lambda W)\log(y)=\frac{1}{\sigma}\left[\log(X)\beta+Z\gamma\right]+\varepsilon^*$$

where $\varepsilon^* \sim N(0,I)$. In this case, the parameters of the DLR model are $\theta=(\sigma,\beta',\gamma',\lambda)'$. Using the DLR model presented in the previous section, we can derive the typical regressor elements for this model as:

$$\frac{\partial f_t(y_t,\theta)}{\partial\sigma}=-\frac{1}{\sigma^2}[\log y_t-\log x_t\beta-z_t\gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial\sigma}=-\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial\beta}=-\frac{1}{\sigma}\log x_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial\beta}=0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial\gamma}=-\frac{1}{\sigma}z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial\gamma}=0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial\lambda}=-\frac{1}{\sigma}\varpi_t\log y_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial\lambda}=-\varpi_t$$

Under the null hypothesis $H_0^{13}:\lambda=0$ assuming $r=1$, the model is now given as linearity with no spatial lag dependence. Thus, the spatial lag dependence needs to be tested while a linear model is known.

$$\frac{1}{\sigma}(I-\lambda W)(y-1)=\frac{1}{\sigma}\left[(X-1)\beta+Z\gamma\right]+\varepsilon^*$$

where $\varepsilon^* \sim N(0,I)$, and $\theta=(\sigma,\beta',\gamma',\lambda)'$.

Typical elements of regressor of the DLR model are:

$$\frac{\partial f_t(y_t,\theta)}{\partial\sigma}=-\frac{1}{\sigma^2}[(y_t-1)-(x_t-1)\beta-z_t\gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial\sigma}=-\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial\beta}=-\frac{1}{\sigma}(x_t-1) \qquad \frac{\partial k_t(y_t,\theta)}{\partial\beta}=0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\frac{1}{\sigma} z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \lambda} = -\frac{1}{\sigma} \varpi_t (y_t - 1) \qquad \frac{\partial k_t(y_t,\theta)}{\partial \lambda} = -\varpi_t$$

Under the hypothesis $H_0^{14}: r = 0$ assuming $\lambda = 0$, the model becomes log-linear assuming no spatial lag dependence. In this case, we need to test for functional form when we assume that there is no spatial lag dependence in the model. The parameter of the DLR is $\theta = (\sigma, \beta', \gamma', r)'$, and the typical elements of the regressor are:

$$\frac{\partial f_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma^2}[\log y_t - \log x_t \beta - z_t \gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \beta} = -\frac{1}{\sigma} \log x_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \beta} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\frac{1}{\sigma} z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial r} = -\frac{1}{2\sigma}[(\log y_t)^2 - (\log x_t)^2 \beta] \qquad \frac{\partial k_t(y_t,\theta)}{\partial r} = \log(y_t)$$

Under the hypothesis $H_0^{15}: r = 1$ assuming $\lambda = 0$, the model becomes linear when we assume no spatial lag dependence. The parameter of the DLR model is $\theta = (\sigma, \beta', \gamma', r)'$ and typical elements of regressor are:

$$\frac{\partial f_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma^2}[(y_t - 1) - (x_t - 1)\beta - z_t \gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \beta} = -\frac{1}{\sigma}(x_t - 1) \qquad \frac{\partial k_t(y_t,\theta)}{\partial \beta} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\frac{1}{\sigma} z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial r} = \frac{1}{\sigma}[y_t(\log y_t - 1) + 1 - (x_t(\log x_t - 1) - 1)\beta] \qquad \frac{\partial k_t(y_t,\theta)}{\partial r} = \log(y_t)$$

2.3. DLR conditional tests

In the practice, the DLR joint test cannot give the "true" model when the null hypothesis is rejected. Meanwhile, the DLR one-direction test ignores the possibility of the functional form being unknown when spatial lag dependence is tested and the spatial lag dependence being unknown when functional form is tested. To deal with these limitations, this subsection presents the derivation of the DLR conditional test that takes into account the possibility of spatial lag dependence when functional form is tested and possibility of misspecification of functional form when spatial lag dependence is tested.

DLR test for spatial lag dependence conditional on a general Box-Cox model

Under the hypothesis $H_0^{16} : \lambda = 0 | unkown \;\; r$, the model becomes a general Box-Cox model with no spatial lag dependence, and its elements are conditionally derived as:

$$\frac{\partial f_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma^2}(y_t^{(r)} - x_t^{(r)}\beta - z_t\gamma) \qquad \frac{\partial k_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \beta} = -\frac{1}{\sigma} x_t^{(r)} \qquad \frac{\partial k_t(y_t,\theta)}{\partial \beta} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\frac{1}{\sigma} z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial\lambda}=-\frac{1}{\sigma}\varpi_t y_t^{(r)}$$

$$\frac{\partial k_t(y_t,\theta)}{\partial\lambda}=-\varpi_t$$

$$\frac{\partial f_t(y_t,\theta)}{\partial r}=\frac{1}{\sigma r^2}\left[[y^r(r\log y-1)+1]-[x^r(r\log x-1)+1]\beta]\right] \qquad \frac{\partial k_t(y_t,\theta)}{\partial r}=\log y_t$$

DLR test for a log-linear model conditional on spatial lag dependence

The test is for the null hypothesis $H_0^{17}: r=0|unkown\ \lambda$. The model is log-linear conditional on spatial lag dependence. Its typical elements are given as:

$$\frac{\partial f_t(y_t,\theta)}{\partial\sigma}=-\frac{1}{\sigma^2}\left[(i_t-\lambda\varpi_t)\log y_t-(\log x_t)\beta-z_t\gamma\right] \qquad \frac{\partial k_t(y_t,\theta)}{\partial\sigma}=-\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial\beta}=-\frac{1}{\sigma}\log x_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial\beta}=0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial\gamma}=-\frac{1}{\sigma}z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial\gamma}=0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial\lambda}=-\frac{1}{\sigma}\varpi_t\log y_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial\lambda}=-\frac{\varpi_t}{1-\lambda\varpi_t}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial r}=\frac{1}{2\sigma}[(i_t-\lambda\varpi_t)(\log y_t)^2-(\log x_t)^2\beta] \qquad \frac{\partial k_t(y_t,\theta)}{\partial r}=\log y_t$$

where i_t is the row t^{th} of the matrix I.

DLR test for a linear model conditional on spatial lag dependence

Under the hypothesis $H_0^{18}: r=1|unkown\ \lambda$, the DLR test for linearity conditional on spatial lag dependence. The typical elements of regressors are as:

$$\frac{\partial f_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma^2}[(i_t - \lambda\varpi_t)(y_t - 1) - (x_t - 1)\beta - z_t\gamma] \qquad \frac{\partial k_t(y_t,\theta)}{\partial \sigma} = -\frac{1}{\sigma}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \beta} = -\frac{1}{\sigma}(x_t - 1) \qquad \frac{\partial k_t(y_t,\theta)}{\partial \beta} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \gamma} = -\frac{1}{\sigma}z_t \qquad \frac{\partial k_t(y_t,\theta)}{\partial \gamma} = 0$$

$$\frac{\partial f_t(y_t,\theta)}{\partial \lambda} = -\frac{1}{\sigma}\varpi_t(y_t - 1) \qquad \frac{\partial k_t(y_t,\theta)}{\partial \lambda} = -\frac{\varpi_t}{1-\lambda\varpi_t}$$

$$\frac{\partial f_t(y_t,\theta)}{\partial r} = \frac{1}{\sigma}[(i_t - \lambda\varpi_t)(y_t(\log y_t - 1) + 1) - (x_t(\log x_t - 1) + 1)\beta] \qquad \frac{\partial k_t(y_t,\theta)}{\partial r} = \log y_t$$

3. Results of DLR test for crime data

This section shows how the DLR tests performed for testing functional form and spatial lag dependence in an empirical example. It presents the results of the DLR tests in comparison to the results of LM counterpart based on the same crime data set.

The previous essay described how to build the spatial weight matrix, which is also used in this essay. The data used in this example is presented in the Table 12.1 Anselin (1988a, p.189). The data was used to investigate how the number of crimes is related to the household incomes and housing values in the 49 neighborhoods in Columbus, Ohio in 1980. Crime was measured as the number of residential burglaries and vehicle thefts per thousand households in the neighborhood, while housing values and incomes were measured in thousands of dollars.

The dependent variable was the number of crimes, and the independent variables were housing values and household incomes. Both dependent and independent variables were subject to the Box-Cox transformation, while the intercept was not. The DLR test statistics for testing functional form and spatial lag dependence based on the crime data set are shown in the Table 2.1.

Table 2. 1: DLR test statistics for functional form and spatial lag dependence

	DLR Test	LM Test**
DLR joint tests		
$H_0^{10}: r=0 \ and \ \lambda=0$	56.406	85.411
	(0.000)*	(0.000)
$H_0^{11}: r=1 \ and \ \lambda=0$	10.552	13.952
	(0.001)	(0.001)
DLR One-Direction tests		
$H_0^{12}: \lambda=0$ assuming $r=0$	1.831	1.940
	(0.176)	(0.164)
$H_0^{13}: \lambda=0$ assuming $r=1$	9.758	13.914
	(0.002)	(0.001)
$H_0^{14}: r=0$ assuming $\lambda=0$	56.235	85. 406
	(0.000)	(0.000)
$H_0^{15}: r=1$ assuming $\lambda=0$	1.465	2.813
	(0.226)	(0.093)
DLR conditional tests		
$H_0^{16}: \lambda=0 \| unkown \ r$	6.449	10.164
	(0.011)	(0.001)
$H_0^{17}: r=0 \| unkown \ \lambda$	55.035	74.537
	(0.000)	(0.000)
$H_0^{18}: r=1 \| unkown \ \lambda$	0.638	0.123
	(0.424)	(0.725)

* Numbers in the parentheses are p-value.

** These numbers come from Baltagi and Li (2004a) with permission from Elsevier Ltd.

The table above provides the DLR and LM test statistics and their p-values. Both the DLR and LM tests are for functional form and spatial lag dependence based on the same crime data set. According to the results, the DLR joint test statistics for the null hypotheses H_0^{10} and H_0^{11} are 56.406 and 10.552, respectively. The results mean that log-linear and linear models with no spatial lag correlation are strongly rejected at the 5% level. The DLR test statistics provide the similar rejection/non-rejection as LM tests based on the same crime data set.

For the DLR one-direction tests, a model assuming log-linearity with no spatial lag dependence under the hypothesis H_0^{12} was not rejected at the 5% level, while a model assuming linearity with no spatial lag correlation under the hypothesis H_0^{13} was significantly rejected. In others words, when assuming log-linearity, one cannot reject the absence of spatial lag dependence while one can reject the presence of spatial lag dependence the linearity is assumed. When no spatial lag dependence is assumed, the log-linear model under the hypothesis H_0^{14} is rejected, while the linear model under the hypothesis H_0^{15} cannot be rejected at the 5% level. The outcomes of the DLR one-direction tests are similar to those of the LM counterpart.

The DLR test statistic of hypothesis H_0^{16} is 6.449 and its p-value is 0.011. It means that a general Box-Cox model with no spatial lag dependence is rejected. The DLR test statistic for a linear model conditional on no spatial lag dependence is 55.035 and has a zero p-value. It strongly rejects the null hypothesis H_0^{17} at the 5% level. Meanwhile, the DLR test statistic for a log-linear model conditional on no spatial lag dependence 0.638 implies the hypothesis H_0^{18} cannot be rejected at the 5% level. For the conditional tests, the DLR test provides the same conclusions in terms of rejection/non-rejection as LM counterparts based on the same crime data set.

The DLR results and LM tests are similar in the crime empirical example. More importantly, all DLR and LM test statistics provide the same rejection and non-rejection decisions. At this point, we can conclude that the linear model with spatial lag dependence is close to the real model, and it should be specified when crime is modeled in its relationship with household incomes and housing values in Columbus neighborhoods in Ohio. However, we are not sure whether spatial error or spatial lag dependence exists in the relationship.

4. Monte Carlo results

This section shows the performance of the DLR tests in the Monte Carlo experiments, which used the general Box-Cox model with spatial lag dependence in equation (2.2):

$$(I - \lambda W) y^{(r)} = X^{(r)} \beta + Z\gamma + \varepsilon$$

We followed the basic setups for Monte Carlo experiments in Anselin and Rey (1991), Anselin et al. (1996), and Baltagi and Li (2004a). The spatial weight matrix was computed by using 49 observations in the Anselin (1988a) crime data set. The number of replications was 1000. Housing values and household incomes in $X^{(r)}$ were generated from a uniform (0, 10) distribution with its coefficient β set to be 1. Both dependent and independent variables were subject to the Box-Cox transformation. The only constant term was in Z with its coefficient γ assumed to be 4, and Z did not follow the Box-Cox transformation. The error terms, ε, were randomly generated to follow the standard normal distribution. All the hypotheses were tested by using the DLR test statistics at the 5% level. The power of the DLR test for functional form and spatial lag dependence is shown in the Figures 2.1 to 2.9 on the following pages.

Figure 2.1 and 2.2 on the next page plot the power of the DLR joint tests under the hypotheses H_0^{10} and H_0^{11}, respectively. The power of the DLR joint tests increases when the functional form and spatial lag coefficients depart from their null hypothesized values. Under the hypothesis H_0^{10}, the power of the DLR test significantly increases when r moves from 0 to 1, and λ moves from 0 to either 1 or -1, and it quickly converges to 100% of rejection. This result implies that the power of the DLR joint test under the log-linear model with no spatial lag dependence is 100% if the "true" model is linearity.

Figure 2. 1: Power of DLR joint test under hypothesis $H_0^{10}: r = 0\ and\ \lambda = 0$

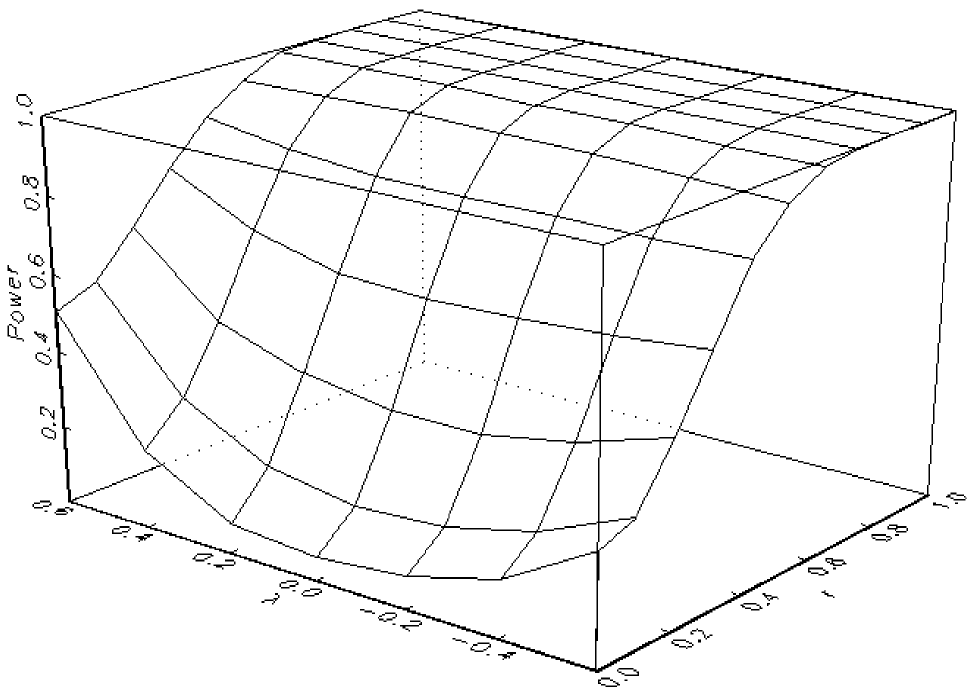
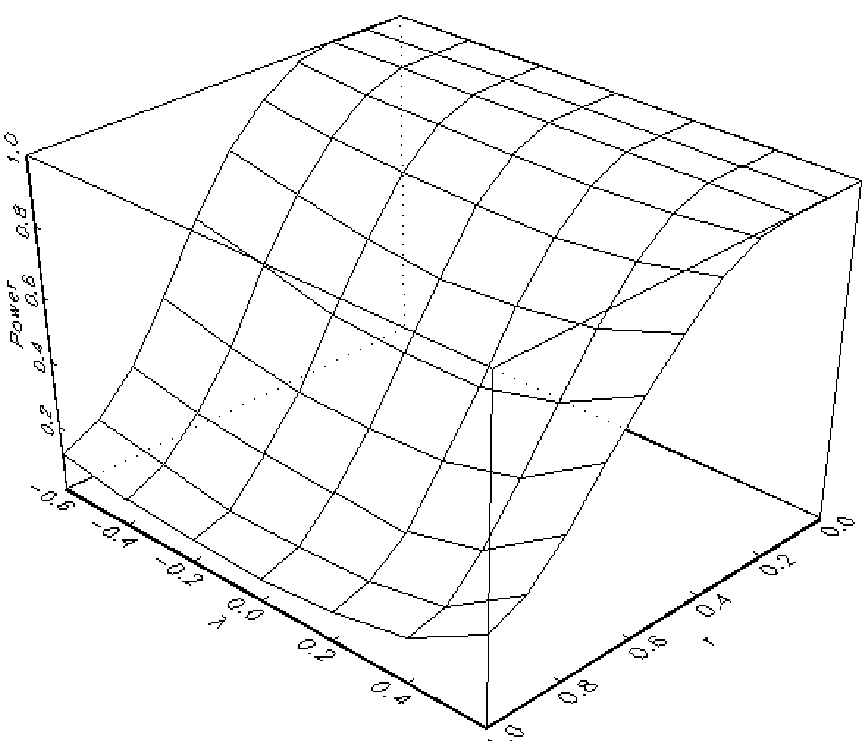


Figure 2. 2: Power of DLR joint test under hypothesis $H_0^{11}: r = 1\ and\ \lambda = 0$



Under the hypothesis H_0^{11}, the power of the DLR joint test increases when r departs toward 0, and λ moves from 0 either 1 or -1. It quickly reaches 100% of rejection as r and λ depart from their hypothesized values. The result implies the power of the DLR joint test under the linear model with spatial lag correlation is 100% if the true model is log-linear.

Figure 2. 3: Power of DLR one-direction test under hypothesis $H_0^{12}: \lambda = 0$ assuming $r = 0$

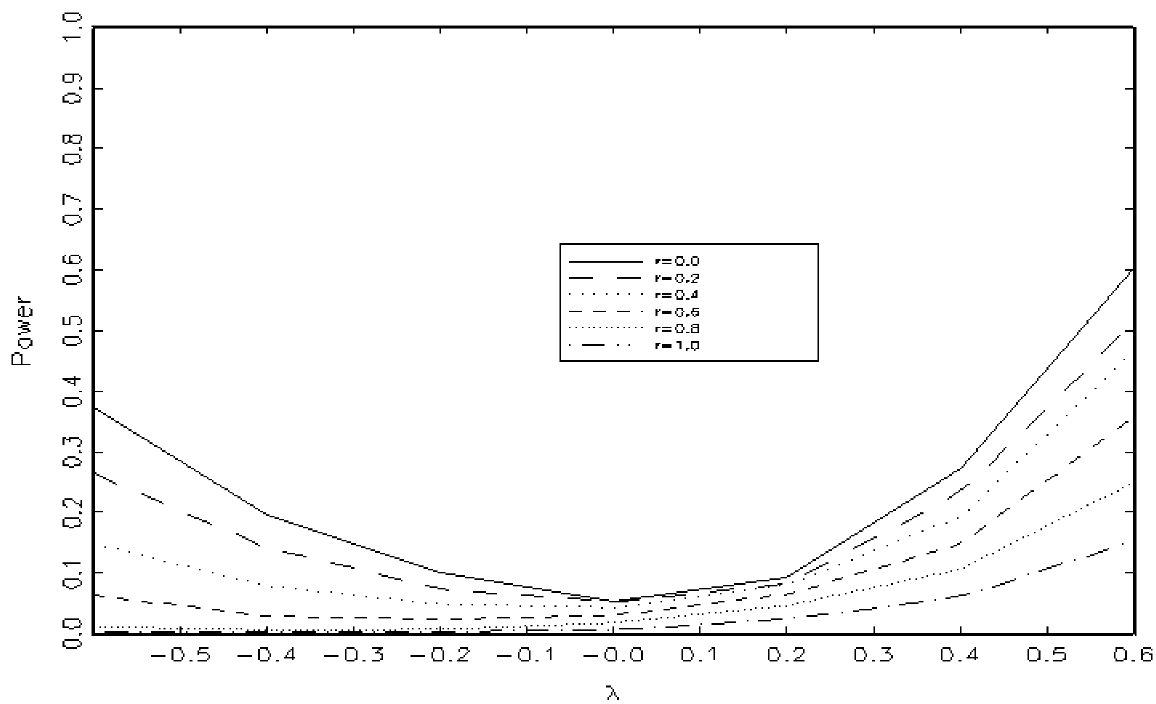


Figure 2.3 shows the power of the DLR one-direction test under the hypothesis H_0^{12}. When log-linearity is assumed, the power of the DLR one-direction test increases when the spatial coefficient moves far away from its hypothesized value of zero, or when the model moves from log-linearity to linearity. The power of the DLR one-direction test under this hypothesis is also sensitive to changes in the value of r.

Figure 2. 4: Power of DLR one-direction test under hypothesis $H_0^{13}: \lambda = 0$ assuming $r = 1$

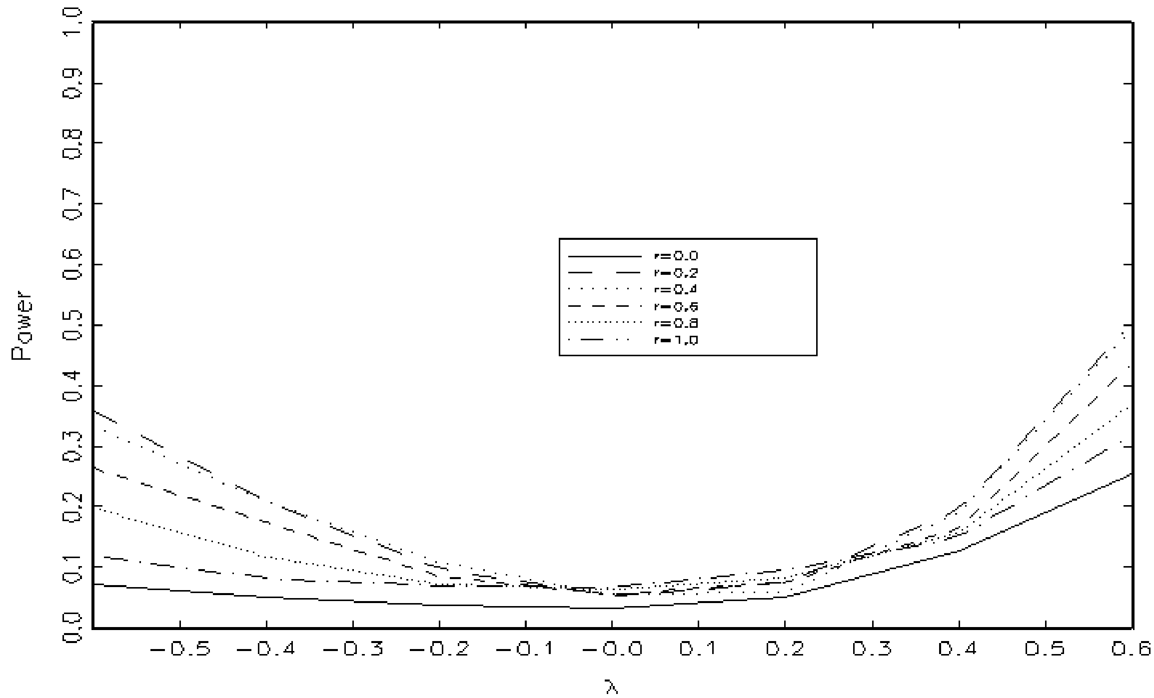


Figure 2. 5: Power of DLR one-direction test under hypothesis $H_0^{14}: r = 0$ assuming $\lambda = 0$

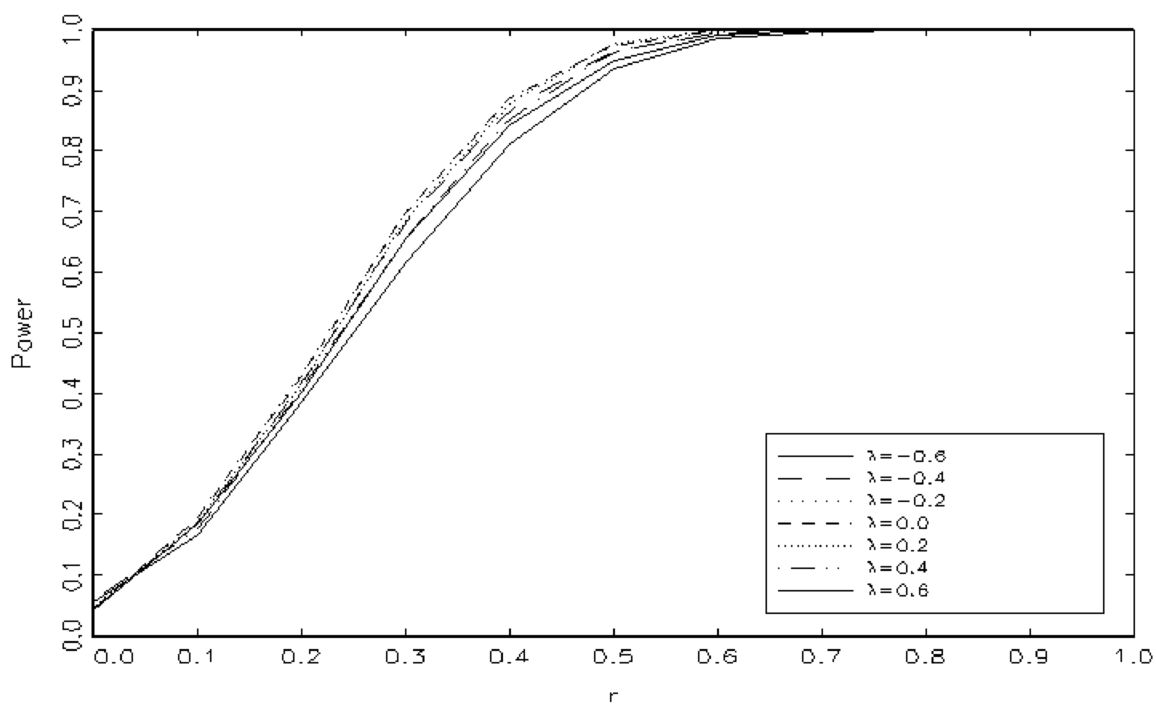


Figure 2.4 on the preceding page presents the power of the DLR one-direction test under the hypothesis H_0^{13} with 1000 replications. When linearity is assumed, the power of this test increases when λ departs from its hypothesized value of zero. More specially, the power of the DLR test under the hypothesis H_0^{13} increases as λ moves from 0 to either 1 or -1. The result also shows that the power of the DLR one-direction test is also sensitive to a change in functional form.

Figure 2.5 on the preceding page plots the power of the DLR one-direction test under the hypothesis H_0^{14}. The power of this test increases when r departs from 0 to 1, and it quickly converges to 100% rejection regardless of the values of spatial lag dependence. However, the power of this test is insensitive to a change in the spatial lag coefficient.

Figure 2. 6: Power of DLR one-direction test under hypothesis $H_0^{15} : r = 1$ assuming $\lambda = 0$

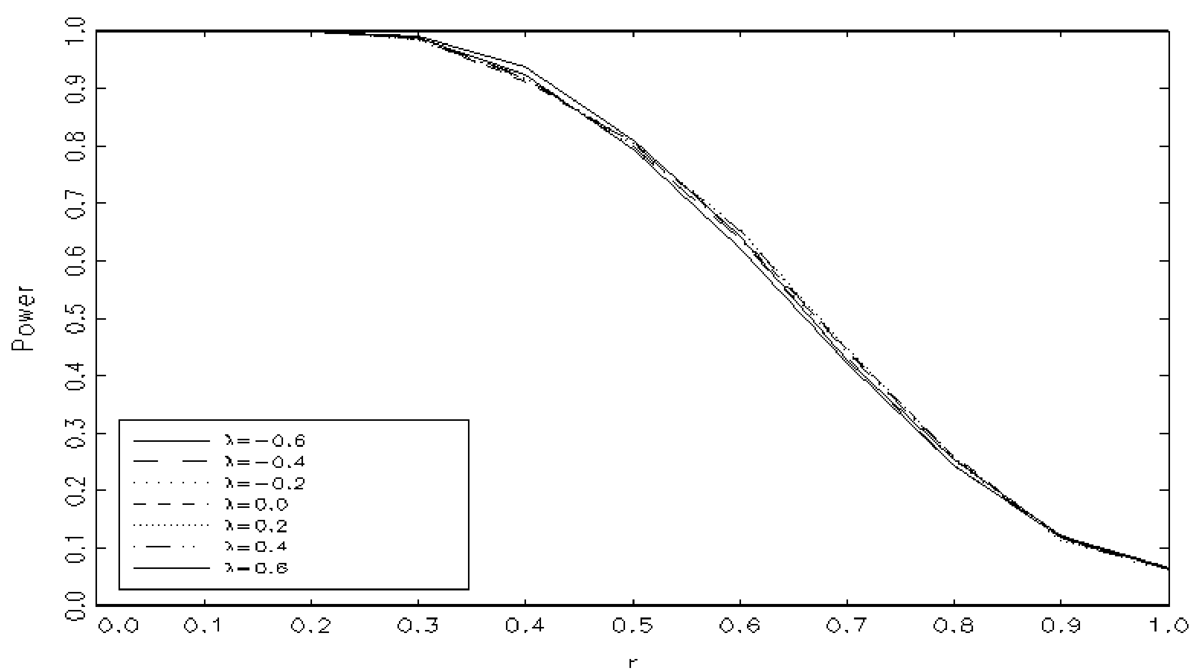


Figure 2.6 shows the power of the DLR one-direction test under the hypothesis H_0^{15} after 1000 replications. Under this hypothesis, the model becomes linear when no spatial lag dependence is assumed. The power of this test increases when r moves from 1 to 0. More clearly, when the functional form changes from linearity toward log-linearity, the power of the DLR one direction test under the hypothesis H_0^{15} converges quickly to 100% rejection regardless of the values of the spatial lag coefficient. Under this hypothesis, the power of the DLR one direction test is insensitive to spatial lag dependence.

Figure 2. 7: Power of DLR conditional test under hypothesis $H_0^{16} : \lambda = 0 | unkown\ r$

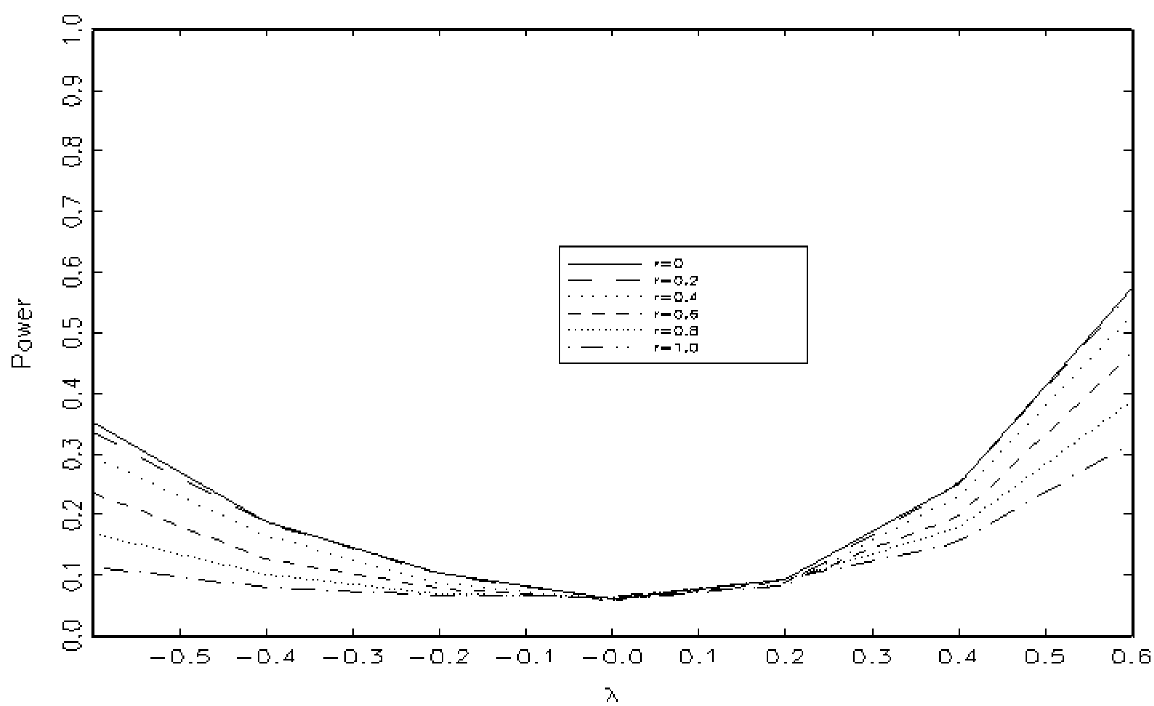


Figure 2.7 gives the power of the DLR conditional test for spatial lag dependence conditional on a general Box-Cox model. The power of the DLR conditional test under the hypothesis H_0^{16} depends on the values the spatial lag coefficient λ. It increases when λ departs from its

hypothesized value of zero to either 1 or -1. Functional form also has a small effect on the power of the DLR conditional test under this hypothesis.

Figure 2. 8: Power of DLR conditional test under hypothesis $H_0^{17}: r = 0 | unkown\ \lambda$

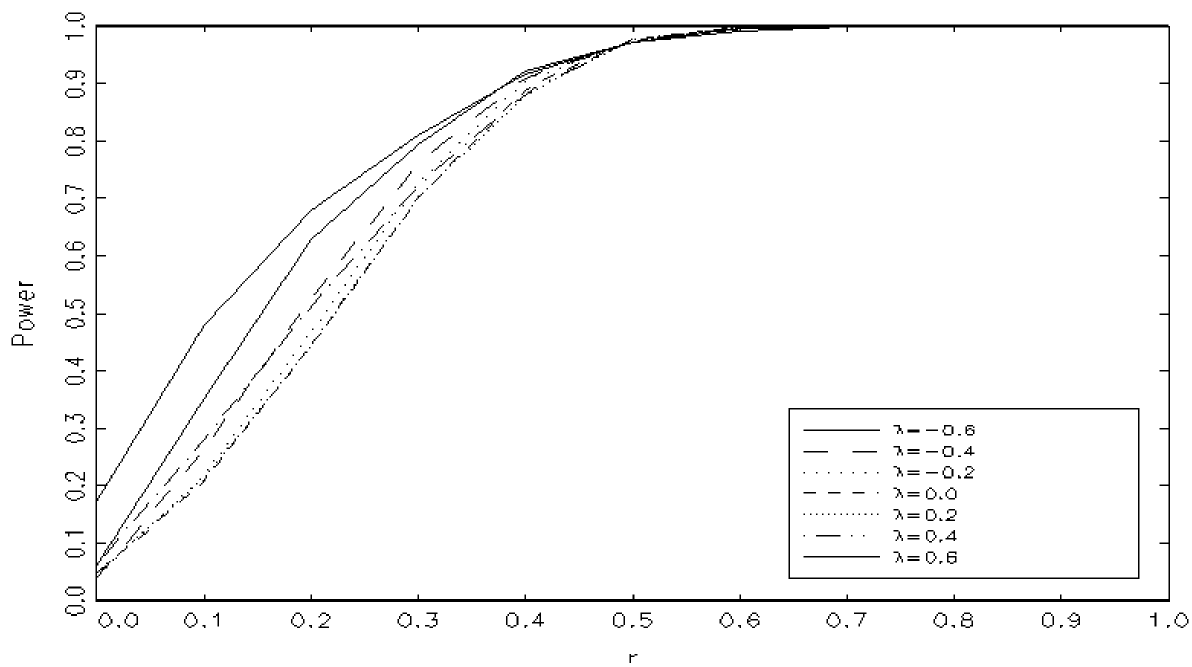
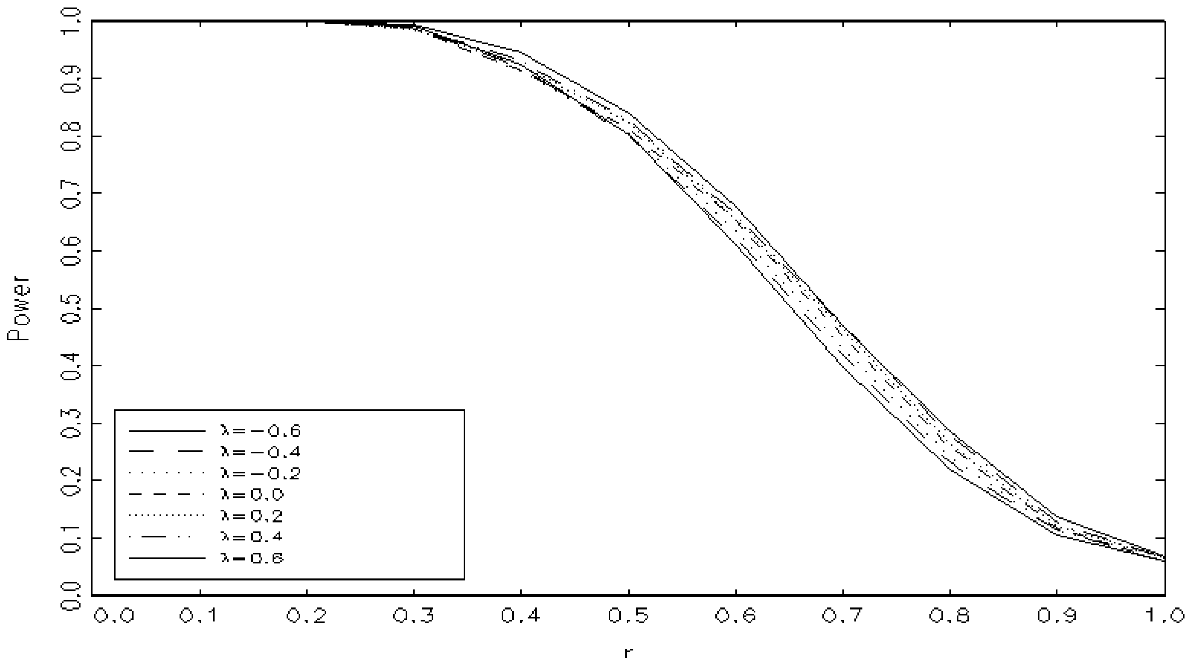


Figure 2.8 plots the power of the DLR conditional test for log-linearity with spatial lag dependence under the hypothesis H_0^{17}. According to the figure, the power of this test increases as r moves from 0 to 1. If the model closes to log-linearity, the power of this test is sensitive to spatial lag dependence. However, the spatial lag coefficient is not important to determine the power of the DLR conditional test as the model closes to linearity.

Figure 2.9 on the next page plots the power of the DLR conditional test for linearity conditional on spatial lag dependence under hypothesis H_0^{18}. The figure shows that the power of this test increases when r departs from 1 to 0. The power of the DLR conditional test under the hypothesis is insensitive to a change in spatial lag coefficients, but it is very sensitive to a change in the values of r.

Figure 2. 9: Power of DLR conditional test under hypothesis $H_0^{18}: r = 1 | unkown\ \lambda$



The Monte Carlo results show that the power of the DLR tests increases as functional forms and/or spatial lag dependence depart from their hypothesized values. The DLR test performs reasonably well and provides test powers very similar to the LM test proposed in the Baltagi and Li (2004a).

5. Conclusions

This essay developed the DLR tests for functional form and spatial lag dependence. Specifically, we developed the DLR joint, one-direction, and conditional tests for testing functional forms and spatial lag dependence. Results from an empirical example and Monte Carlo simulations show that the DLR and LM tests performed similarly in the empirical example. They had the same rejection and non-rejection decisions for all hypotheses. The DLR and LM tests also worked similarly in Monte Carlo simulations, and they had similar responses of test power associated with a change in functional form and spatial lag dependence. The DLR tests, however, only needed the first-order derivatives of the log-likelihood function, while the LM tests required the Hessian matrix (the second-order derivatives of the log-likelihood function). The DLR tests provide practitioners an easy-to-use alternative method for testing functional form and spatial lag dependence in empirical studies.

Essay 3: Effects of fertility on parental labor force participation and labor supply: Evidence from Vietnam

1. Introduction

Studies of labor supply have attracted much attention from economists and demographers in the last few decades. Many theoretical models have been developed to investigate the linkages of a household's behavior such as childbearing, the number of children, and child gender, to labor supply of the household members (e.g. DeTray, 1973; Ashenfelter and Heckman, 1974, Cogan, 1977; Rosenzweig and Wolpin, 1980, Salkerve, 1982; Robinson and Tomes, 1982, and Schultz, 1990). Other papers have considered empirical models that explore the relationship between fertility and the female labor supply for married women, unmarried women, and women in general (e.g. Carliner et al., 1980; Gregory, 1982; Leher, 1992; Chun and Oh, 2002; Rica and Ferrero, 2003; Cristia, 2006; and Cruces and Galiani, 2007). Few papers, however, have examined the effects of fertility on paternal labor force participation and labor supply (e.g. Pencavel, 1986, Carlin and Flood, 1997; Angrist and Evans, 1998; Lundberg and Rose, 2002; Lundberg, 2005; Choi et al., 2005; and Kim and Aassve, 2006).

Studies of children and parental labor supply in the economic literature can be categorized as to how the studies dealt with endogeneity of fertility. Some studies assumed that fertility was an exogenous variable, and they used Ordinary Least Squares (OLS) to estimate the effects of fertility on labor supply (Gronau, 1973; Heckman, 1974; Heckman and Willis, 1977; Carlin and Flood, 1997; and Cristia, 2006). Other studies found that fertility was an endogenous variable and suggested ways to remedy the endogeneity problem. For example, some research used simultaneous equation models, in which fertility and women's labor supply were two endogenous variables in the system (Cain and Dooley, 1976; Schultz, 1978; and Fleisher and Rhodes, 1979). Other research dealt with the endogeneity of fertility by using instrumental variables that helped to generate exogenous variations in fertility. Many instrumental variables have been used in the literature. For instance, many studies have suggested that a twin first birth is a factor that produces exogenous variations in fertility (Rosenzweig and Wolpin, 1980; Bronars and Grogger, 1994; Jacosen et al., 1999; Oyama, 2001; and Black et al., 2007). Bloom et al. (2007) used abortion legislation as an instrumental variable for fertility, while Kim and Aassve (2006) used the contraceptive choice of couples to generate the exogenous variations in fertility. Angrist and Evans (1998) used a mixed sibling-sex composition as the instrumental variable. Other studies used the gender of the first child as the instrumental variable (Choi and Oh, 2002 and Lundberg and Rose, 2002).

Almost all studies in the literature on motherhood and the labor market behavior of women have found that having children substantially reduced women's labor force participation rate and women's labor supply. Using data from the 1980 and 1990 Census Public Use Micro Sample, Angrist and Evans (1998) suggested that having a third child reduced the probability of working of married women by 17%, and caused their labor supply to fall by 8-9 weeks per year or 6-7 hours per week. When estimating the effects of fertility on the labor force participation of married women in Korea, Chun and Oh (2002) found that an additional child reduced female labor force participation probability by 27.5%. Using data adapted from the National Survey of Family Growth, Cristia (2006) found that having a first child younger than one-year old reduced the maternal probability of participating in the labor force by 26%. Bloom et al. (2007) suggested that the negative effect of fertility on female labor force participation ranges from 5% to 15%, depending on the age groups to which mothers belong. This inverse relationship between fertility and female labor force participation and the negative effect of fertility on the female labor supply is also found in many other studies (Gronau, 1973; Heckman, 1974; Heckman and Willis, 1977; Carliner et al., 1980; Rosenweig and Wolpin, 1980; Robinson and Tomes, 1982; Gregory, 1982; and Waldfogel, 1998).

Unlike the obviously negative effects of fertility on female labor force participation and labor supply, the average effect of fertility on paternal labor force participation and labor supply is ambiguous. Although the burden of childbearing mainly falls on women, an increase in family size reallocates time among household members. Two opposite effects of fertility on the paternal labor supply have been presented in the literature. Becker (1985) argued that an increase in family size would lead women to spend more time and energy on supplying child services because of the increase in the intensity of childcare, but men are likely to spend more energy and time in the labor market due to the higher return of labor on the labor market. These responses to an expansion of family size are called the specialization effect. Lundberg and Rose (1999) argued that more children increase value of parental time as an input of household production by supplying child services; thus, parents are likely to reduce their labor supply. This effect of fertility is called the home-intensive effect. According to both arguments, the probability of women joining the labor force and the time they spend in the labor market is reduced, while the male labor force participation and labor supply may increase or decrease, depending on which of the two effects dominates.

In the economics literature, some studies found that the presence of children had a weakly positive or no significant association with the probability of paternal work participation and labor supply (Pencavel, 1986; Bloomquist and Hannson, 1990; Carlin and Flood. 1997; and Angrist and Evans, 1998). Other papers found a positive relationship between fertility and paternal labor supply. Using data from the Panel Survey of Income Dynamics, Lundberg and Rose (2002) suggested that children significantly increased their fathers' annual work hours and hourly wage rate. Choi et al. (2005) found that a new child could increase the number of hours a father works, and the first son could cause an increase in the father's hours of more than 100 hours per year than the first daughter for German men could cause. Kim and Aassve (2006), using the data from Indonesia, found that fertility had positive effects on male labor supply in rural areas, but it was not statistically significant for men in urban areas.

Using data from the Vietnamese Household Living Standard Survey in 2004, this essay investigates the effects of fertility on parental labor force participation and labor supply in Vietnam. Specifically, I use instrumental variable (IV) probit models to estimate the effects of fertility on parental labor force participation and the IV models to estimate the effect of fertility on parental labor supply. This essay further aims to investigate the effects of fertility on the labor force participation and labor supply of both men and women, and how the effects of fertility on parental labor force participation and labor supply vary across urban and rural areas. Using the gender of the first child and whether the first two children have the same gender as two instrumental variables, this essay finds a negative effect of fertility on maternal labor force participation and labor supply and a positive effect on paternal labor force participation and labor supply for both rural and urban areas. Having an additional child reduces the probability of female participation in the labor force by 26.4% for the whole sample and by 29.3% for rural mothers. It also causes a decline of 0.35 work hours per day for the whole sample and 0.40 work hours for rural mothers per day. On the other hand, the results suggest that fertility has positive effects on male labor force participation and labor supply. Fathers increased their probability of going to work by 21.7% after a new child was born in their households, but that probability is dramatically different in urban and rural areas: 26.8% for urban men and 19.3% for rural men. Having one more child, fathers worked 0.32 hours more than before, while urban fathers increased their work hours per day by 0.47 hours and rural fathers by 0.30 hours per day. These results imply that having an additional child had specialization effects on the parental labor supply.

The remainder of this essay is organized as follows. Section 2 describes the methodology. Section 3 provides some background information about Vietnam and the data set. Section 4 presents the empirical results. Section 5 concludes the essay.

2. Methodology

This section describes the methodology used in this study. In particular, it presents empirical models for estimating the effects of fertility on paternal labor force participation and labor supply. Additionally, this section discusses how instrument variables are chosen and used to generate exogenous variations in fertility.

2.1. Empirical models

The objective of this research was to estimate the effects of fertility on the parental labor force participation and labor supply. Because of the possible endogeneity of fertility, this essay first tested whether endogeneity of fertility existed in the parental labor force participation and labor supply. Since fertility, parental labor force participation, and labor supply were simultaneously determined, the OLS approach was inappropriate, and its estimates are biased and inconsistent. This research used the IV probit and IV models to deal with the problems of endogeneity.

In the first stage, the demand for fertility specified in equation (3.1) below was used to estimate and predict number of children as a function of instrument variables and other exogenous variables. The equation is given as:

$$n = \phi_n + Z\varphi_n + W\alpha_n + M\beta_n + H\gamma_n + \varepsilon_n \tag{3.1}$$

where n is the number of children in the households; Z is a vector of the instrumental variables; W is a wage vector of both the mother's and father's earnings from labor market; M is a vector of parental characteristics such as age, age-squared, education levels, whether the parent is a household head, and whether the work was a non-farm job; H is a vector of other household and labor market characteristics such as location of the household, the non-labor household income, whether a grandparent was living in the household, and local annual employment; and ε_n is the error term. The first-stage model was also used to test for validity of instrumental variables and justify how instruments affect fertility.

The parental labor force participation was estimated using the following probit model:

$$P(y = 1 \mid \hat{n}, W, M, H) = \Phi(\phi_y + \varphi_y \hat{n} + W\alpha_y + M\beta_y + H\gamma_y + \varepsilon_y) \tag{3.2}$$

where Φ is a probability distribution function. y takes value of one or zero depending on whether the parent works or not. If the father or mother participated in the labor market, $y = 1$, and if not $y = 0$. $\hat{n}$ was the predicted number of children estimated from equation (3.1). ε_y was the error term. To estimate this model, we used the IV probit estimation technique focusing mainly on how fertility affects the decision to work for both mother and father.

Another objective of this research was to estimate the effects of fertility on the parental labor supply. To do that, the following linear model was estimated:

$$L = \phi_l + \varphi_l \hat{n} + W\alpha_l + M\beta_l + H\gamma_l + \varepsilon_l \qquad (3.3)$$

where the dependent variable L is labor supply of the father/mother. In this research, the parental labor supply was measured by work hours per day. A key independent variable was fertility. Using IV estimation procedure, we measured how fertility affects parental work hours per day.

2.2. Instrumental variables

The choice of appropriate instrumental variables is important because these can affect the reliability of estimates and inferences. Valid and strong instrumental variables must satisfy two conditions: an instrumental variable should be uncorrelated with the error term and it should be highly correlated with the right-hand-side endogenous regressor(s). In this research, that means that the instrumental variables have no correlation with factors that directly affect parental labor force participation and labor supply and that the instruments are correlated with fertility.

Two instrumental variables were used to generate exogenous variations in fertility in this research. The first instrumental variable was the gender of the first child. If gender of the first child is random, the first condition of instruments would be automatically satisfied (Chun and Oh, 2002). In Vietnam, son preference is strong for various economic, religious, and social reasons, and it has a significant positive effect on fertility, meaning that when a couple has a daughter, they are more likely to try again in hope of having a son. Thus, having a son negatively affected the decision to have a second child in a household (Haughton and Haughton, 1995 and Belanger, 2002). Normally, the gender of a child is a random variable, and it is uncorrelated with parental labor force participation and labor supply. One may claim that, however, son preference makes parents choose gender of their kids by aborting unborn girls. Fortunately, abortion was not gender-selective in Vietnam. Hieu et al. (1993) and Vach and Bishop (1996) found that more than 58% of abortions were due to poor sex education and failure of contraceptive. Sixty percent of abortions occurred during the first six weeks of pregnancy, while 92% occurred during the twelve weeks after the last menstrual period. Abortion during that time does not affect randomness of fertility because the gender of the unborn child cannot yet be known[3]. Furthermore, since the 1989 Law on Protection of the People's Health was issued, the Vietnamese government has not allowed hospitals and health care centers to reveal the gender of unborn children. This prohibition was reconfirmed in the 2003 Ordinance on Population. In addition, we found that the boy-to-girl ratio of the first child was 1.05 in our sample, which is close to the natural ratio. Thus, the gender of the first child is a valid instrumental variable.

The second instrumental variable is whether the first two children are of the same gender. Using data from the 1980 and 1990 Census Public Use Micro Sample, Angrist and Evans (1998) found that parents prefer a mixed sibling-sex composition, and parents who first had two girls or boys had a higher probability of having additional children. Using the 1988 and 1989 Demographic and Health Survey of Vietnam, Allman et al. (1991), found that couples with more than two children wanted to have at least one son in their family. Haughton and Haughton (1998) used a hazard model for the Vietnam Living Standard Survey 1992-1993 and found a strong son preference among women who had at least one child. They also found a mixed sibling-sex preference among those who had more than two sons in the family. Hollander (1996) showed that having one son reduced the likelihood of another birth by 21% to 30% among women with at least one child. Among women with more than two children,

[3] Before 18 weeks of last menstrual period, the gender of an aborted child is unknown.

the likelihood of another birth was 28% if they had a son, 34% if they had two sons and 11% if they had three sons. This evidence implies that siblings with mixed genders are desirable among Vietnamese families.

Following the work done by Angrist and Evans (1998), we generated the same gender dummy variable by using the gender variables of the first two children. The same gender variable is an interaction of the gender of the first and second children. It consists of two first boys and two first girls. Assuming the gender of the first child is s_1, and that of the second child is s_2, the same gender variable (sg) can be calculated by $sg = s_1 s_2 + (1 - s_1)(1 - s_2)$. If the first two children have the same gender, the same gender variable equals one ($sg = 1$); zero ($sg = 0$) otherwise.

We used an approach suggested by Stock et al. (2002) in which the first-stage F-statistics of instrumental variables must be larger than 10 for the instruments to be strong. The F-test statistics for significance of the two instrumental variables-the gender of first child and same gender of first two children-in the first stage varied from 44.75 to 243.79 for parental labor force participation and labor supply. The results implied that the gender of the first child and the same gender of the first two children are strong instruments.

The gender of the first child and the same gender of the first two children variables meet the two conditions required of a valid and strong instrument, and they can serve as instruments to generate exogenous variations in fertility.

3. Background and the data

Over the past two decades, fertility has been decreasing as the labor force participation rates of women in most developing and advanced countries has been increasing (Kim and Aassve, 2006 and Del Boca et al., 2005). This change implies the changing roles of women and changes in the time allocation among household members in both work activities and fertility behavior. We also observed this pattern in Vietnam. Table 3.1 summarized some important structural changes in the Vietnamese economy and society after the Doi Moi "renovation" process launched in 1986.

Table 3. 1: Total fertility rate, labor force participation, GDP per capita, and urbanization in Vietnam

Year	Fertility	Labor force participation (%)	Urban Population (%)	GDP per capita (USD)
1986	4.2	46.8	19.8	202.8
1988	3.8	47.1	20.0	210.4
1990	3.6	47.3	20.3	226.9
1992	3.2	47.7	21.0	250.6
1994	2.6	48.1	21.8	283.7
1996	2.5	48.7	22.6	327.8
1998	2.4	49.5	23.5	364.1
2000	2.0	50.6	24.3	397.0
2002	1.9	51.5	25.3	443.7
2004	1.8	52.5	26.2	502.0
2006	1.8	52.8	27.1	587.4

Source: The World Bank (2006) for the year 1986 to 2004 and General Statistic Office (2007) for the year 2006.

For the last two decades, the fertility rates of Vietnamese women fell by 57%, while the labor force participation rates for the whole population steadily increased to 52.8% in 2006. A decline in fertility also accompanied an increase in income. During the period from 1986 to 2006, while fertility dramatically decreased, GDP per capita increased 2.9 times to 587.4 USD per capita. This pattern is consistent with microeconomic predictions: higher income leads to a reduction in fertility and the inverse relationship of fertility and labor force participation (Becker and Lewis, 1973 and Willis, 1973). Another obvious phenomenon is the rapid urbanization in Vietnam. During the past two decades, the proportion of population living in urban areas increased 1.4 times. This process has led to a change in the structure of the work force. Urbanization has reduced the number of households working on farms where work is labor-intensive, and it has possibly affected the way time is allocated among household members. Noticeably urbanization, however, has not occurred equally across regions. It has happened mostly in areas near urban centers. As a result, there is variation in regions, especially between urbanized and rural areas in terms of labor market behavior and fertility decisions. Thus, it is important to understand the effects of fertility and labor market behavior of households in both urban and rural areas.

The data used in this paper came from the Vietnamese Household Living Standard Survey 2004, which was conducted by the Vietnamese General Statistical Office (GSO) with technical support from The World Bank. The survey sample was randomly selected to represent the whole country, taking into account urban and rural structures, geographical conditions, regional issues, ethnic differences, and provincial representation. It was carried out in 300 rural hamlets and 80 urban blocks in 3,063 communities around the country. The survey sample consisted of 9,192 households with 42,839 individuals. The survey collected information about the following: household information, education, health, employment, migration, housing, fertility and family planning, incomes, expenditures, borrowing, lending, and savings.

Only households with at least one child under 18 years old and households with a mother and father younger than 60 and 65 years of age, respectively, at the time of the interview were included in this research. There are 3,985 households in the sample used for this research. Table 3.2 on the next page provides a summary of the descriptive statistics used in this research. The dependent variables were the working status and hours worked per day by each of the parents.

Table 3. 2: Descriptive and data statistics

Variables	Mean	Std. Dev.	Min.	Max.
Mothers' characteristics				
Work status	0.954	0.210	0	1
Work hours per day	6.468	1.917	0	16
Age	35.994	6.660	19	60
Hourly wages	2.138	7.184	0	369.6
Head of household	0.087	0.282	0	1
No education	0.041	0.199	0	1
Primary school	0.282	0.450	0	1
Middle school	0.317	0.466	0	1
High school	0.250	0.433	0	1
College and higher	0.108	0.311	0	1
Work on farms	0.784	0.412	0	1
Fathers' characteristics				
Work status	0.978	0.111	0	1
Work hours per day	7.711	1.786	1	17
Age	38.637	6.931	20	65
Hourly wages	2.978	5.599	0	165.4
No education	0.035	0.185	0	1
Primary school	0.274	0.446	0	1
Middle school	0.330	0.470	0	1
High school	0.240	0.427	0	1
College and higher	0.121	0.326	0	1
Work on farms	0.630	0.483	0	1
Children statistics				
Number of children	2.552	1.239	1	15
Two first boys or girls	0.412	0.492	0	1
First boy	0.512	0.499	0	1
Household characteristics				
Non-labor household income	1.404	1.845	-2.9	40.3
Grandparents living in household	0.113	0.317	0	1
Grandparents' age	71.444	10.349	39	98
Urban	0.213	0.410	0	1
Local economic conditions				
Annual new jobs	28.269	32.218	3.5	220

According to Table 3.2, 97.8% of fathers worked in the interview year, and on average, they worked 7.7 hours per day. For women, 95.4% of mothers worked in the interview year, and on average, they worked 6.5 hours per day. The explanatory variables are the number of children in the household, parents' wages, parental characteristics, characteristics of the households, and local labor market characteristics. Table 3.2 shows that each household had an average of 2.5 children; the wages paid to men were higher than wages paid to women; that on average, men received 3.0 thousands Vietnamese Dong (VND) per hour, while women received 2.1 thousand per hour; and 63% of working fathers and 78.4% of working wives worked on farms.

Table 3.2 also shows that a higher percentage of men than women had a college education and that lower percentage had no education. Of the household heads, 91.3% were male. The households' characteristics were represented by the location of the household, non-labor household income, and presence of a grandparent in the household. According to Table 3.2, 21.3% of households were located in urban areas, and the average non-labor household income was 1.4 million VND per year. The data also showed that 11.3% of the sampled households had grandparents living in them. Local labor market characteristics were represented by annual new jobs. On average each province created 28.3 thousands new jobs annually.

Two instrumental variables were the gender of the first child and the same gender of the first two children in the households. The gender of the first child equaled 1 if the household had a male first child and 0 otherwise. The same gender of the first two children equaled 1 if the first two children had the same gender, and 0 otherwise. According to Table 3.2, 51.2% of sampled households had a male first child and 41.2% of households had two first children whose gender was the same.

4. Empirical results

This section provides the empirical results of our investigation. We first tested for the endogeneity of fertility using the Hausman test. Then we estimated the probit and IV probit models to measure the effects of fertility on the parental labor force participation for the whole, urban, and rural samples. Third, we estimated the effects of fertility on the parental labor supply using OLS and IV models. We also estimated models for samples of households with at least two children and used their results for comparison and making inferences.

4.1. Hausman tests

One of the concerns when estimating the effects of fertility on the parental labor force participation and labor supply is the endogeneity of fertility. Because fertility and parental working status are simultaneously determined, and fertility and parental labor supply are simultaneously determined, regular probit and OLS regression techniques do not yield unbiased and consistent estimates. To test for the possible endogeneity of fertility, this study used the Hausman test. The null hypothesis is that fertility is exogenous. If the null hypothesis is rejected, fertility is an endogenous variable. In this case, the probit and OLS techniques were not appropriate techniques to estimate effects of fertility on parental labor force participation and labor supply. Instead, the IV Probit and 2SLS/IV methods were used.

This research found that fertility is endogenous in the parental labor force participation and in the female labor supply at the 5% level. The Hausman test statistics for testing endogeneity of fertility in female and male labor force participation were 66.91 with zero p-value and 43.42 with p-value of 0.0028, respectively. Thus, the hypothesis of an exogenous fertility in parental labor force participation was rejected. Similarly, the exogenous fertility in female labor supply was also rejected, and the Hausman test statistic of 39.72 with p-value of

0.0116 suggested the endogeneity of fertility in the female labor supply. However, the Hausman test statistic does not suggest rejecting exogenous fertility in the male labor supply.

4.2. Fertility and parental labor force participation

The model presented in equation (3.2) suggests that parental labor force participation depends on fertility, parental wages, the characteristics of parents, and the characteristics of household and local labor market. Using the gender of first child and same gender of first two children in the households as two instrumental variables, we estimated the effects of fertility on labor force participation. Two separate models, one for men and one for women, were estimated. The dependent variables were parental labor force participation. The dependent variable equaled 1 if the father (for male labor force participation) or mother (for female labor force participation) worked in the interviewed year. Tables 3.3 and 3.4 on the next two pages report the marginal effects in the parental labor force participation models, while the coefficients are presented in appendices 1 and 2.

Table 3. 3: Marginal effects of fertility on maternal labor force participation

	Pool		Urban		Rural	
	Probit	IV-Probit	Probit	IV-Probit	Probit	IV-Probit
Fertility	-0.0417***	-0.2637***	0.0475	0.0820	-0.0490***	-0.2932***
Mothers' characteristics						
Age	-0.0209	0.0290	-0.0466	-0.0519	-0.0193	0.0357
Age squares	0.0002	-0.0004	0.0004	0.0005	0.0002	-0.0005
Household head	0.0624	-0.0346	-0.0212	-0.0148	0.0967	-0.0442
Primary school	0.4940***	0.2350***	0.2109*	0.2161*	0.5671***	0.1693***
Middle school	0.5037***	0.2918***	0.3591***	0.3671***	0.6843***	0.2247***
High school	0.5323***	0.2977***	0.3975***	0.4051***	0.5821***	0.2209***
College and higher	0.5331***	0.3283***	0.4962***	0.5037***	0.5227***	0.2277***
On farm	0.1029***	0.1551***	0.1594***	0.1497**	0.1238***	0.1563***
Fathers' characteristics						
Age	-0.0859***	-0.0108	0.0633	0.0572	-0.1257***	-0.0087
Age squares	-0.0016	-0.0017	-0.0046*	-0.0047*	0.0026	0.0011
Hourly wage	0.0011***	0.0002	-0.0006	-0.0006	0.0015***	0.0001
Primary school	0.100*	0.0386	0.3758***	0.3816***	0.0313	-0.0140
Middle school	0.2226***	0.0998*	0.4060***	0.4138***	0.1594**	0.0264
High school	0.1414**	0.0535	0.3403**	0.3486***	0.0749	-0.0040
College and higher	0.0031	-0.0091	0.2069	0.2159	-0.0649	-0.0307
On farm	-0.1037***	-0.0382	-0.0479	-0.0480	-0.0881**	-0.0018
Household's characteristics						
Non-labor Household income	-0.0235***	-0.0079	-0.0166**	-0.0181**	-0.0462***	-0.0195***
Grandparent	0.9470**	0.7480*	0.5183	0.5275	0.9768**	0.6772*
Grandparent's age	-0.1569***	-0.0786**	-0.1022**	-0.1045**	-0.1579***	-0.0469*
Grp. age squares	0.0012***	0.0006**	0.0010***	0.0010***	0.0012***	0.0003*
Urban	-0.0270	-0.1182***				
Local labor market situation						
Annual new jobs	-0.0004	-0.0006**	-0.0010**	-0.0010**	0.0009	-0.0005
Observations	3935	3935	832	832	3103	3103

Significant levels of 0.1 are denoted by *; 0.05 by **; and 0.01 by ***.

Table 3. 4: Marginal effects on paternal labor force participation

	Pool		Urban		Rural	
	Probit	IV-Probit	Probit	IV-Probit	Probit	IV-Probit
Fertility	0.0192***	0.2165***	0.0119	0.2678***	0.0188***	0.1935***
Mothers' characteristics						
Age	0.0271***	0.0172	0.0345***	0.0190	0.0238***	0.0245
Age squares	-0.0003***	-0.0002	-0.0004***	-0.0001	-0.0003***	-0.0002
Hourly wage	0.0023*	0.0056**	0.0025	0.0037	0.0013	0.0050
Household head	-0.1207***	-0.0761**	-0.1357***	-0.0895*	-0.0904***	-0.0668
Primary school	0.0019	0.0571	0.0344	0.2729**	-0.0282	-0.0310
Middle school	-0.0400	0.0161	-0.0027	0.2058*	-0.0574*	-0.0513
High school	0.0035	0.0395	0.0484*	0.3487***	-0.0399	-0.0744
College and higher	0.0164	0.0344	0.0341*	0.2735***	-0.0065	-0.0407
On farm	0.1083***	0.0675**	0.1017***	0.0953*	0.0884***	0.0742**
Fathers' characteristics						
Age	0.0325***	0.0330**	0.0489***	0.0643	0.0251***	0.0276*
Age squares	-0.0004***	-0.0004***	-0.0006***	-0.0008*	-0.0003***	-0.0004*
Primary school	0.0259***	0.1208***	-0.0403	0.0660	0.0268***	0.1094***
Middle school	0.0261**	0.1254***	-0.0311	0.1118	0.0231**	0.0949***
High school	0.0269***	0.1315***	-0.0077	0.1456	0.0266***	0.1083***
College and higher	0.0310***	0.1668***	0.0254	0.2652***	0.0287***	0.1183***
On farm	0.0301***	0.0327	0.0287**	0.0681	0.0267***	0.0353
Household's characteristics						
Non-labor Household income	0.0108***	0.0183**	0.0078*	0.0014	0.0145**	0.0304**
Grandparent	-0.0572	0.0642	0.0255	0.2673	-0.0480	0.0317
Grandparent's age	0.0053*	-0.0120*	-0.0008	-0.0324*	0.0048	-0.0060
Grp. age squares	-0.0001**	0.0001	0.0000	0.0003*	-0.0001**	0.0000
Urban	-0.1108***	-0.0779***				
Local labor market situation						
Annual new jobs	-0.0003***	-0.0003	-0.0002**	-0.0004	-0.0003**	-0.0002
Observations	3800	3800	771	771	3029	3029

Significant levels of 0.1 are denoted by *; 0.05 by **; and 0.01 by ***.

As shown in Tables 3.3 and 3.4, the marginal effects of fertility on parental labor force participation from IV probit models are larger in absolute values than probit estimates. According to the IV probit results, the marginal effect of fertility on female working participation was negative and statistically significant at the 5% level for the whole sample. This result means that having one additional child decreased the probability of female labor force participation by 26.4%, all other factors being equal. The sign and magnitude of the effects on maternal labor force participation of these results were similar to the estimates of Chun and Oh (2002)[4]. When estimates were made for urban and rural women separately, fertility had no significant impact on the probability of urban mothers' working decisions, while one additional child reduced the probability of labor force participation by 29.3% for rural women. In contrast, an additional child had a positive and statistically significant effect on paternal labor force participation at the 5% level. For the whole sample, an additional child increased the probability of paternal labor force participation by 21.7%. When estimates were made for urban and rural men separately, an additional child increased the probability of urban fathers working by 26.8% while it increased the likelihood of rural fathers by 19.3%.

Rural parents worked mostly on farms, and their schedules were much more flexible. In addition, the productivity of rural labor and the returns to rural labor were low, while urban parents had non-farm jobs with higher pay. So having a new child increased the value of the mother's time as an input of home production. Rural women were most likely to stay home because their return on their labor was low and their jobs were flexible (no constraints and no

[4] Fertility reduced the female probability of participating in the labor force by 27.5% for married Korean women. Moreover, IV probit estimates from a sample of households with at least two children showed that having an additional child decreased the female probability of work participation by 17.3% and had no effects of fertility on male labor force participation. These results are close to the effects of fertility on the female labor force participation of U.S. women (decreased by 17%) and no statistically significant effect on the male probability of work participation in Angrist and Evans (1998) for the U.S.

contracts); meanwhile, urban men were more willing to work because of the higher returns in urban jobs. The results imply that value of male time on the local labor market is relatively higher than that of women, and women have a higher productivity in household production by providing child services. This is consistent with the findings of Becker (1985) who found that men would concentrate more on wage-paid activities, while women would concentrate on home production after a child was born.

As expected, education had a positive and statistically significant effect on the labor force participation of parents. The results show that higher-educated parents were more likely to participate in the labor force than lower-educated ones were. One other interesting result was that spousal characteristics did not have statistically significant effects on probability of participating labor force for Vietnamese parents when a new child was born. Tables 3.3 and 3.4 also showed that presence of grandparents in households had positive but insignificant effects on parental labor force participation at 5%. Local labor market conditions, measured by the number of annual new jobs, had little effect on parental labor force participation. Table 3.3 showed that the number of annual new jobs negatively and significantly affected the work decision of women for the whole and urban sample but not for rural sample. The magnitude of that effect, however, was relatively small. Meanwhile, Table 3.4 suggested that the number of new jobs created annually did not have statistically significant effect on the decision by men to work regardless of location at 5%.

This study has shown that fertility has a negative effect on maternal labor force participation and a positive effect on paternal labor force participation in Vietnam. This result is consistent with Becker’s (1985) findings that there is the specialization effect of having a new child in households.

4.3 Fertility and parental labor supply

The effects of fertility on parental labor supply were measured with the estimating models given by equation (3.3). The dependent variable was female/male work hours per day. As discussed in the earlier sections, fertility and parental labor supply were simultaneously determined, or in other words, there was endogeneity of fertility in the parental labor supply equation. Thus, the OLS coefficients are likely to be biased if used to measure the effects of fertility on the parental labor supply. In this situation, IV methods were a better method to give unbiased estimates. Table 3.5 on the next page shows the effects of fertility on female labor supply, while Table 3.6 on the following page presents the effects of fertility on the male labor supply in Vietnam. Each table provides IV coefficients in comparison to OLS estimates and their t-values in parentheses for the whole, urban, and rural samples.

According to Tables 3.5 and 3.6, IV estimates are different from OLS estimates. This occurs because there are biases in the OLS estimation results. The OLS estimates are biased because fertility picks up effects of unobserved variables such as the traditional values of Vietnamese families: extended family system (Altonji et al., 1989; Groot and Van den Brink, 1992; and Lacroix et al. 1997) and the social values of the traditional family norm (Vach and Bishop, 1996). Other unobserved variables could be the education quality of the parents, changes in the patterns of family formation, changes in the values and attitudes of women towards a less traditional role of women within the family and society (Hakim, 2003 and Gilbert, 2005), and the morality and poor healthcare system in Vietnam (Phai, 1999 and Huan, 1997). In terms of econometrics, the models omit some relevant variables. The omission of the unobserved relevant variables, thus, leads to the biased estimates of the parameters. In this case, the sign of the bias is a product of how the extended family system affects parental labor supply and how the extended family affects fertility.

Table 3. 5: Fertility and maternal work hours per day

	Pool		Urban		Rural	
	OLS	IV	OLS	IV	OLS	IV
Fertility	0.041	-0.353***	-0.021	-0.225	0.041	-0.399***
	(1.522)	(-3.288)	(-0.264)	(-0.890)	(1.480)	(-3.457)
Mothers' characteristics						
Age	-0.112**	-0.045	-0.071	-0.043	-0.123**	-0.047
	(-2.076)	(-0.780)	(-0.465)	(-0.272)	(-2.198)	(-0.766)
Age squares	0.001*	0.000	0.001	0.000	0.001*	0.000
	(1.694)	(0.431)	(0.382)	(0.175)	(1.929)	(0.556)
Hourly wage	-0.034***	-0.035***	-0.020***	-0.020***	-0.105***	-0.110***
	(-8.206)	(-8.222)	(-4.048)	(-4.047)	(-10.791)	(-10.750)
Household head	0.376***	0.271**	0.331*	0.294*	0.383***	0.245*
	(3.543)	(2.409)	(1.940)	(1.667)	(2.760)	(1.646)
Primary school	-1.351***	-1.447***	-1.920***	-1.985***	-1.242***	-1.345***
	(-7.394)	(-7.632)	(-3.400)	(-3.467)	(-6.646)	(-6.846)
Middle school	-1.093***	-1.245***	-1.623***	-1.710***	-0.981***	-1.157***
	(-5.937)	(-6.432)	(-2.903)	(-2.996)	(-5.183)	(-5.722)
High school	-1.743***	-1.796***	-1.773***	-1.860***	-1.787***	-1.820***
	(-9.446)	(-9.436)	(-3.176)	(-3.263)	(-9.390)	(-9.171)
College and higher	-0.762***	-0.762***	-1.842***	-1.914***	-0.489**	-0.468**
	(-3.886)	(-3.778)	(-3.138)	(-3.213)	(-2.414)	(-2.215)
On farm	-1.107***	-0.976***	-0.741***	-0.694***	-1.461***	-1.321***
	(-13.099)	(-10.446)	(-4.410)	(-3.906)	(-14.749)	(-12.103)
Fathers' characteristics						
Age	-0.067	0.000	0.058	0.096	-0.073	0.004
	(-1.284)	(0.005)	(0.356)	(0.566)	(-1.386)	(0.069)
Age squares	0.001	0.000	-0.001	-0.001	0.001	0.000
	(1.269)	(0.059)	(-0.274)	(-0.471)	(1.246)	(-0.119)
Hourly wage	-0.013**	-0.013**	-0.005	-0.004	0.002	0.003
	(-2.226)	(-2.052)	(-0.525)	(-0.403)	(0.211)	(0.328)
Primary school	-0.243	-0.317*	-1.250**	-1.291**	-0.112	-0.194
	(-1.389)	(-1.751)	(-2.373)	(-2.431)	(-0.622)	(-1.033)
Middle school	-0.106	-0.206	-1.234**	-1.300**	0.058	-0.045
	(-0.600)	(-1.123)	(-2.364)	(-2.453)	(0.316)	(-0.236)
High school	-0.065	-0.163	-0.793	-0.858	-0.030	-0.137
	(-0.367)	(-0.881)	(-1.510)	(-1.609)	(-0.161)	(-0.709)
College and higher	-0.070	-0.100	-1.178**	-1.238**	0.084	0.078
	(-0.373)	(-0.519)	(-2.189)	(-2.271)	(0.425)	(0.379)
On farm	-0.186***	-0.144**	-0.256	-0.245	-0.218***	-0.176**
	(-2.641)	(-1.973)	(-1.513)	(-1.436)	(-2.860)	(-2.200)
Households' characteristics						
Non-labor H income	0.037**	0.048***	0.074***	0.083***	-0.019	-0.014
	(2.203)	(2.749)	(2.755)	(2.862)	(-0.834)	(-0.579)
Grandparents	0.068	-0.150	0.256	0.176	-0.078	-0.329
	(0.203)	(-0.429)	(0.292)	(0.198)	(-0.222)	(-0.882)
Grandparents' age	-0.025*	0.013	-0.035	-0.018	-0.018	0.026
	(-1.765)	(0.743)	(-0.906)	(-0.415)	(-1.224)	(1.343)
Gra. Age squares	0.000	0.000	0.000	0.000	0.000	0.000
	(1.335)	(-0.896)	(0.535)	(0.138)	(0.956)	(-1.375)
Urban	1.052***	0.938***				
	(13.301)	(10.803)				
Local labor market situation						
Annual new jobs	0.001	0.000	0.002	0.002	-0.001	-0.002
	(0.926)	(0.347)	(1.325)	(1.225)	(-0.587)	(-1.361)
Observations	3754	3754	756	756	2998	2998

Numbers in parentheses are t-values. Significant levels of 0.1 are denoted by *; 0.05 by **; and 0.01 by ***.

Table 3. 6: Fertility and paternal work hours per day

	Pool		Urban		Rural	
	OLS	IV	OLS	IV	OLS	IV
Fertility	0.074***	0.317***	0.014	0.467**	0.076***	0.303***
	(3.042)	(3.282)	(0.190)	(2.008)	(2.964)	(2.908)
Mothers' characteristics						
Age	-0.035	-0.076	-0.008	-0.072	-0.048	-0.087
	(-0.707)	(-1.449)	(-0.060)	(-0.492)	(-0.925)	(-1.574)
Age squares	0.000	0.001	0.000	0.001	0.000	0.001
	(0.131)	(0.904)	(-0.053)	(0.421)	(0.349)	(1.018)
Hourly wage	0.007*	0.008**	0.005	0.006	0.018**	0.020**
	(1.838)	(1.990)	(1.237)	(1.247)	(1.965)	(2.192)
Household head	-0.049	0.016	-0.060	0.021	-0.087	-0.016
	(-0.501)	(0.158)	(-0.393)	(0.127)	(-0.678)	(-0.119)
Primary school	-0.499***	-0.440***	-0.498	-0.355	-0.472***	-0.419**
	(-2.992)	(-2.580)	(-0.979)	(-0.673)	(-2.718)	(-2.359)
Middle school	-0.496***	-0.402**	-0.052	0.142	-0.565***	-0.474***
	(-2.948)	(-2.310)	(-0.103)	(0.269)	(-3.214)	(-2.596)
High school	-0.476***	-0.444***	-0.069	0.125	-0.580***	-0.563***
	(-2.825)	(-2.592)	(-0.138)	(0.238)	(-3.281)	(-3.141)
College and higher	-0.050	-0.050	-0.366	-0.206	0.078	0.067
	(-0.276)	(-0.273)	(-0.692)	(-0.374)	(0.416)	(0.353)
On farm	0.108	0.028	-0.375**	-0.480***	0.342***	0.270***
	(1.399)	(0.329)	(-2.479)	(-2.933)	(3.719)	(2.738)
Fathers' characteristics						
Age	0.187***	0.146***	0.139	0.055	0.211***	0.171***
	(3.946)	(2.884)	(0.953)	(0.352)	(4.294)	(3.236)
Age squares	-0.002***	-0.002***	-0.002	-0.001	-0.002***	-0.002***
	(-3.783)	(-2.802)	(-0.895)	(-0.345)	(-4.101)	(-3.121)
Hourly wage	-0.047***	-0.047***	-0.037***	-0.040***	-0.050***	-0.050***
	(-8.597)	(-8.558)	(-4.178)	(-4.307)	(-6.568)	(-6.552)
Primary school	-2.111***	-2.065***	-2.734***	-2.641***	-2.043***	-2.001***
	(-13.214)	(-12.687)	(-5.764)	(-5.398)	(-12.274)	(-11.788)
Middle school	-2.168***	-2.106***	-2.792***	-2.644***	-2.094***	-2.041***
	(-13.454)	(-12.768)	(-5.937)	(-5.417)	(-12.406)	(-11.823)
High school	-2.471***	-2.411***	-2.765***	-2.622***	-2.523***	-2.468***
	(-15.189)	(-14.486)	(-5.843)	(-5.339)	(-14.788)	(-14.132)
College and higher	-1.794***	-1.775***	-2.852***	-2.718***	-1.501***	-1.498***
	(-10.453)	(-10.201)	(-5.883)	(-5.414)	(-8.220)	(-8.096)
On farm	-1.055***	-1.080***	-0.836***	-0.861***	-1.114***	-1.136***
	(-16.434)	(-16.427)	(-5.489)	(-5.486)	(-15.760)	(-15.717)
Households' characteristics						
Non-labor H income	0.091***	0.084***	0.114***	0.094***	0.067***	0.065***
	(5.918)	(5.322)	(4.736)	(3.513)	(3.202)	(3.031)
Grandparents	0.416	0.550*	1.912**	2.092**	0.007	0.137
	(1.354)	(1.745)	(2.418)	(2.560)	(0.022)	(0.406)
Grandparents' age	0.014	-0.010	-0.034	-0.071*	0.029**	0.006
	(1.049)	(-0.623)	(-0.965)	(-1.774)	(2.100)	(0.370)
Gra. Age squares	0.000	0.000	0.000	0.001	0.000	0.000
	(-0.688)	(0.792)	(0.878)	(1.620)	(-1.619)	(-0.114)
Urban	0.451***	0.521***				
	(6.237)	(6.676)				
Local labor market situation						
Annual new jobs	0.002**	0.002**	0.002**	0.003**	0.001	0.002
	(2.116)	(2.440)	(1.965)	(2.113)	(0.839)	(1.273)
Observations	3754	3754	756	756	2998	2998

Numbers in parentheses are t-values. Significant levels of 0.1 are denoted by *; 0.05 by **; and 0.01 by ***.

The OLS and IV estimates of fertility shown in Table 3.5 suggested that there was a positive bias in the OLS estimate of fertility on female labor supply. This positive bias was caused by omission of relevant variables or unobserved variables in the sample. To provide an appropriate explanation of the bias, I dug in the literature, data sample, and Vietnam context and found that the extended family system in Vietnam was one of the possible reasons that created the bias in estimated results. The extended family system was one of the reasons that normally increased the demand for children. This system reduced household's costs of rearing children because women could receive assistance from other family members such as grandparents, other children, and their relatives. Therefore, women are likely to have more children or higher fertility. On the other hand, the extended family also helped to reduce the housewife's work since women could receive help from other family members like grandparents, other children, and relatives, so women are more likely to supply more time in the labor market (Ketkar, 1979). The extended family system, thus, increases the female labor supply. Therefore, the extended family system in Vietnam is one of the reasons creating the positive bias of OLS estimates on female labor supply.

According to Table 3.6 on the preceding page, OLS and IV estimates imply a negative bias in the OLS estimate of fertility on male labor supply. This bias occurs because fertility picks up a negative bias of unobserved variables on male labor supply, for example an unobserved variable of extended family system in Vietnam. This family system and its interactions played an important role in supplying labor of its members (Groot and Van den Brink, 1992). Using theoretical models and empirical analysis, Lacroix et al. (1997) found that the extended family had a statistically significant and negative effect on paternal labor supply and had no significant effects on labor supply of maternal and other household members. In the context of Vietnam, a survey of the General Statistical Office of Vietnam

(2000) reported that a man living in the extended family worked 0.4 hours per day less than but 11.8 days per year more than a man living in a nuclear family did. A possible reason was that living in the extended family, men can pool their labor in forms such as rotating arrangements and labor gangs. Labor polling allows workers, especially farmers, to seek assistance from their extended family members, and so they can reduce their labor supply measured by work hours per day. In this sense, the extended family system can be a possible reason to create a negative bias in OLS estimates of fertility on male work hours per day.

Table 3.5 on page 75 showed that the IV coefficient of fertility was -0.353 and its t-value of -3.288. This means that an additional child caused a decline in the number hours worked by a woman by 0.35 hours per day. Meanwhile, Table 3.6 on page 76 showed that an additional child increased male labor supply by 0.317 hours per day. The effects of fertility on parental labor supply, however, differed from rural to urban areas. Rural women decreased their work hours per day by 0.4 hours, but urban women were not significantly affected. Urban men worked 0.47 hours more per day in response to an increase in family size, while rural men increased their labor supply by 0.30 hours per day[5]. The estimated results affirmed the specialization-effect argument of Becker (1985) that women are likely to work less and focus on providing child service, while men are likely to focus on labor market activities to respond to an increase in family size.

Educated parents worked less than non-educated parents did, but the results did not confirm that higher educated parents work less. Another consistent result is that the characteristics of the spouses had insignificant effects on their partners' labor supply. For example, the

[5] I also estimated the effects of fertility on the parental labor supply for families having at least two children. The IV results confirmed that there was a negative effect of fertility on female labor supply and a positive effect of fertility on male labor supply in Vietnam. The effect of fertility on parental labor supply, however, was much larger than the effect in the household with at least one child. The addition of a child decreased female labor supply by 1 hour per day and increased male labor supply by 0.89 hours per day.

mother's hourly wage, age, age-squared, and education level did not significantly affect paternal labor supply, while father's characteristics did not have significant effects on maternal work hours per day as well. Tables 3.5 and 3.6 also suggested that grandparents did not significantly help to increase parental labor supply, but they helped to increase work hours of urban men at the 10% level[6]. The location of the households had significant effects on parental labor supply, and on the average, urban parents tended to work longer hours than rural parents did[7]. Meanwhile, the parental labor supply responded differently to local labor market situations. If the number of new jobs created annually was higher, men tended to work longer (extra hours or multiple jobs), but women did not change their work hours.

An interesting result is a positive relationship between household non-labor income and parental labor supply. According to the regression results, parental labor supplies increase by 0.08 hours per day for men and 0.05 hours per day for women when non-labor household income increases by one million of Vietnamese Dong. To give an appropriate reason, I examined the literature and found the response of parental labor supplies strictly depends on how the household uses up household non-labor income. If the household uses this amount of income primarily for consumption, the parental labor supply reduces (Dinerman, 1982 and Airola, 2008). On the other hand, if the household uses its non-labor household income mostly for investment or for promoting household farms or small household firms, the parental labor supplies increase including self-employment (Woodruff and Zenteno, 2001 and Funkhouser, 1992). In the Vietnam context, most parents are characterized by low-income,

---

[6] We found that sampled grandparents had the average age of 71.4 years (presented in Table 3.2). At this age, it is hard for them to provide assistance and they even need to be provided with grandparent care.

[7] These results are true for a general case. In the crop seasons, however, the rural people may have longer working days than urban people may do.

but not hungry[8], working on-farm, or running small household businesses. Those parents are not likely to use non-labor income up for consumption, but they use it for investment to promote their household businesses or expanding/cultivating their farms. The non-labor household income plays an important role in encouraging non-farm household businesses in Vietnam (Tran et al. 2004) and eventually promoting parental labor supplies.

One striking finding in this study is that there was a negative effect of wage on the labor supply. According to Table 3.5, female hourly wages negatively and significantly affected the female labor supply in all samples: the whole, urban, and rural samples. The estimated coefficients of female hourly wage were negative for both the OLS and IV approaches. The similar results were found for the male hourly wage and labor supply presented in Table 3.6. This negative relationship between wage and labor supply for the Vietnamese is consistent with the findings of many studies for advanced countries and most of the developing countries (Sharif, 2003). Vietnam is a poor country with annual GDP per capita of 587.4 USD in 2006, almost all of the people are the working poor, and 73% of the population lives in rural areas and does farm work. Most Vietnamese worked unusually long hours in physically demanding jobs[9], but they still had difficulties meeting basic needs. Based on these facts, this study strongly suggested that the explanations of negative relationship between wage and labor supply come from economic suffering rather than a perverse mentality of the working poor in Vietnam.

This study has found a negative effect of fertility on maternal labor supply and a positive effect of fertility on paternal work hours per day. These results imply that having one

---

[8] Low-income parents do not mean they are hungry. According to the General Statistic Office (2007), the percentage of hungry in 2004 was 3.7%, while the poor were 18.4%.

[9] According to the sample data, 12.2% of working mothers and 16.3% of working sampled fathers worked more than 8 hours a day. Women received an average hourly wage of 2.14 thousand Vietnamese Dong (VND) (approximately 15 U.S. cents), while men received 2.98 thousand VND (about 20 U.S. cents) per hour.

more child in households creates a specialization effect. In other words, women are likely to work less and focus on home production by supplying child service, while men tend to concentrate on labor market activities and work more when a new child is born.

5. Conclusions

This essay estimated the effects of fertility on the parental labor force participation and labor supply in Vietnam. It used the gender of the first child and the same gender of the first two children in households as two instrumental variables in IV-probit and IV models. The empirical results shed light on how fertility affects the parental labor force participation and labor supply.

The results from the IV-Probit and IV models showed that Vietnamese couples responded differently when there was a new birth. Although the Probit and OLS techniques underestimated effects of fertility on the parental labor force participation and labor supply, fertility had negative effects on the female labor force participation and labor supply and positive effects on the paternal labor force participation and labor supply. A new child decreased the probability of mothers participating in the labor force by 26.4% and reduced maternal work hours per day by 0.35 hours. The estimated results also showed that the effects were more pronounced for rural and less well educated women. An equally important finding is that fertility had positive effects on the paternal labor force participation and labor supply. The positive effects can be found in the whole sample and the two sub-samples: urban and rural areas. Our results showed that the probability that Vietnamese men participate in the labor force increased by 21.6% and by 0.32 hours per day in response to an increase in family size. These effects appeared to be more pronounced for less well educated and urban parents.

The specialization effect of having a new child existed among Vietnamese couples. This result implies that Vietnamese men are more likely to concentrate on labor market activities rather than household production because their productivity in household production is relatively low compared to their productivity in local labor markets. Meanwhile, women are more likely to focus on home production, in which they have higher productivity than what they may obtain from the labor market. This finding also provides some evidence supporting the negative effects of fertility on female labor force participation and the labor supply and the positive effects of fertility on paternal labor force participation and the labor supply in the developing countries like Vietnam.

References

Airola, J. (2008). Labor supply in response to remittance income: The case of Mexico, *Journal of Developing Areas*, The Spring.

Allman, J., V. Q. Nhan, N. M. Thang, P. B. San, and V. D. Man (1991), Fertility and Family Planning in Vietnam, *Studies in Family Planning*, Vol. 22, No. 5, pp. 308-317.

Altonji, J.G., F. Hayashi, L.J. Kotlikoff (1989). Is the extended family altruistically linked? Direct test using micro data, *NBER working paper series,* No. 3046.

Angrist, J. D. and W. N. Evans (1998), Children and Their Parents' Labor Supply: Evidence From Exogenous Variation in Family Size, *American Economic Review,* Vol. 88, No. 3, pp. 450-477.

Anselin, L. (1981), Small Sample Properties of Estimators for the Linear Model with a Spatial Autoregressive Structure in the Disturbance. In: *Proceedings of the Pittsburgh Conference on Modeling and Simulation,* Vol. 12, pp. 899-90.

Anselin, L. (1985), Specification Tests and Model Selection for Spatial Interaction and the Structure of Spatial Dependence, *Final Report to the National Science Foundation*, The Ohio State University Research Foundation.

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

Anselin, L. (1988b), Estimation Methods for Spatial Autoregressive Structures, *Regional Science Dissertation and Monograph,* Series 8, Cornell University, Ithaca, N.Y.

Anselin, L. (1995), Spatial Patterns in Appalachian Growth and Development, *Research Paper 9512, Regional Research Institute*, West Virginia University.

Anselin, L. (1999), Spatial Econometrics, *Bruton Center*, School of Social Sciences University of Texas at Dallas.

Anselin, L. (2001), Rao's Score Test in Spatial Econometrics, *Journal of Statistical Planning and Inference,* Vol. 97, No. 1, pp. 113-39.

Anselin, L. (2006), Spatial Econometrics in RSUE Retrospect and Prospect, *National Center for Supercomputing Application*, University of Illinois.

Anselin, L. and A. Can (1998), Spatial Effects in Models of Mortgage Origination. In: *Proceedings of the 1998 Annual Mid Year Meeting of the American Real Estate and Urban Economics Association*. Washington, DC.

Anselin, L. and S. Rey (1991), The Performance of Tests for Spatial Dependence in a Linear Regression. *Technical Report 91-13, National Center for Geographic Information and Analysis*, University of California.

Anselin, L., A.K. Bera, R. Florax, and M.J. Yoon (1996), Simple Diagnostic Tests for Spatial Dependence, *Regional Science and Urban Economics*, Vol. 26, pp. 77-104.

Anselin, L., J. Le Gallo, and H. Jayet (2007), Spatial Panel Econometrics, In: *The Econometrics of Panel Data, Fundamentals, and Recent Developments in Theory and Practice,* Dordrecht, Kluwer.

Ashenfelter, O. and J. J. Heckman (1974), The Estimation of Income and Substitution Effects in a Model of Family Labor Supply, *Econometrica,* Vol. 42, No. 1, pp. 73-85.

Baltagi, B.H., S.H. Song, and W. Koh (2002), Testing Panel Data Regression Models with Spatial Error Correlation, *Journal of Econometrics*, Vol. 117. No. 1, pp. 123-50.

Baltigi, B.H. and D. Li (2001), LM Tests for Functional Form and Spatial Error Correlation, *International Regional science Review*, Vol. 24, pp. 194-225.

Baltigi, B.H. and D. Li (2004a), Testing for Linear and Log-linear Models against Box-Cox Alternatives with Spatial Lag Dependence, *Spatial and Spatiotemporal Econometrics Advances in Econometrics*, Vol. 18, pp. 35-74.

Baltagi, B.H. and D. Li (2004b), Prediction in Panel Data Model with Spatial Correlation. In: *Advances in spatial econometrics: Methodology, tools and application*, edited by Anselin, L., G.M. Florax, and S. Rey, 283-95. Heidelberg, Germany.

Becker G.S. and H.G. Lewis (1973), On the Interaction Between the Quantity and Quality of Children, *Journal of Political Economy,* Vol. 81, No. 2, pp. S279-S288.

Becker, G.S. (1985), Human Capital, Effort, and the Sexual Division of Labor, *Journal of Labor Economics,* Vol. 3, No. 1, pp. S33-58.

Belanger, D. (2002), Son Preference in a Rural Village in North Vietnam, *Studies in Family Planning,* Vol. 33, No. 4, pp, 321-334.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2007), Older and Wiser? Birth Order and IQ of Young Men, *Working Papers,* Center for Advanced Studies in Economics.

Blomquist, N. S. and B. U. Hansson (1990), The Effect of Taxes on Male and Female Labor Supply in Sweden, *Journal of Human Resources,* Vol. 25, No. 3, pp. 317-357.

Bloom, D. E., D. Canning, G. Fink, and J. E. Finlay (2007), Fertility, Female Labor Force Participation, and the Demographic Dividend, *NBER Working Papers,* National Bureau of Economic Research.

Bodson, P. and D. Peeters (1975), Estimation of Coefficients of a Linear Regression in the Presence of Spatial Autocorrelation: An Application to Belgian Labor demand Function, *Environment and Planning*, Vol. 7, pp. 455-72.

Box, G.E.P. and D.R. Cox (1964), An Analysis of Transformations, *Journal of the Royal Statistical Society*, Series B 26, pp. 211-52.

Boxall, P.C., W.H. Chan, M.L. McMillan (2005), The Impact of Oil and Natural Gas Facilities on Rural Residential Property Values: A Spatial Hedonic Analysis, *Resource and Energy Economics*, Vol. 27, No. 3, pp. 248-69.

Bronars, S. G. and J. Grogger (1994), The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment, *American Economic Review,* Vol. 84, No. 5, pp. 1141-1156.

Cain, G. G. and M. D. Dooley (1976), Estimation of a Model of Labor Supply, Fertility, and Wages of Married Women, *Journal of Political Economy,* Vol. 84, No. 4, pp. S179-99.

Can, A. and I. Megbolugbe (1997), Spatial and Housing Price Index construction, *Journal of Real Estate Finance and Economics,* Vol. 14, pp. 2003-22.

Carlin, P. S. and L. Flood (1997), Do Children Affect the Labor Supply of Swedish Men? Time Diary vs. Survey Data, *Labour Economics,* Vol. 4, No. 2, pp. 167-183.

Carliner, G., C. Robinson, and N. Tomes (1980), Female Labour and Fertility in Canada, *The Canadian Journal of Economics,* Vol. 13, No. 1, pp. 46-64.

Cassel, E. and R. Mendelsohn (1985), The Choice of Functional Form for Hedonic Price Equations: Comment, *Journal of Urban Economics*, Vol. 18, pp. 135-42.

Choi, H. J., J. Joesch, and S. Lundberg (2005), Work and Family: Marriage, Children, Child Gender and The Work Hours and Earnings of West German Men, *IZA Discussion Paper*.

Chun, H. and J. Oh (2002), An Instrumental Variable Estimate of The Effect of Fertility on the Labour Force Participation of Married Women, *Applied Economics Letters,* Vol. 9, No. 10, pp. 631-634.

Cliff, A.D and J.K. Ord (1972), Testing for Spatial Autocorrelation among Regression Residuals, *Geographical Analysis,* Vol. 4, pp. 267–84.

Cliff, A.D. and J.K. Ord (1981), *Spatial Processes, Models and Applications*, London: Pion.

Cogan, J. F. (1977), Labor Supply with Time and Money Costs of Participation. *RAND R-20044-HEW,*

Craig, S.G, J.E. Kohlharse, and D.H. Papell (1991), Chaos Theory in Microeconomics: An Application to Model Specification and Hedonic Estimation, *Reviews of Economics and Statistics,* Vol. 73, pp. 208-15.

Cristia, J. P. (2006), The Effects of The First Child on Female Labor Supply: Evidence from Women Seeking Fertility Services, *Working Paper Series,*

Cropper, M.L., L.B. Deck, and K.E. Mc Connell (1988), On the Choice of Functional Form for Hedonic Price Functions, *Review of Economics and Statistics*, Vol. 70, pp. 668-75.

Cruces, G. and S. Galiani (2007), Fertility and Female Labor Supply in Latin America: New Causal Evidence, *Labour Economics,* Vol. 14, No. 3, pp. 565-573.

Crushing, B. (2005), Specification of Functional Form in Models of Population Migration, for presentation at *the Southern Regional Science Association Conference*, Arlington, VA.

Dacey, M. (1968), A Review of Measures of Contiguity for Two and K-color Maps, *In Spatial Analysis: A reader in statistical Geography*, Englewood Cliffs, N.J: Prentice Hall, pp. 479-95.

Davidson, R. and J.G. MacKinnon (1984), Model Specification Tests Based on Artificial Linear Regressions, *International Economic Review*, Vol. 25, pp. 485-502.

Davidson, R. and J.G. MacKinnon (1985), Testing Linear and Log-linear Regressions against Box-Cox Alternatives, *Canadian Journal of Economics*, Vol. 18, pp. 499-517.

Davidson, R. and J.G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press, New York.

Del Boca, D., S. Pasqua, and C. Pronzato (2005), Why Are Fertility and Women's Employment Rates so Low in Italy? Lessons from France and the U.K., *Labour,* Vol. 19, No.1, pp. 51-79.

DeTray, D. N. (1973), Child Quality and the Demand for Children, *Journal of Political Economy,* Vol. 81, No. 2, pp. S70-95.

Dinerman, I.R. (1982). Migrants and Stay-at-Homes: A comparative study of rural migration from Michoacán, Mexico, *Monographs in U.S. Mexican Studies*, No. 5, U.C. San Diego.

Elad, R.L., I.D. Clifton, and J.E. Epperson (1994), Hedonic Estimation Applied to the Farmland Market in Georgia, *Journal of Agricultural and Applied Economics*, Vol. 26, pp. 351-66.

Ertur, C., J. L. Gallo, and C. Baumont (2006), The European Regional Convergence Process, 1980-1995: Do Spatial Regimes and Spatial Dependence Matter? *International Regional Science Review*, Vol. 29, No. 1, pp. 3-34.

Fattouh, B., L. Harris, and P. Scaramozzino (2005), Non-linearity in the Determinants of Capital Structure: Evidence from UK Firms, *Social Science Research Network*.

Fik, T.J. and G.F. Mulligan (1998), Functional Form and Spatial Interaction Models, *Environment and Planning,* A 30, pp. 1497-1507.

Fleisher, B. M. and G.F. Rhodes (1979), Fertility, Women's Wage Rates, and Labor Supply, *The American Economic Review,* Vol. 69, No. 1, pp. 14-24.

Funkhouser, E. (1992). Migration from Nicaragua: Some Recent Evidence, *World Development,* Vol. 20 No. 8, pp. 1209-1218.

Gatrell, A.C. (1989), On the Spatial Representation and Accuracy of Address-Based Data in the United Kingdom. *International Journal of Geographical Information Systems,* Vol. 3, No. 4, pp. 335-48.

General Statistical Office (2000). *Labor and job survey in 1999*, [Dieu tra ve nganh nghe va lao dong nam 1999], Statistical Publishing House, Hanoi.

General Statistical Office (2007). *Statistical Year Book*, [Nien Giam Thong Ke], Statistical Publishing House, Hanoi

Gilbert, N. (2005), What Do Women Really Want?, *http://www.thepublicinterest.com-current/-article2.html*.

Graaf, T., R. Florax, P. Nijkamp, and A. Reggiani (2001), Misspecification Tests for Spatial Regression Model, *Journal of Regional Science*, Vol. 42, pp. 255-76.

Gregory, P. R. (1982), Fertility and Labor Force Participation in the Soviet Union and Eastern Europe, *Review of Economics and Statistics,* Vol. 64, No. 1, pp. 18-31.

Griffth, D.A, H.P. Paelinck, and R.A. Van Gasten (1998), The Box-Cox Transformation: Computational and Interpretation Features of Parameter, *In Econometrics Advances in Spatial Modeling and Methodology: Essays in Honor of Jean Paelinck*, The Netherlands: Kluwer.

Gronau, R. (1973), The Intra-Family Allocation of Time: The Value of The Housewives' Time, *American Economic Review,* Vol. 63, No. 4, pp. 634-651.

Groot, W.J.N. and H. M. Van Den Brink (1992). Labor supply and non parental child care demand, Research Memorandum University of Amsterdam,

Hakim, C. (2003), *Models of the Family in Modern Societies: Ideals and Realities*, June, Ashgate.

Haughton, J. and D. Haughton (1995), Son Preference in Vietnam, *Studies in Family Planning,* Vol. 26, No. 6, pp. 325-337.

Haughton, J. and D. Haughton (1998), Are Simple Tests of Son Preference Useful? An Evaluation Using Data from Vietnam, *Journal of Population Economics,* Vol. 11, No. 4, pp. 495-516.

Heckman, J. J. (1974), Shadow Prices, Market Wages, and Labor Supply, *Econometrica,* Vol. 42, No. 4, pp. 679-694.

Heckman, J. J. and R. J. Willis (1977), A Beta-Logistic Model for the Analysis of Sequential Labor Force Participation by Married Women, *Journal of Political Economy,* Vol. 85, No. 1, pp. 27-58.

Hieu, D. T., J. Stoeckel, and N. V. Tien (1993), Pregnancy Termination and Contraceptive Failure in Vietnam, *Asia-Pacific Population Journal*, Vol. 4, pp. 3-18.

Hollander, D. (1996), Son Preference Remains Strong among Vietnamese Women, but Has only a Modest Effect on Fertility, *International Family Planning Perspectives,* Vol. 22, No. 2, PP. 84-85.

Huan, N. D. (1997), A Socioeconomic Analysis of the Determinants of Fertility: The Case of Vietnam, *Journal of Population Economics*, Vol. 10, No. 3, pp. 251-71.

Huh, S. and S.J. Kwak (1997), The Choice of Functional Form and Variables in Hedonic Price Model in Seoul, *Urban Studies*, Vol. 34, No. 7, pp. 989-89.

Jacobsen, J. P., J. W. Pearce, and J. L. Rosenbloom (1999), The Effects of Child-Bearing on Married Women's Labor Supply and Earnings: Using Twin Births as a Natural Experiment, *The Journal of Human Resources,* Vol. 34, No. 3, pp. 449-474.

Ketkar, S.L. (1979). Determinants of Fertility in a Developing Society: The Case of Sierra Leone, *Population Studies,* Vol. 33, No. 3, pp. 479-488

Kim, J., and Aassve, A. (2006), Fertility and Its Consequences on Family Labor Supply, *IZA Discussion Paper, 2162*.

Koch, W. (2006), Growth and Spatial Dependence in Europe, *RePEc: lat:legeco: 2006-02.*

Lacombe, D.J. and T.M. Shaughnessy (2007), Accounting for Spatial Error Correlation in the 2004 Presidential Popular Vote, *Public Finance Review*, Vol. 35, No. 4, pp. 480-99.

Lacroix, G., C. Sofer, and M. Picot (1997). The Extent of Labour Specialization in the Extended Family: A Theoretical and Empirical Analysis. Social Science Research Network Working Paper Series.

Le, C.Q., D. Li, (2008), Double-Length Regression tests for testing functional forms and spatial error dependence, *Economics Letters,* Vol. 101, No. 3, pp 253-257.

Ledent, J. (1986), A Model of Urbanization with Non-linear Migration Flows, *International Regional Science Reviews*, Vol. 10, pp. 221-42.

Lehrer, E. L. (1992), The Impact of Children on Married Women's Labor Supply, *The Journal of Human Resources,* Vol. 27, No. 3, pp. 422-444.

LeSage, J.P. (1998), *Spatial Econometrics*, University of Toledo.

Longhi, S. and P. Nijkamp (2007), Forecasting Regional Labor Market Developments under Spatial Autocorrelation, *International Regional Science Review*, Vol. 30, No. 2, pp. 100-119.

Lundberg, J. (2004), Using Spatial Econometrics to Analyze Local Growth in Sweden, *ERSA Conference Papers ERSA04P25*.

Lundberg, J. (2006), Spatial Interaction Model of Spillovers from Locally Provided Public Services, *Regional Studies*, Vol. 40, No. 6, pp. 631-44.

Lundberg, S. (2005), The Division of Labor by New Parents: Does Child Gender Matter? *Working Papers, Department of Economics, University of Washington,*

Lundberg, S. and E. Rose (1999), The Determinants of Specialization within Marriage, Mimeo, *University of Washington*.

Lundberg, S. and E. Rose (2002), The Effects of Sons And Daughters on Men's Labor Supply and Wages, *Review of Economics and Statistics,* Vol. 84, No. 2, pp. 251-268.

Manski, C.F. (1993), Identification of Endogenous Social Effects: The Reflection Problems, *Review of Economics Studies*, Vol. 60, pp. 531-42.

McMillen, D.P. (1992), Probit with Spatial Autocorrelation, *Journal of Regional Science*, Vol. 32, pp. 335-48.

Miyazaki, K. (2005), Box-Cox Transformation and Option Pricing, *Japanese Journal of Applied Statistics,* Vol. 34. No. 2, pp. 1-23.

Mobley, L., E. Root, L. Anselin, N. Lozano, and J. Koschinsky (2006), Spatial Analysis of Elderly Access to Primary Care Services, *International Journal of Health Geographics,* Vol. 5, No. 19.

Mok, H., P. Chan and Y. S. Cho (1995), A Hedonic Price Model for Private Properties in Hong Kong, *Journal of Real Estate Finance and Economics*, Vol. 10, pp. 37-48.

Moran, P.A. (1948), The Interpretation of Statistical Maps. *Biometrika,* Vol. 35, pp. 255-60.

Moran, P.A. (1950), A Test for the Serial Dependence of Residuals. *Biometrik,* Vol. 37, pp. 178-81.

Ord, J.K. (1975), Estimation Methods for Models of Spatial Interaction, *Journal of American Statistical Association,* Vol. 70, pp. 120-26.

Oyama, M. (2001), Fertility Decline and Female Labor Force Participation in Japan, *Working Paper, University of Chicago,*

Pace, R.K. and R. Barry, V.C. Slawson, and C.F. Sirmans (1999), Simultaneous Spatial and Functional Form Transformations, in *Advances in Spatial Econometrics*, Anselin, L., G.M. Florax, and J. R. Sergio, Springer-Verlag New York, LLC.

Pencavel, J. (1986), Labor Supply of Men: A survey. *Handbook of Labor Economics* (pp. 3-102), North-Holland, New York.

Phai, V. N. (1999), The 1997 Vietnam Demographic and Health Survey: Fertility and Fertility Regulation, *Vietnam Population News*, Vol. 10, Jan-Mar, 2-3.

Pinkse, J. (1998), Asymptotic Properties of the Moran and Related Tests and a Test for Spatial Correlation in Probit Models. *Working Paper*, University of British Columbia, Vancouver, BC.

Pinkse, J. and M.E. Slade (1998), Contracting in Space: An Application of Spatial Statistics to Discrete-choice Models, *Journal of Econometrics,* Vol. 85, pp. 125-54.

Rey, S. and L. Anselin (2006), Recent Advances in Software for Spatial Analysis in the Social Sciences, *Geographical Analysis,* Vol. 38, No. 1, pp. 1-4.

Rica, S. D. and D. M. Ferrero (2003), The Effect of Fertility on Labour Force Participation: The Spanish Evidence, *Spanish Economic Review,* Vol. 5, No. 2, pp. 153-172.

Robinson, C. and N. Tomes (1982), Family Labour Supply and Fertility: A Two-Regime Model, *The Canadian Journal of Economics,* Vol. 15, No. 4, pp. 706-734.

Rosenzweig, M. R. and K. I. Wolpin (1980), Life-Cycle Labor Supply and Fertility: Causal Inferences from Household Models, *Journal of Political Economy*, Vol. 88, No. 2, pp. 328-348.

Salkever, D. S. (1982), Children’s Health Problems: Implications for Parental Labor Supply and Earnings, *Economic aspects of health* (pp. 221-251).

Schultz, P. T. (1978), The Influence of Fertility on Labor Supply of Married Women: Simultaneous Equation Estimates. *Aspects of labor economics* (pp. 273-351). Greenwich, CT: JAI Press:

Schultz, P. T. (1990), Testing the Neoclassical Model of Family Labor Supply and Fertility, *Journal of Human Resources,* Vol. 25, No. 4, pp. 599-634.

Sharif, M. (2003), *Work Behavior of the World's Poor: Theory, Evidence and Policy*, British of Congress cataloguing in publication data.

Shimizu, C., K.G. Nishimura, and K. Karato (2007), Non-linearity of Housing Price Structure, *CSIS-Discussion Paper No. 86.*

Stock, J. H., J.H. Wright, and M. Yogo (2002), A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments, *Journal of Business and Economic Statistics,* Vol. 20, No. 4, pp. 518-529.

Tran, Q.T., Nguyen, T.T., Tran, Q.T., Le, Q.C. (2004). Non-farm households' development in Vietnam, *Vietnam Economic Research Network.*

Upton, G.J. and B. Fingleton (1985), *Spatial Data Analysis by Example*, Vol. 1, Chichester, Wiley.

Vach, H. T. and A. Bishop (1996), Pregnancy Testing in Vietnam to Reduce Unnecessary Abortion, *American Public Health Association Conference,* University of Chicago.

Waldfogel, J. (1998), The Family Gap for Young Women in the United States and Britain: Can Maternity Leave Make a Difference? *Journal of Labor Economics,* Vol. 16, No. 3, pp. 505-545.

Willis, R. (1973), A New Approach to the Economic Theory of Fertility Behavior, *Journal of Political Economy*, Vol. 81, pp. S14-64.

Woodruff, C. and R. Zenteno (2001). Remittances and Microenterprises in Mexico, *Pacific Studies working paper*.

World Bank (2006). World Development Indicators, Washington, D.C.

Yang, Z., C. Li, and Y.K. Tse (2006), Functional Form and Spatial Dependence in Dynamic Panels, *Economics Letters*, Vol. 91, pp. 138-45.

Appendix

Appendix 1: Fertility and maternal labor force participation

	Pool		Urban		Rural	
	Probit	IV Probit	Probit	IV Probit	Probit	IV Probit
Fertility	-0.118***	-0.803***	0.119	0.206	-0.151***	-0.885***
	(-3.917)	(-22.441)	(1.360)	(0.788)	(-4.373)	(-40.860)
Mothers' characteristics						
Age	-0.059	0.073	-0.117	-0.130	-0.059	0.090*
	(-0.850)	(1.442)	(-0.801)	(-0.867)	(-0.671)	(1.823)
Age squares	0.001	-0.001*	0.001	0.001	0.001	-0.001*
	(0.596)	(-1.691)	(0.510)	(0.589)	(0.578)	(-1.845)
Household head	0.171	-0.087	-0.053	-0.037	0.277	-0.111
	(1.373)	(-0.946)	(-0.311)	(-0.210)	(1.357)	(-0.929)
Primary school	1.353***	0.615***	0.557*	0.569*	1.588***	0.437***
	(10.711)	(4.832)	(1.697)	(1.728)	(11.070)	(3.504)
Middle school	1.465***	0.805***	1.080***	1.101***	2.245***	0.585***
	(13.207)	(5.557)	(3.222)	(3.249)	(12.942)	(4.040)
High school	1.531***	0.808***	1.273***	1.292***	1.648***	0.589***
	(11.583)	(6.082)	(3.805)	(3.832)	(10.802)	(4.544)
College and higher	1.565***	0.912***	2.862***	2.864***	1.425***	0.603***
	(10.205)	(6.535)	(5.152)	(5.174)	(8.650)	(4.541)
On farm	0.306***	0.398***	0.402***	0.378**	0.423***	0.399***
	(3.659)	(6.285)	(2.768)	(2.320)	(3.856)	(5.716)
Fathers' characteristics						
Age	-0.243***	-0.027	0.159	0.143	-0.386***	-0.022
	(-3.201)	(-0.480)	(1.056)	(0.916)	(-3.962)	(-0.393)
Age squares	-0.005	-0.004	-0.012*	-0.012*	0.008	0.003
	(-0.834)	(-0.983)	(-1.781)	(-1.814)	(0.584)	(0.365)
Hourly wages	0.003***	0.000	-0.002	-0.001	0.005***	0.000
	(3.216)	(0.676)	(-0.865)	(-0.752)	(3.818)	(0.470)
Primary school	0.277*	0.097	1.035***	1.047***	0.095	-0.035
	(1.802)	(0.780)	(2.951)	(2.981)	(0.529)	(-0.282)
Middle school	0.610***	0.251*	1.093***	1.113***	0.471**	0.066
	(3.742)	(1.895)	(3.130)	(3.163)	(2.409)	(0.505)
High school	0.385**	0.134	0.897**	0.918***	0.222	-0.010
	(2.392)	(1.038)	(2.545)	(2.581)	(1.158)	(-0.078)
College and higher	0.009	-0.023	0.532	0.554	-0.211	-0.077
	(0.053)	(-0.172)	(1.423)	(1.467)	(-1.066)	(-0.561)
On farm	-0.289***	-0.096	-0.120	-0.120	-0.263**	-0.004
	(-3.483)	(-1.521)	(-0.756)	(-0.760)	(-2.347)	(-0.069)
Households' characteristics						
Non-labor H income	-0.066***	-0.020	-0.042**	-0.045**	-0.142***	-0.049***
	(-4.538)	(-1.609)	(-2.101)	(-2.027)	(-5.281)	(-2.737)
Grandparents	9.939**	5.279*	1.918	1.949	10.996**	3.647*
	(2.254)	(1.884)	(0.519)	(0.529)	(2.331)	(1.775)
Grandparents' age	-0.444***	-0.198**	-0.256**	-0.262**	-0.485***	-0.118*
	(-3.579)	(-2.422)	(-2.307)	(-2.350)	(-3.610)	(-1.921)
Grandparent Age squares	0.003***	0.001**	0.002***	0.002***	0.004***	0.001*
	(4.026)	(2.520)	(2.671)	(2.717)	(3.990)	(1.813)
Urban	-0.077	-0.301***				
	(-0.869)	(-4.571)				
Local labor market situation						
Annual new jobs	-0.001	-0.001**	-0.003**	-0.003**	0.003	-0.001
	(-1.272)	(-2.041)	(-2.396)	(-2.398)	(1.137)	(-0.964)
Observations	3935	3935	832	832	3103	3103

Appendix 2: Fertility and paternal labor force participation

	Pool		Urban		Rural	
	Probit	IV Probit	Probit	IV Probit	Probit	IV Probit
Fertility	0.271***	0.908***	0.188	1.091***	0.281***	0.847***
	(4.128)	(24.732)	(1.557)	(20.253)	(3.377)	(12.743)
Mothers' characteristics						
Age	0.383***	0.059	0.546***	0.051	0.356***	0.107
	(5.984)	(1.195)	(3.173)	(0.481)	(5.048)	(1.576)
Age squares	-0.005***	-0.001	-0.007***	0.000	-0.004***	-0.001
	(-5.511)	(-0.827)	(-3.003)	(-0.258)	(-4.516)	(-1.238)
Hourly wage	0.032*	0.019**	0.040	0.010	0.020	0.022
	(1.848)	(2.012)	(1.479)	(0.860)	(0.886)	(1.425)
Household head	-0.886***	-0.242**	-1.130***	-0.234*	-0.739***	-0.260
	(-7.062)	(-2.416)	(-6.121)	(-1.808)	(-3.741)	(-1.559)
Primary school	0.027	0.203	0.827	0.840**	-0.364	-0.132
	(0.088)	(1.137)	(1.528)	(2.559)	(-0.910)	(-0.481)
Middle school	-0.476	0.056	-0.043	0.582*	-0.670*	-0.216
	(-1.568)	(0.307)	(-0.082)	(1.815)	(-1.669)	(-0.762)
High school	0.051	0.139	0.919*	1.032***	-0.466	-0.300
	(0.165)	(0.778)	(1.706)	(3.173)	(-1.164)	(-1.096)
College and higher	0.284	0.122	0.948*	0.870**	-0.091	-0.166
	(0.873)	(0.645)	(1.649)	(2.532)	(-0.216)	(-0.587)
On farm	0.910***	0.220**	1.192***	0.253*	0.781***	0.293**
	(7.549)	(2.337)	(5.530)	(1.836)	(5.112)	(2.068)
Fathers' characteristics						
Age	0.460***	0.113**	0.773***	0.171	0.375***	0.121*
	(7.681)	(2.285)	(4.516)	(1.562)	(5.721)	(1.799)
Age squares	-0.006***	-0.002***	-0.009***	-0.002*	-0.005***	-0.002**
	(-8.086)	(-2.601)	(-4.521)	(-1.730)	(-6.266)	(-2.115)
Primary school	0.677***	0.526***	-0.427	0.183	0.908***	0.742***
	(3.126)	(3.715)	(-0.721)	(0.551)	(3.673)	(3.925)
Middle school	0.591**	0.553***	-0.353	0.321	0.603**	0.583***
	(2.317)	(3.507)	(-0.593)	(0.971)	(2.527)	(3.320)
High school	0.749***	0.593***	-0.110	0.432	0.881***	0.728***
	(3.399)	(4.118)	(-0.180)	(1.291)	(3.524)	(3.852)
College and higher	1.461***	0.873***	0.945	0.948***	1.308***	0.868***
	(5.559)	(4.885)	(1.422)	(2.645)	(4.239)	(3.524)
On farm	0.382***	0.111	0.477**	0.183	0.349***	0.150
	(3.501)	(1.605)	(2.354)	(1.597)	(2.585)	(1.492)
Households' characteristics						
Non-labor H income	0.153***	0.063**	0.124*	0.004	0.217**	0.133**
	(2.942)	(2.088)	(1.746)	(0.118)	(2.453)	(2.182)
Grandparents	-0.547	0.237	0.675	0.886	-0.502	0.147
	(-0.404)	(0.336)	(0.222)	(0.777)	(-0.350)	(0.165)
Grandparents' age	0.075*	-0.041*	-0.013	-0.087*	0.072	-0.026
	(1.710)	(-1.723)	(-0.109)	(-1.854)	(1.546)	(-0.815)
Grandparent age squares	-0.001**	0.000	0.000	0.001*	-0.001**	0.000
	(-2.207)	(1.523)	(0.199)	(1.660)	(-2.029)	(0.540)
Urban	-0.927***	-0.252***				
	(-8.177)	(-2.780)				
Local labor market situation						
Annual new jobs	-0.004***	-0.001	-0.004**	-0.001	-0.005**	-0.001
	(-3.628)	(-1.205)	(-2.462)	(-1.095)	(-2.571)	(-0.563)
Observations	3800	3800	771	771	3029	3029

Numbers in parenthesis are t-values. Significant levels of 0.1 are denoted by *; 0.05 by **; and 0.01 by ***.

Appendix 3: Permission from Sage Publication Inc. to use the Table 1 in Baltagi and Li (2001)

Customer Information

Customer: Canh Le
Account Number: 3000143904
Email: catfishneu@yahoo.com
Phone: 1-785-395-5780
Payment Method: Credit Card ending in 6720

Order Details

INTERNATIONAL REGIONAL SCIENCE REVIEW
Order detail ID: 22208579

ISBN/ISSN: 01600176
Publication Year: 2001
Publisher: SAGE PUBLICATIONS, INC.
Rightsholder: Sage Publications Inc. Journals
Author/Editor: Badi Baltagi and Dong Li

Permission Status: ✔ **Granted**
Permission type: Republish into a book, journal, newsletter...
Requested use: Dissertation
Republishing organization: KSU
Organization status: Not for profit
Republication date: 05/31/2009
Circulation/Distribution: 1
Type of content: Figure, diagram, or table
Description of requested content: LM Test for Functional Form and Spatial Error Correlation
Page range(s): 203
Requested content's publication date: 04/01/2001

Billing Status: **Charged to Credit Card**

Rightsholder terms apply

Appendix 4: Permission from Elsevier Jai. to use the Table 1 in Baltagi and Li (2004a)

Customer Information

Customer: Canh Le
Account Number: 3000143904
Email: catfishneu@yahoo.com
Phone: 1-785-395-5780
Payment Method: Credit Card ending in 6720

Order Details

SPATIAL AND SPATIOTEMPORAL ECONOMETRICS
Order detail ID: 22628001

ISBN/ISSN: 9780762311484
Publication Year: 2004
Publisher: ELSEVIER JAI
Rightsholder: Emerald Group Publishing Limited
Author/Editor: Badi H. Baltagi and Dong Li

Permission Status: ✔ **Granted**
Permission type: Republish into a book, journal, newsletter...
Requested use: Dissertation
Republishing organization: KANSAS STATE UNIVERSITY
Organization status: Not for profit
Republication date: 05/31/2009
Circulation/Distribution: 1
Type of content: Figure, diagram, or table
Description of requested content: The Table 1
Page range(s): 45
Requested content's publication date: 11/04/2004

Billing Status: **Charged to Credit Card**

Rightsholder terms apply

Appendix 5: A part of permission from Elsevier Jai. to use the paper of Le and Li (2008)

ELSEVIER LICENSE
TERMS AND CONDITIONS

Nov 18, 2008

This is a License Agreement between Canh Q Le ("You") and Elsevier ("Elsevier"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

Supplier	Elsevier Limited The Boulevard,Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Canh Q Le
Customer address	1700 Hillcrest Dr
	Manhattan, KS 66502
License Number	2071990993377
License date	Nov 18, 2008
Licensed content publisher	Elsevier
Licensed content publication	Economics Letters
Licensed content title	Double-Length Regression tests for testing functional forms and spatial error dependence
Licensed content author	Canh Quang Le and Dong Li
Licensed content date	December 2008
Volume number	
Issue number	
Pages	0
Type of Use	Thesis / Dissertation
Portion	Full article
Format	Both print and electronic
You are an author of the Elsevier article	Yes
Are you translating?	No
Purchase order number	
Expected publication date	May 2009
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
Value added tax 0.0%	0.00 USD
Total	0.00 USD