ISSN: 1525-3066

Center for Policy Research
Working Paper No. 84

PREDICTION IN THE PANEL DATA MODEL WITH SPATIAL CORRELATION: THE CASE OF LIQUOR

Badi H. Baltagi and Dong Li

Center for Policy Research
Maxwell School of Citizenship and Public Affairs
Syracuse University
426 Eggers Hall
Syracuse, New York 13244-1020
(315) 443-3114 | Fax (315) 443-1081
e-mail: ctrpol@syr.edu

July 2006

$5.00

CENTER FOR POLICY RESEARCH – Spring 2006

Timothy Smeeding, Director
Professor of Economics & Public Administration

Associate Directors

Margaret Austin
Associate Director
Budget and Administration

Douglas Wolf
Professor of Public Administration
Associate Director, Aging Studies Program

John Yinger
Professor of Economics and Public Administration
Associate Director, Metropolitan Studies Program

SENIOR RESEARCH ASSOCIATES

Badi Baltagi	Economics
Pablo Beramendi	Political Science
Dan Black	Economics
Lloyd Blanchard	Public Administration
William Duncombe	Public Administration
Gary Engelhardt	Economics
Deborah Freund	Public Administration
Madonna Harrington Meyer	Sociology
Christine Himes	Sociology
William C. Horrace	Economics
Bernard Jump	Public Administration
Duke Kao	Economics
Eric Kingson	Social Work
Thomas Kniesner	Economics
Tom Krebs	Economics
Jeffrey Kubik	Economics
Andrew London	Sociology
Len Lopoo	Public Administration
Jerry Miner	Economics
Jan Ondrich	Economics
John Palmer	Public Administration
Lori Ploutz-Snyder	Exercise Science
David Popp	Public Administration
Grant Reeher	Political Science
Stuart Rosenthal	Economics
Ross Rubenstein	Public Administration
Margaret Usdansky	Sociology
Michael Wasylenko	Economics
Janet Wilmoth	Sociology

GRADUATE ASSOCIATES

Javier Baez	Economics
Sonali Ballal	Public Administration
Dana Balter	Public Administration
Jason Bladen	Economics
Jesse Bricker	Economics
Maria Brown	Social Science
Yong Chen	Economics
Ginnie Cronin	Social Science
Ana Dammert	Economics
Mike Eriksen	Economics
Katie Fitzpatrick	Economics
Alexandre Genest	Public Administration
Julie Anna Golebiewski	Economics
Nadia Greenhalgh-Stanley	Economics
Yue Hu	Economics
Becky Lafrancois	Economics
Joseph Marchand	Economics
Neelakshi Medhi	Social Science
Larry Miller	Public Administration
Wendy Parker	Sociology
Emily Pas	Economics
Shawn Rohlin	Economics
Cynthia Searcy	Public Administration
Talisha Searcy	Public Administration
Jeff Thompson	Economics
Wen Wang	Economics
Yoshikun Yamamoto	Public Administration

STAFF

Kelly Bogart	Administrative Secretary
Martha Bonney	Publications/Events Coordinator
Karen Cimilluca	Administrative Secretary
Kati Foley	Administrative Assistant
Kitty Nasto	Administrative Secretary
Candi Patterson	Computer Consultant
Mary Santy	Administrative Secretary
Tammy Tanner	Librarian/Office Coordinator

Abstract

This paper considers the problem of prediction in a panel data regression model with spatial autocorrelation in the context of a simple demand equation for liquor. This is based on a panel of 43 states over the period 1965-1994. The spatial autocorrelation due to neighboring states and the individual heterogeneity across states is taken explicitly into account. We compare the performance of several predictors of the states demand for liquor for one year and five years ahead. The estimators whose predictions are compared include OLS, fixed effects ignoring spatial correlation, fixed effects with spatial correlation, random effects GLS estimator ignoring spatial correlation and random effects estimator accounting for the spatial correlation. Based on RMSE forecast performance, estimators that take into account spatial correlation and heterogeneity across the states perform the best for one year ahead forecasts. However, for two to five years ahead forecasts, estimators that take into account the heterogeneity across the states yield the best forecasts.

JEL Code: C21, C23, C53
Keywords: Prediction, spatial correlation, panel data, liquor demand

We thank Lance Bachmeier, the editor and an anonymous referee for helpful comments and suggestions.

1 Introduction

This paper focuses on prediction in a simple demand equation for liquor based on a panel of 43 states over the period 1965-1994. The spatial autocorrelation due to neighboring states and the individual heterogeneity across states is taken explicitly into account. In order to explain how spatial autocorrelation may arise in the demand for liquor, we note that liquor prices vary among states primarily due to variation in state taxes on liquor. For example, in 1983, state excise taxes ranged from $1.50 per gallon in a low state tax like Maryland to $6.50 per gallon in Florida. In 1984, apparent per capita consumption of alcohol for persons 14 years and older in New Hampshire was 4.91 gallons, a little less than twice the national median of 2.64 gallons per capita. This does not imply that New Hampshire residents are heavy drinkers. Carlson (1985, p.31) reports that "about 55% of New Hampshire's $155 million in annual liquor sales is to out of state tipplers." Border effect purchases not explained in the demand equation can cause spatial autocorrelation among the disturbances. [1]At the county level, liquor is not available in dry counties and consumers are forced to buy it from adjacent wet counties. The availability and pricing of liquor also varies as we move from private licensed states to monopoly states. Private licensed states are states with privately owned liquor stores that are licensed by the state. Monopoly states have a legal monopoly on the wholesale or retail of liquor. For more on the effects of the two governmental systems on prices, revenues and consumption of liquor, see Simon (1966) and Zardkoohi and Sheer (1984).

This paper models the demand for liquor as follows:

$$y_{it} = x'_{it}\beta + \varepsilon_{it} \qquad i = 1, ..., N;\; t = 1, ..., T \tag{1}$$

where y_{it} denotes the real per capita consumption of liquor measured in gallons of distilled spirits by persons of drinking age (16 years and older). The explanatory variables include the average retail price of a 750 ml of Seagram 7 expressed in real terms, and the real per capita disposable income of each state and a time trend. All variables, except the time trend, are expressed in logarithms and the estimated coefficients represent elasticities. Per capita consumption of liquor is obtained from the *Distilled Spirits Institute,* the price series is obtained from various issues of *The Liquor Handbook* and updated using the

[1]In fact, Baltagi and Griffin (1995) used the minimum price in neighboring states to capture border effects purchases.

price of alcoholic beverages from the inter city cost of living index published quarterly by the American Chamber of Commerce Researchers Associates. Per capita disposable income data on a state basis are published in various issues of the *Survey of Current Business.* Population data are obtained from various issues of the *Current Population Reports.* Price deflators are obtained from the *Bureau of Labor Statistics.* $N = 43$ states and $T = 30$ years. We only use the first 25 years for estimation and reserve the last 5 years for out of sample forecasts. For data sources, see Baltagi and Griffin (1995). Here, we update the data 12 years from 1983 to 1994. The disturbance term follows an error component model with spatially autocorrelated residuals, see Anselin (1988, p 152). The disturbance vector for time t is given by

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\mu} + \boldsymbol{\phi}_t \tag{2}$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$, $\boldsymbol{\mu} = (\mu_1, ..., \mu_N)'$ denotes the vector of state effects and $\boldsymbol{\phi}_t = (\phi_{1t}, ..., \phi_{Nt})'$ are the remainder disturbances which are independent of $\boldsymbol{\mu}$. The $\boldsymbol{\phi}_t$'s follow the spatial error dependence model

$$\boldsymbol{\phi}_t = \lambda \boldsymbol{W} \boldsymbol{\phi}_t + \boldsymbol{\nu}_t \tag{3}$$

where $\boldsymbol{W}$ is the matrix of known spatial weights of dimension $N \times N$ and λ is the spatial autoregressive coefficient. $\boldsymbol{\nu}_t = (\nu_{1t}, ..., \nu_{Nt})'$ is $iid(0, \sigma_\nu^2)$ and is independent of $\boldsymbol{\phi}_t$ and $\boldsymbol{\mu}$. The spatial matrix $\boldsymbol{W}$ is constructed as follows: a neighboring state takes the value 1, otherwise it is zero. The rows of this matrix are normalized so that they sum to one. The $\boldsymbol{\mu}_i$'s are the unobserved state specific effects which can be fixed or random, see Hsiao (1986). State specific effects include but are not limited to the following: (i) States like Montana, New Mexico and Arizona with Indian reservations sell tax-exempt liquor. (ii) States like Florida, Texas, Washington and Georgia with tax exempt military bases. (iii) Utah, a state with a high percentage of Mormons (a religion which forbids drinking) had an adult per capita consumption of liquor in 1994 of 1.2 gallons. This is much less than the national average of 1.82 gallons per adult. (iv) Nevada, a highly touristic state, has per capita consumption of liquor in 1994 of 4.68 gallons which is more than twice the national average. Not accounting for these state specific effects may lead to biased estimates. There are also numerous government interventions and restrictions as well as health warnings and Surgeon General's reports. These include the Alcohol Traffic Safety Act of 1983 which provided states with the financial incentives to enforce stringent drunk driving laws. Also the Federal Uniform Drinking Age Act of 1984 passed by Congress to pressure all states into raising the drinking age to 21. Numerous warning labels on all alcoholic beverages warning pregnant women about the dangers of drinking and the public about the dangers of drinking and driving. Kenkel (1993) reports that between 1981 and 1986, 729 state laws were enacted pertaining to drunk driving.

2 Estimation

Table 1 reports the estimates of a simple, albeit naive demand model for liquor using pooled OLS.[2] These estimates ignore the states heterogeneity and the spatial autocorrelation. The price elasticity estimate is -0.77, while the income elasticity estimate is 1.47 and both are statistically significant. Next, we take into account the spatial autocorrelation, and estimate the model using maximum likelihood method described in Anselin (1988). This assumes normality of the disturbances but ignores the heterogeneity across states. The resulting estimates are reported as pooled spatial in Table 1. This yields a slightly higher price (-0.82) and income elasticities (1.61) than OLS ignoring the spatial correlation. Both elasticities are significant. The estimate of λ is 0.34.[3] In addition, we conducted a grid search procedure over λ to ensure a global maximum. The likelihood ratio test for $\lambda = 0$ yields a value of 102.9 which is asymptotically distributed as χ^2_1 under the null hypothesis. The null is rejected justifying concern over spatial autocorrelation.

Table 2 allows for different parameter (heterogeneous) estimates for each year. The first set of estimates give the cross-sectional demand equation estimates using OLS for each year. The price elasticity estimates varied between -1.41 in 1989 to 0.09 in 1977, while the income elasticity estimates varied between 0.97 in 1989 to a high of 1.75 in 1971. Pesaran and Smith (1995) suggested averaging these heterogeneous estimates to obtain a pooled estimator. This yields a price elasticity estimate of -0.58 and an income elasticity estimate of 1.45, both of which are significant. These are reported as average heterogeneous OLS in Table 1. These individual cross-section regressions and their average do not take the spatial autocorrelation into account. Using the normality assumption, we

[2]For a dynamic demand model of liquor, see Baltagi and Griffin (1995).

[3]This was obtained using the Constrained Optimum (CO) module with GAUSS version 4.0.28.

re-estimate these cross-sectional demand equations using the maximum likelihood estimates (MLE) described in Anselin (1988) which account for spatial autocorrelation in the disturbances. These heterogeneous spatial estimates are reported in Table 2 along with the corresponding estimate of λ. We also report for each year the Lagrange Multiplier test for $\lambda = 0$, given in Anselin and Bera (1998). The spatial coefficients estimates are insignificant at the 5% level in 15 out of the 25 years used for estimation. They are significant in 1975, 1979, 1981, 1982, 1984, 1985, 1986, 1987, 1988 and 1989. The heterogeneous MLE estimates accounting for spatial autocorrelation do not differ much from the heterogeneous OLS estimates ignoring spatial autocorrelation. The price elasticity estimates varied from a low of -0.16 in 1983 to a high of -1.56 in 1989, while the income elasticity estimates varied from a low of 0.97 in 1989 to a high of 1.88 in 1971. The average pooled spatial heterogeneous MLE estimator yields a price elasticity estimate of -0.77 and an income elasticity estimate of 1.59 with a spatial autocorrelation parameter estimate of λ of 0.31, all of which are significant. These are reported in Table 1 as the average spatial maximum likelihood estimates. Note that these estimates are slightly higher than the average heterogeneous OLS estimates ignoring spatial autocorrelation.

Next, we account for heterogeneity across states by using the fixed effects (FE) estimator. This model assumes that the μ_i's are fixed parameters to be estimated. The F-statistic for testing the significance of the state dummies yields a value of 165.79 which is statistically significant. Note that if these state effects are ignored, the OLS estimates and their standard errors in Table 1 would be biased and inconsistent, see Moulton (1986).[4] Ignoring the spatial effects, the FE estimator can be obtained by running the regression with state dummy variables or by performing the within transformation and then running OLS, see Hsiao (1986). Denote these estimates by $\hat{\boldsymbol{\beta}}_{FE}$. These are reported in Table 1 as FE. Compared to the OLS estimates, the price elasticity estimate drops to -0.68 and the income elasticity estimate to 0.94 and both are significant.

This FE estimator does not take into account the spatial autocorrelation. This paper estimates the fixed effects with spatial autocorrelation using MLE.[5] In addition, we checked this global maximum using a grid search procedure over λ. [6] The estimates

[4]Note that prices vary across states mainly due to tax changes across states. To the extent that endogeneity in prices is due to its correlation with the state effects makes the fixed effects estimator a viable estimator which controls for endogeneity by wiping out the state effects.

[5]This was obtained using the Constrained Optimum (CO) module with GAUSS version 4.0.28.

[6]In fact, Figure 1 shows that the maximum likelihood function is well behaved for values of λ around

are reported in Table 1 as FE-Spatial. These results yield a much lower price elasticity estimate of -0.31 and a lower income elasticity estimate of 0.61 than the FE estimator. Both estimates are statistically significant. The λ estimate is 0.62. The likelihood ratio test for $\lambda = 0$, yields a χ_1^2 test statistic of 445.4. This is statistically significant and rejects the null of $\lambda = 0$ in the FE model.

For the random effects model, the μ_i's are $iid(0, \sigma_\mu^2)$ and are independent of the ϕ_{it}'s, see Anselin (1988). For this model, we need to derive the variance-covariance matrix. Let $\boldsymbol{B} = \boldsymbol{I}_N - \lambda \boldsymbol{W}$, then the disturbances in equation (3) can be written as follows: $\boldsymbol{\phi}_t = (\boldsymbol{I}_N - \lambda \boldsymbol{W})^{-1} \boldsymbol{\nu}_t = \boldsymbol{B}^{-1} \boldsymbol{\nu}_t$. Substituting $\boldsymbol{\phi}_t$ in (2), we get

$$\boldsymbol{\varepsilon} = (\boldsymbol{\iota}_T \otimes \boldsymbol{I}_N)\boldsymbol{\mu} + (\boldsymbol{I}_T \otimes \boldsymbol{B}^{-1})\nu \tag{4}$$

where $\boldsymbol{\iota}_T$ is a vector of ones of dimension T and $\boldsymbol{I}_N$ is an identity matrix of dimension N. The variance covariance matrix is

$$\boldsymbol{\Omega} = \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma_\mu^2(\boldsymbol{\iota}_T \boldsymbol{\iota}_T' \otimes \boldsymbol{I}_N) + \sigma_\nu^2(\boldsymbol{I}_T \otimes (\boldsymbol{B}'\boldsymbol{B})^{-1}) \tag{5}$$

In this case, GLS on (1) using $\boldsymbol{\Omega}^{-1}$ derived by Anselin (1988, p.154) yields $\hat{\boldsymbol{\beta}}_{GLS}$.

If $\lambda = 0$, so that there is no spatial autocorrelation, then $\boldsymbol{B} = \boldsymbol{I}_N$ and $\boldsymbol{\Omega}$ from (5) becomes the usual error component variance-covariance matrix

$$\boldsymbol{\Omega}_{RE} = \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma_\mu^2(\boldsymbol{\iota}_T \boldsymbol{\iota}_T' \otimes \boldsymbol{I}_N) + \sigma_\nu^2(\boldsymbol{I}_T \otimes \boldsymbol{I}_N) \tag{6}$$

Applying GLS using this $\boldsymbol{\Omega}_{RE}$ yields the random effects (RE) estimator which we will denote by $\hat{\boldsymbol{\beta}}_{RE}$. The one-sided Breusch and Pagan (1980) test for $\sigma_\mu^2 = 0$ yields a $N(0,1)$ test statistic of 97.3 which is statistically significant. Feasible GLS is based on Amemiya's (1971) method of estimating the variance components. This is an analysis of variance method that uses FE residuals in place of the true disturbances. The results are reported as RE in Table 1. In fact, the price elasticity estimate is -0.68 and the income elasticity estimate is 0.96 and both are significant. These RE estimates are close to those of the FE estimator. In fact, a Hausman (1978) test statistic for misspecification based on the difference between the FE and RE estimators of $\boldsymbol{\beta}$ yield a χ_3^2 test statistic of 3.36 which is statistically insignificant. The null hypothesis is not rejected and we conclude that the

the global maximum.

RE estimator is consistent.

If $\lambda \neq 0$, MLE under normality of the disturbances using this error component model with spatial autocorrelation is derived in Anselin (1988). Here we apply this MLE using the CO module of GAUSS version 4.0.28. In addition, we checked the global maximum by running a grid search procedure over λ and $\rho = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\nu^2)$. The latter is a positive fraction allowing a grid search over values of ρ between zero and one. [7] The results are reported in Table 1 as RE-Spatial. These results yield a much lower price elasticity estimate of -0.32 and a lower income elasticity estimate of 0.65 than the RE estimator. Both estimates are statistically significant. The λ estimate is 0.61 which is close to that of the FE-spatial model. The likelihood ratio test for $\lambda = 0$, yields a χ_1^2 test statistic of 423.1. This is statistically significant and rejects that $\lambda = 0$ in the RE model.

We now turn to comparing these various estimators using five years ahead forecasts. These are out of sample predictions for 1990, 1991, .., and 1994.

3 Prediction

Goldberger (1962) showed that, for a given $\mathbf{\Omega}$, the best linear unbiased predictor (BLUP) for the ith state at a future period $T+S$ is given by

$$\widehat{y}_{i,T+S} = \boldsymbol{x}'_{i,T+S}\hat{\boldsymbol{\beta}}_{GLS} + \boldsymbol{\omega}'\mathbf{\Omega}^{-1}\hat{\boldsymbol{\varepsilon}}_{GLS} \tag{7}$$

where $\boldsymbol{\omega} = \mathrm{E}(\varepsilon_{i,T+S}\boldsymbol{\varepsilon})$ is the covariance between the future disturbance $\varepsilon_{i,T+S}$ and the sample disturbances $\boldsymbol{\varepsilon}$. $\hat{\boldsymbol{\beta}}_{GLS}$ is the GLS estimator of $\boldsymbol{\beta}$ from (1) based on $\mathbf{\Omega}$, and $\hat{\boldsymbol{\varepsilon}}_{GLS}$ denotes the corresponding GLS residual vector.

For the error component model without spatial autocorrelation ($\lambda = 0$), Taub (1979) derived this BLUP and showed that it reduces to

$$\widehat{y}_{i,T+S} = \boldsymbol{x}'_{i,T+S}\hat{\boldsymbol{\beta}}_{GLS} + \frac{\sigma_\mu^2}{\sigma_1^2}(\boldsymbol{\iota}'_T \otimes \boldsymbol{l}'_i)\hat{\boldsymbol{\varepsilon}}_{GLS} \tag{8}$$

where $\sigma_1^2 = T\sigma_\mu^2 + \sigma_\nu^2$ and $\boldsymbol{l}_i$ is the ith column of $\boldsymbol{I}_N$. The typical element of the last term of (8) is $\frac{T\sigma_\mu^2}{\sigma_1^2}\bar{\varepsilon}_{i.,GLS}$ where $\bar{\varepsilon}_{i.,GLS} = \sum_{t=1}^{T}\hat{\varepsilon}_{ti,GLS}/T$. Therefore, the BLUP of $y_{i,T+S}$ for the

[7] Figure 2 shows that the maximum likelihood function is well behaved for values of λ and ρ around the global maximum.

RE model modifies the usual GLS forecasts by adding a fraction of the mean of the GLS residuals corresponding to the ith state. In order to make this forecast operational, $\hat{\boldsymbol{\beta}}_{GLS}$ is replaced by its feasible GLS estimate $\hat{\boldsymbol{\beta}}_{RE}$ reported in Table 1 and the variance components are replaced by their feasible estimates. The corresponding predictor is labeled the RE predictor in Table 3.

Baltagi and Li (1999) derived the BLUP correction term when both error components and spatial autocorrelation are present. In this case the predictor reduces to

$$\widehat{y}_{i,T+S} = \boldsymbol{x}'_{i,T+S}\hat{\boldsymbol{\beta}}_{GLS} + T\theta \sum_{j=1}^{N} \delta_j \bar{\varepsilon}_{j.,GLS} \tag{9}$$

where $\theta = \frac{\sigma_\mu^2}{\sigma_\nu^2}$, δ_j is the jth element of the ith row of $\boldsymbol{V}^{-1}$ with $\boldsymbol{V} = T\theta \boldsymbol{I}_N + (\boldsymbol{B}'\boldsymbol{B})^{-1}$and $\bar{\varepsilon}_{j.,GLS} = \sum_{t=1}^{T} \hat{\varepsilon}_{tj,GLS}/T$. In other words, the BLUP of $y_{i,T+S}$ adds to $\boldsymbol{x}'_{i,T+S}\hat{\boldsymbol{\beta}}_{GLS}$ a weighted average of the GLS residuals for the N regions averaged over time. The weights depend upon the spatial matrix $\boldsymbol{W}$ and the spatial autocorrelation coefficient λ. To make this predictor operational, we replace $\hat{\boldsymbol{\beta}}_{GLS}$, θ and λ by their estimates from the RE-spatial MLE reported in Table 1. The corresponding predictor is labeled RE-spatial in Table 3.

When there is no spatial autocorrelation, i.e., $\lambda = 0$, the BLUP correction term given in (9) reduces to the RE predictor term given in (8). Also, when there are no random state effects, so that $\sigma_\mu^2 = 0$, then $\theta = 0$ and the BLUP prediction term in (9) drops out completely from equation (7). In this case, $\boldsymbol{\Omega}$ in (5) reduces to $\sigma_\nu^2(\boldsymbol{I}_T \otimes (\boldsymbol{B}'\boldsymbol{B})^{-1})$ and GLS on this model, based on the MLE of λ, yields the pooled spatial estimator reported in Table 1. The corresponding predictor is labeled the pooled spatial predictor in Table 3.

If the fixed effects model without spatial autocorrelation is the true model, then the BLUP is given by

$$\tilde{y}_{i,T+S} = \boldsymbol{x}'_{i,T+S}\tilde{\boldsymbol{\beta}}_{FE} + \tilde{\mu}_i \tag{10}$$

see Baillie and Baltagi (1998), with μ_i estimated as $\tilde{\mu}_i = \bar{y}_{i.} - \bar{x}'_{i.}\hat{\boldsymbol{\beta}}_{FE}$ and $\bar{y}_{i.} = \sum_{t=1}^{T} y_{it}/T$ and $\bar{\boldsymbol{x}}_{i.}$ similarly defined. Note that in this case, $\lambda = 0$, so that ϕ_{it} in (3) reduces to ν_{it} and the latter are not serially correlated over time. Therefore, $\boldsymbol{\omega} = \mathrm{E}(\nu_{i,T+S}\boldsymbol{\nu}) = 0$, and the last term of (7) for the FE model is zero. However, the $\tilde{\mu}_i$ appear in the predictions as shown in (10). The corresponding predictor is labeled the FE predictor in Table 3.

If the fixed effects model with spatial autocorrelation is the true model, then the problem is to predict

$$y_{i,T+S} = \boldsymbol{x}'_{i,T+S}\boldsymbol{\beta} + \mu_i + \phi_{i,T+s} \tag{11}$$

with $\boldsymbol{\phi}_{T+S} = \lambda \boldsymbol{W}\boldsymbol{\phi}_{T+S} + \boldsymbol{v}_{T+s}$ obtained from (3). Unlike the previous case, $\lambda \neq 0$ and the μ_i's and $\boldsymbol{\beta}$ have to be estimated from MLE, i.e., using the FE-spatial estimates. The disturbance vector from (3) can be written as $\boldsymbol{\phi} = (\boldsymbol{I}_T \otimes \boldsymbol{B}^{-1})\boldsymbol{v}$, so that $\boldsymbol{\omega} = \mathrm{E}(\phi_{i,T+S}\boldsymbol{\phi}) = 0$ since the $\boldsymbol{v}$'s are not serially correlated over time. So the BLUP for this model looks like that for the FE model without spatial correlation given in (10) except that the μ_i's and $\boldsymbol{\beta}$ are estimated assuming $\lambda \neq 0$. The corresponding predictor is labeled the FE-spatial predictor in Table 3.

Table 3 gives the RMSE for the one year, two year,.., and five year ahead forecasts along with the RMSE for all 5 years. These are out of sample forecasts from 1990 to 1994. Each year's RMSE is obtained from 43 state by state predictions. We compare the forecasts for all 5 years. The pooled OLS predictor in Table 3 is computed as $\widehat{y}_{i,T+S} = \boldsymbol{x}'_{i,T+S}\hat{\boldsymbol{\beta}}_{OLS}$. Pooled OLS, which ignores spatial autocorrelation and heterogeneity across the states gives a RMSE over the 5 years of 0.2590. Accounting for spatial autocorrelation using the pooled spatial estimator increases this RMSE to 0.2678. This predictor replaces the OLS estimator of $\boldsymbol{\beta}$ by that of pooled spatial MLE reported in Table 1. Substituting the average heterogeneous OLS estimator (which ignores spatial autocorrelation but allows for parameter heterogeneity across time) more than triples the RMSE of pooled OLS yielding a RMSE to 0.8781. This forecast performance is not improved by accounting for spatial autocorrelation. Substituting the average heterogeneous spatial MLE yields a RMSE of 0.9678. Parameter heterogeneity is costly in terms of RMSE forecast performance and is beaten by estimators that rely on parameter homogeneity. A substantial improvement in RMSE forecast performance occurs when one uses a simple FE or RE estimators. In fact, the simple FE estimator without spatial autocorrelation yields a RMSE of 0.1360 followed closely by the RE estimator without spatial autocorrelation with a RMSE of 0.1367. These predictors were described in (10) and (8), respectively. Taking into account both heterogeneity and spatial autocorrelation, the best forecast performance for one year ahead is obtained by the RE estimator with spatial autocorrelation which yields a RMSE of 0.1207, followed closely by the FE with spatial autocorrelation estimator with a RMSE of 0.1213. The FE-spatial predictor is obtained as in (10) but with the FE-spatial estimates from Table 1 replacing the FE estimates. The RE-spatial predictor is obtained from (9) by substituting the RE-spatial estimates from Table 1. For two or more years ahead forecasts, the FE estimator without spatial correlation performs the best followed by the RE estimator.

In sum, for the simple liquor demand model chosen to illustrate our forecasts, taking into account the heterogeneity across states by a FE or RE estimators yields the best out of sample RMSE forecast performance. The FE estimator gives the lowest RMSE for 1991, 1992, 1993 and 1994 and is only surpassed by the RE-spatial estimator in the first year, 1990. Overall, both the FE and RE estimators perform well in predicting liquor demand. Adding the spatial correlation in the model does not improve prediction except for the first year. However, as we show next, the difference in forecast accuracy between FE and FE-spatial or RE-spatial is not significant.

To compare the out-of-sample forecast performance, we conduct the Diebold and Mariano (1995) test on the eight forecasts considered in this paper. The null hypothesis is that there is no difference in forecast accuracy of the two competing forecasts. Briefly, this test considers two forecast errors series, $\{\widehat{e}_t\}$ and $\{\widetilde{e}_t\}$, based on two competing methods. Suppose the loss function is $g(\cdot)$. In our case, $g(e_t) = e_t^2$. Diebold and Mariano (1995) proposed an asymptotic test

$$S = \frac{\bar{d}}{\sqrt{\mathrm{Var}(\bar{d})}} \tag{12}$$

where $\bar{d} = \frac{1}{T}\sum_{t=1}^{T}[g(\widehat{e}_t) - g(\widetilde{e}_t)]$ and $\mathrm{Var}(\bar{d})$ is the Hubert/White robust variance of the numerator. The test statistic $S \sim N(0,1)$. See Diebold and Mariano (1995) and West (2005) for more details. We conduct this test for pair-wise comparisons based on the eight forecasts. The results for the five year forecast averaged over the 43 states are reported in Table 4. For example, the FE model is significantly better than the pooled OLS, the pooled spatial, the average heterogeneous OLS, the average spatial MLE, and the RE models in terms of out-of-sample forecast performance. However, the difference in forecast accuracy between FE and FE-spatial or RE-spatial is insignificant. These results are in agreement with the findings in the last column of Table 3. We also conduct the Diebold-Mariano test on a year by year basis. The results are consistent with the first five columns in Table 3 and are available upon request from the authors. We add the caveat that comparisons of forecast accuracy is but one of many diagnostics that should be considered, and that the superiority of a particular forecast does not necessarily mean that other forecasts contain no additional information, see Diebold and Mariano (1995).

Some of the limitations of our study is that we used a simple static model of liquor demand when a dynamic liquor demand may be more appropriate. However, the latter model introduce additional econometric complications for our forecasting illustrations and these are beyond the scope of this paper. Despite these limitations, this paper lays out a simple methodology for forecasting with panel data models that are spatially autocorrelated.

REFERENCES

Amemiya, T. (1971), The estimation of the variances in a variance components model, *International Economic Reviews,* 12, 1-13.

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

Anselin, L. and A. Bera (1998), Spatial dependence in linear regression models with an introduction to spatial econometrics, in the *Handbook of Applied Economic Statistics*, A. Ullah and D. Giles, eds., Marcel Dekker, New York, pp.237-289.

Baillie, R. and B. Baltagi (1998), Prediction from the regression model with one-way error components, in *Analysis of Panel Data and Limited Dependent Variable Models*, H. Pesaran, K. Lahiri, C. Hsiao and L-F. Lee, eds.,Cambridge University Press, Cambridge.

Baltagi, B.H. and J. M. Griffin (1995), A dynamic demand model for liquor: the case for pooling, *Review of Economics and Statistics*, 77, 545-554.

Baltagi, B.H. and D. Li (1999), Prediction in the spatially autocorrelated error component model, *Econometric Theory,* problem *99.2.4, 15, 259.*

Breusch, T.S. and A. Pagan, 1980, The Lagrange multiplier tests and its applications to model specification in econometrics, *Review of Economic Studies*, 47, 239-253.

Carlson, E. (1985), New Hampshire tries harder, marketing liquor with spirit, *Wall Street Journal*, 8, 31.

Diebold, F.X. and R.S. Mariano (1995), Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 253-263.

Goldberger, A.S. (1962) Best linear unbiased prediction in the generalized linear regression model, *Journal of the American Statistical Association* 57, 369–375.

Hausman, J.A. (1978), Specification tests in econometrics, *Econometrica*, 46, 1251-1271.

Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge, Cambridge University Press.

Kenkel, D. S. (1993), Drinking, driving, and deterrence: the effectiveness and social costs of alternative policies, *Journal of Law and Economics*, 36, 877-913.

Moulton, B. R. (1986), Random group effects and the precision of regression estimates, *Journal of Econometrics*, 32, 385-397.

Pesaran, M.H. and R. Smith (1995), Estimating long-run relationships from dynamic heterogeneous panels, *Journal of Econometrics*, 68, 79-113.

Simon, J. L. (1966), The economic effects of state monopoly of packaged- liquor retailing, *Journal of Political Economy*, 74, 188-194.

Taub, A.J. (1979), Prediction in the context of the variance components model, *Journal of Econometrics*, 10, 103-107.

Wansbeek, T. and A. Kapteyn (1983), A note on spectral decomposition and maximum likelihood estimation of ANOVA models with balanced data, *Statistics and Probability Letters*, 1, 213-215.

West, K.D. (2005), Forecasting evaluation, Handbook of Economic Forecasting (Edited by G. Elliot, C.W.J. Granger, and A. Timmermann), Elsevier.

Zardkoohi, A. and A. Sheer (1984), Public versus private liquor retailing: an investigation into the behavior of the state governments, *Southern Economic Journal*, 50, 1058-1076.

Table 1: Estimates of Liquor Demand
(Based on 25 years, standard errors in parentheses)

	Price	Income	Year
Pooled OLS	−0.774 (0.088)	1.468 (0.065)	−0.062 (0.004)
Pooled Spatial	−0.819 (0.093)	1.605 (0.070)	−0.067 (0.004)
Average Heterogeneous OLS	−0.584 (0.064)	1.451 (0.041)	
Average Spatial MLE	−0.766 (0.062)	1.589 (0.044)	
FE	−0.679 (0.044)	0.938 (0.063)	−0.049 (0.002)
FE-Spatial	−0.314 (0.044)	0.612 (0.075)	−0.029 (0.002)
RE	−0.682 (0.044)	0.959 (0.062)	−0.049 (0.002)
RE-Spatial	−0.317 (0.045)	0.654 (0.075)	−0.030 (0.002)

† The numbers in parentheses are standard errors.
* The F test for $H_0; \mu = 0$ in FE model is $F(42, 1029) = 165.79$ with $p = 0.000$.
** The Breusch-Pagan test for $H_0; \sigma_\mu^2 = 0$ in RE model is 97.30 with $p = 0.000$.
*** The Hausman test based on FE and RE yields χ_3^2 of 3.36 with $p = 0.339$.

Table 2: Heterogeneous Estimates of Liquor Demand

	Heterogeneous OLS		Heterogeneous Spatial			LM^*
	Price	Income	Price	Income	λ	
1965	−0.354	1.454	−0.371	1.559	0.238	1.437
	(0.524)	(0.286)	(0.496)	(0.324)	(0.195)	(0.231)
1966	−0.299	1.447	−0.393	1.551	0.260	1.893
	(0.523)	(0.291)	(0.499)	(0.325)	(0.189)	(0.169)
1967	−0.405	1.543	−0.487	1.642	0.251	1.679
	(0.506)	(0.294)	(0.480)	(0.325)	(0.192)	(0.195)
1968	−0.489	1.633	−0.657	1.749	0.314	2.724
	(0.507)	(0.291)	(0.484)	(0.320)	(0.185)	(0.099)
1969	−0.367	1.640	−0.443	1.740	0.244	1.681
	(0.533)	(0.312)	(0.505)	(0.340)	(0.189)	(0.195)
1970	−0.992	1.664	−1.084	1.752	0.245	1.889
	(0.569)	(0.319)	(0.553)	(0.342)	(0.183)	(0.169)
1971	−1.306	1.749	−1.424	1.884	0.264	2.297
	(0.584)	(0.331)	(0.577)	(0.364)	(0.182)	(0.130)
1972	−1.230	1.693	−1.434	1.854	0.298	3.111
	(0.596)	(0.380)	(0.613)	(0.413)	(0.177)	(0.078)
1973	−0.933	1.419	−1.171	1.578	0.308	3.730
	(0.590)	(0.390)	(0.578)	(0.417)	(0.169)	(0.053)
1974	−1.054	1.576	−1.263	1.723	0.304	3.419
	(0.571)	(0.400)	(0.570)	(0.423)	(0.172)	(0.064)
1975	−1.098	1.528	−1.556	1.843	0.392	5.006
	(0.605)	(0.418)	(0.599)	(0.454)	(0.169)	(0.025)
1976	−0.081	1.401	−0.459	1.640	0.267	2.314
	(0.704)	(0.459)	(0.731)	(0.504)	(0.178)	(0.128)
1977	0.091	1.329	−0.313	1.603	0.280	2.347
	(0.680)	(0.443)	(0.706)	(0.499)	(0.181)	(0.126)
1978	0.047	1.305	−0.244	1.587	0.281	2.417
	(0.622)	(0.437)	(0.604)	(0.488)	(0.179)	(0.120)
1979	−0.424	1.374	−0.718	1.650	0.364	5.132
	(0.605)	(0.397)	(0.531)	(0.426)	(0.162)	(0.023)
1980	−0.313	1.225	−0.498	1.392	0.287	3.042
	(0.614)	(0.372)	(0.564)	(0.390)	(0.170)	(0.081)
1981	−0.271	1.383	−0.442	1.528	0.336	4.435
	(0.515)	(0.362)	(0.456)	(0.364)	(0.163)	(0.035)
1982	−0.294	1.415	−0.377	1.553	0.334	4.660
	(0.476)	(0.355)	(0.424)	(0.355)	(0.162)	(0.031)
1983	−0.161	1.471	−0.157	1.586	0.302	3.542
	(0.504)	(0.351)	(0.468)	(0.353)	(0.169)	(0.060)
1984	−0.617	1.463	−0.713	1.548	0.354	5.593
	(0.531)	(0.361)	(0.493)	(0.355)	(0.157)	(0.018)
1985	−0.698	1.443	−0.884	1.497	0.337	4.764
	(0.540)	(0.341)	(0.511)	(0.338)	(0.160)	(0.029)
1986	−0.273	1.476	−0.526	1.517	0.366	5.214
	(0.522)	(0.314)	(0.508)	(0.311)	(0.158)	(0.022)
1987	−0.478	1.557	−0.711	1.621	0.358	4.744
	(0.487)	(0.278)	(0.469)	(0.279)	(0.162)	(0.029)
1988	−1.178	1.127	−1.255	1.153	0.344	4.487
	(0.446	(0.255	(0.424	(0.266	(0.165)	(0.034)
1989	−1.409	0.966	−1.564	0.967	0.477	9.732
	(0.367)	(0.253)	(0.320)	(0.258)	(0.145)	(0.002)

*This gives the LM statistic for $H_0; \lambda = 0$ and the corresponding p-value in parenthesis.

Table 3: RMSE Performance of Out-of-Sample Forecasts
(Estimation sample of 25 years. Prediction sample of 5 years)

	1990	1991	1992	1993	1994	5 Years
Pooled OLS	0.2485	0.2520	0.2553	0.2705	0.2678	0.2590
Pooled Spatial	0.2548	0.2594	0.2638	0.2816	0.2783	0.2678
Average Heterogeneous OLS	0.7701	0.8368	0.8797	0.9210	0.9680	0.8781
Average Spatial MLE	0.8516	0.9237	0.9715	1.0142	1.0640	0.9678
FE	0.1232	0.1351	0.1362	0.1486	0.1359	0.1360
FE-Spatial	0.1213	0.1532	0.1529	0.1655	0.1605	0.1515
RE	0.1239	0.1356	0.1368	0.1493	0.1366	0.1367
RE-Spatial	0.1207	0.1517	0.1513	0.1633	0.1581	0.1497

Table 4: Diebold-Mariano Test

	1	2	3	4	5	6	7	8
(1) Pooled OLS								
(2) Pooled Spatial	2.667							
(3) Average Heterogeneous OLS	10.209	10.009						
(4) Average Spatial MLE	11.170	10.978	15.176					
(5) FE	-2.741	-2.984	-12.651	-13.490				
(6) FE-Spatial	-2.535	-2.766	-12.751	-13.626	1.076			
(7) RE	-2.729	-2.972	-12.645	-13.484	2.606	-1.004		
(8) RE-Spatial	-2.562	-2.792	-12.765	-13.638	0.971	-2.629	0.899	

The test statistic follows a standard normal distribution asymptotically.

Figure 1. MLE for FE-Spatial

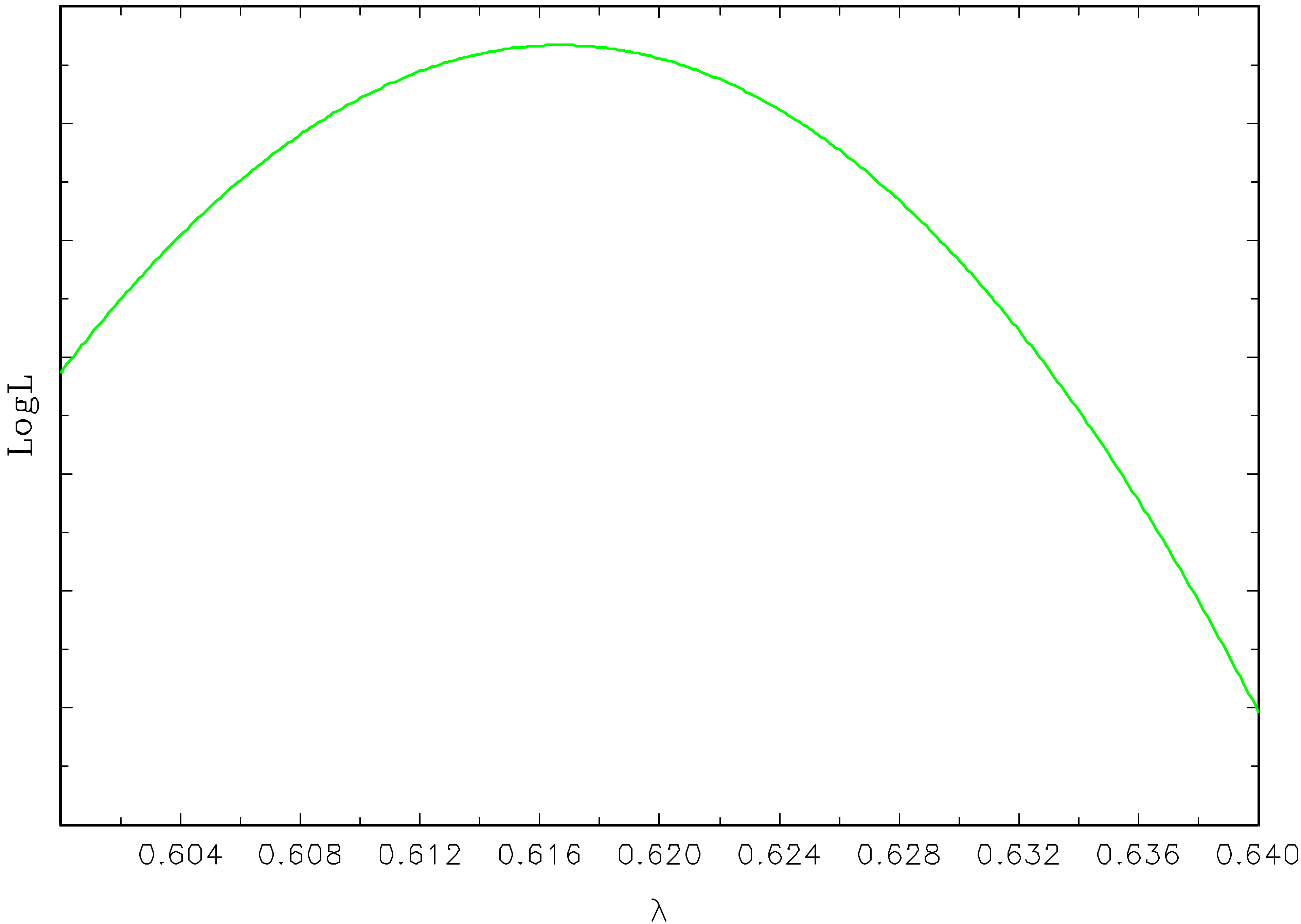

Figure 2. MLE for RE-Spatial