NHH

INSTITUTT FOR SAMFUNNSØKONOMI

DEPARTMENT OF ECONOMICS

Discussion paper

After "Raising the Bar": applied maximum likelihood estimation of families of models in spatial econometrics

BY
Roger Bivand

This series consists of papers with limited circulation, intended to stimulate discussion.

Norges Handelshøyskole

NORWEGIAN SCHOOL OF ECONOMICS

After "Raising the Bar": applied maximum likelihood estimation of families of models in spatial econometrics[*]

Roger Bivand[†]

October 2011

Abstract

Elhorst (2010) shows how the recent publication of LeSage and Pace (2009) in his expression "raises the bar" for our fitting of spatial econometrics models. By extending the family of models that deserve attention, Elhorst reveals the need to explore how they might be fitted, and discusses some alternatives. This paper attempts to take up this challenge with respect to implementation in the R **spdep** package for the maximum likelihood case, using a smaller data set to see whether earlier conclusions would be changed when newer techniques are used, and two larger data sets to examine model fitting issues.

1   Background

In an interesting review, Elhorst (2010) "raises the bar" to place the general spatial autoregressive model and the Spatial Durbin model in a shared context, such as that of the model proposed by Manski (1993). Fitting these models by maximum likelihood makes it possible to start trying to investigate which augmented forms of commonly used spatial econometric models may be of use in empirical work.

It is not intended to make reference here to the burgeoning spatial econometrics literature on the properties of estimators, including maximum likelihood estimators. We will also focus exclusively on maximum likelihood estimators, although

---

[†]Department of Economics, NHH Norwegian School of Economics, Helleveien 30, N-5045 Bergen, Norway; E-mail: `Roger.Bivand@nhh.no`

extensions to Bayesian estimators are clearly of interest (LeSage and Pace, 2009). Similarly, we will not consider GM estimators, but acknowledge that the sets of instruments that may be used in fitting the general two parameter spatial autoregressive model are not limited to low-order spatial lags of the explanatory variables (Kelejian and Prucha, 1999; Drukker et al., 2011, and references therein).

This analysis extends work presented in Bivand (2010a), which included comparisons of model fitting for spatial error, spatial lag between implementations in R, OpenGeoDa running under Wine, and these and spatial Durbin models with the Spatial Econometrics toolbox running under Octave.

In Bivand (2010a), two data sets distributed with the R **spdep** package are used (R Development Core Team, 2011); both originated from the Spatial Econometrics toolbox, and are provided in **spdep** with pre-built lists of spatial neighbours. Here, a third, smaller data set is used for convenience first, permitting comparison with spatial econometrics functions in `sppack` in Stata$^{\text{TM}}$ 11.2 on a smaller system running Windows XP with 1GB RAM. Its use also permits us to examine whether legacy results require revision when confronted with newer estimation techniques. A broad survey of the analysis of spatial data in the R environment is given by Bivand (2006) and Bivand et al. (2008).

1.1   Garbage in — garbage out?

The smaller data set is used to revisit conclusions drawn in Bivand and Szymanski (2000), and concerns a study of the impact of introducing compulsory competitive tendering in garbage collection in England at the district level. The larger data sets could not be fitted with Stata on the platform available, as Stata uses dense matrix techniques for maximum likelihood fitting.

The underlying research question in Bivand and Szymanski (2000) was firstly whether the change in policy regime had affected the net real costs of garbage collection in local government districts, and secondly whether the fit of models including standard explanatory variables had changed with respect to those variables. The net real cost of garbage collection varies with the number of collection points (units), the proportion of dwelling house units, the density of units by district surface area, dummies for London and other metropolitan areas, and the real wage level in the region to which the district belongs. The introduction of compulsory competitive tendering would be expected not only to lower net real costs, but also to sharpen the impact of cost-shaping variables, such as density and wages. The data for 324 out of 366 districts was split into two sets, one with the values of real net cost and real wages for each district in the pre-CCT year, and a second in the post-CCT year, with the actual year of adoption dropped.

After hearing a presentation of some preliminary results of his aspatial study (Szymanski, 1996), I asked Stefan Szymanski whether he had tested the residu-

als for spatial autocorrelation as a standard specification check. He asked me to collaborate in carrying this out, and we found that the residuals were significantly spatially autocorrelated. The results of the robust Lagrange Multiplier tests seemed to indicate that a spatial error model is preferable to a spatial lag model in both cases.

We spent some time discussing why the spatial autocorrelation was observed, and reached a working hypothesis that before the introduction of CCT, the district principals might "mimic" the costings of their near neighbours because they had few other sources of information about the costs of garbage collection, and that this "mimicking" should abate following the introduction of CCT. This principal-agent model and a first cut at a spatial analysis are reported in Bivand and Szymanski (1997), fitting a spatial error model.

The principal-agent model would have been better suited by a spatial lag model, with the observed cost levels of proximate neighbours influencing the principals' decisions directly, but the model diagnostics seemed to suggest otherwise. Continuing to work on this contradiction, we noticed that the kinds of yardstick comparisons that might be occurring should probably be politically "coloured", and decided to include party political control in districts in the analysis, as described in Bivand and Szymanski (2000). The neighbours used here are those from the original paper, defined as a graph on all 366 districts, and subsetted to remove missing districts.

1.2 US 1980 election turnout data set

The US county data set with 3107 observations includes a 1980 Presidential election turnout variable with a single county (Hinsdale County, CO) with a value over unity — most likely from cross-border voting in this remote rural area. We define a formula relating this variable to income ($1000) per inhabitant over age 19, the number with college degrees as a proportion of all over age 19, and homeownership as a proportion of all over age 19. The right hand side variables are taken as logarithms, as in the file `data/elect.txt` in the Spatial Econometrics toolbox.

The data set provided in **spdep** includes a number of `nb` objects listing the neighbours of the counties in the data set using different definitions. Here we will use a Queen contiguity scheme constructed using a shapefile from the USGS National Atlas site, file: `co1980p020.tar.gz`. This object contains four counties with no neighbours, and because of this, an option is set to permit computations under the assumption that the lagged value of a variable for a county with no neighbours may be set to zero (Bivand and Portnov, 2004).

1.3 Lucas County, OH, housing data set

The Lucas County, Ohio, housing data set has 25,357 observations of single family homes sold 1993–1998, and is fully described in the file `data/house.txt` in the Spatial Econometrics toolbox. It is used here to supplement conclusions drawn for the 1980 US election turnout data set, which is of a size that permits dense matrix methods, since only sparse or approximate methods are feasible for larger N. The dependent variable is the logarithm of the selling price. The right hand side variables include the age, squared age, and cubed age of the house, sale year dummies, the logarirms of lot size and total living area, and numbers of rooms and bedrooms. No contextual variables about the neighbourhood of the houses are available, so one would expect a strong spatial autocorrelation reflecting this misspecification.

The list of neighbours provided with the data set in **spdep** is a sphere of influence graph constructed from a triangulation of the point coordinates of the houses after projection to the Ohio North NAD83 (HARN) Lambert Conformal Conical specification (EPSG:2834). It is relatively sparse, with less than three neighbours per observation on average.

2 Candidate models

The spatial lag model (Cliff and Ord, 1973; Ord, 1975; Bivand, 1984; Anselin, 1988; LeSage and Pace, 2009) is the most frequently encountered specification in spatial econometrics:

$$\mathbf{y} = \rho\mathbf{Wy} + \mathbf{X}\beta + \varepsilon,$$

where $\mathbf{y}$ is an $(N \times 1)$ vector of observations on a dependent variable taken at each of N locations, $\mathbf{X}$ is an $(N \times k)$ matrix of exogenous variables, β is an $(k \times 1)$ vector of parameters, ε is an $(N \times 1)$ vector of independent and identically distributed disturbances and ρ is a scalar spatial lag parameter.

In the spatial Durbin model, the spatially lagged exogenous variables are added to the model:

$$\mathbf{y} = \rho\mathbf{Wy} + \mathbf{X}\beta + \mathbf{WX}\gamma + \varepsilon,$$

where γ is an $((k-1) \times 1)$ vector of parameters where $\mathbf{W}$ is row-standardised, and a $(k \times 1)$ vector otherwise. It is clear that these two models are estimated in the same way.

The spatial error model may be written as (Cliff and Ord, 1973; Ord, 1975; Ripley, 1981; Anselin, 1988; LeSage and Pace, 2009):

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}, \qquad \mathbf{u} = \lambda\mathbf{Wu} + \varepsilon,$$

where λ is a scalar spatial error parameter, and $\mathbf{u}$ is a spatially autocorrelated disturbance vector with constant variance and covariance terms specified by a fixed spatial weights matrix and a single coefficient λ:

$$\mathbf{u} \sim N(0, \sigma^2(\mathbf{I} - \lambda\mathbf{W})^{-1}(\mathbf{I} - \lambda\mathbf{W}')^{-1})$$

When the Common Factor condition is met: $\beta = -\rho\gamma$, the spatial Durbin and spatial error models are equivalent. Note that here we use the notation of Anselin (1988); LeSage and Pace (2009), with ρ the spatial autoregressive coefficient of the dependent variable, and λ the spatial autoregressive coefficient of the disturbance; the usage is reversed in other parts of the spatial econometrics literature.

The general two parameter spatial Durbin autoregressive model, includes the spatially lagged dependent variable, spatially lagged explanatory variables, and a spatially autoregressive disturbance. It has been variously termed as the Manski, SARAR Durbin or SAC Durbin, with the latter two terms extending the SARAR description used by Kelejian and Prucha (1999, and subsequent papers), and the SAC term used as a function name by LeSage and Pace (2009). The SARAR and SAC terms refer to the general model described by Anselin (1988, pp. 64–65, 182–183), with two spatial process parameters, but no spatially lagged explanatory variables. Here we will use the term SAC Durbin; the model may be written (assuming the use of the same weight matrix $\mathbf{W}$ in all spatial processes):

$$\mathbf{y} = \rho\mathbf{Wy} + \mathbf{X}\beta + \mathbf{WX}\gamma + \mathbf{u}, \qquad \mathbf{u} = \lambda\mathbf{Wu} + \varepsilon,$$

This representation forks to the general model (SARAR/SAC) by setting $\gamma = 0$, to the spatial Durbin by setting $\lambda = 0$, and to the error Durbin model by setting $\rho = 0$ (the error Durbin model includes the spatially lagged explanatory variables and a spatial autoregressive error process). In the general model case, when the weights matrices in both processes are the same, the identification of ρ and λ depends on $\beta \neq 0$.

Elhorst (2010) suggests that it may be appropriate to fit a general, inclusive model first, here the SAC Durbin model, and to test restrictions on that model. Some tests are available for simpler comparisons between models using ordinary least squares residuals, but so far none are defined for the more complex models. Consequently, after fitting the pairs of models to be compared by maximum likelihood, it is possible to use likelihood ratio tests, and this approach will be followed here. If one wished to accommodate situations in which the assumptions required for use of maximum likelihood were not met, Bayesian model comparison would be a possible alternative. It is not yet clear whether a J-test approach could be used for GMM-fitted models (Kelejian, 2008; Kelejian and Piras, 2011).

3   Maximum likelihood estimation

The log-likelihood function for the spatial lag model is:

$$\ell(\beta, \rho, \sigma^2) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma^2 + \ln|\mathbf{I} - \rho\mathbf{W}|$$
$$-\frac{1}{2\sigma^2}\left[((\mathbf{I} - \rho\mathbf{W})\mathbf{y} - \mathbf{X}\beta)'((\mathbf{I} - \rho\mathbf{W})\mathbf{y} - \mathbf{X}\beta)\right]$$

and by extension the same framework is used for the spatial Durbin model when $[\mathbf{X}(\mathbf{WX})]$ are grouped together. Since β can be expressed as $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \rho\mathbf{W})\mathbf{y}$, all of the cross-product terms can be pre-computed as cross-products of the residuals of two ancilliary regressions: $\mathbf{y} = \mathbf{X}\beta_1$ and $\mathbf{Wy} = \mathbf{X}\beta_2$, and the sum of squares term can be calculated much faster than the log determinant (Jacobian) term of the $N \times N$ sparse matrix $\mathbf{I} - \rho\mathbf{W}$; see LeSage and Pace (2009) for details.

The log-likelihood function for the spatial error model is:

$$\ell(\beta, \lambda, \sigma^2) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma^2 + \ln|\mathbf{I} - \lambda\mathbf{W}|$$
$$-\frac{1}{2\sigma^2}\left[(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{I} - \lambda\mathbf{W})'(\mathbf{I} - \lambda\mathbf{W})(\mathbf{y} - \mathbf{X}\beta)\right]$$

β may be concentrated out of the sum of squared errors term, for example as:

$$\ell(\lambda, \sigma^2) = -\frac{N}{2}\ln 2\pi - \frac{N}{2}\ln\sigma^2 + \ln|\mathbf{I} - \lambda\mathbf{W}|$$
$$-\frac{1}{2\sigma^2}\left[\mathbf{y}'(\mathbf{I} - \lambda\mathbf{W})'(\mathbf{I} - \mathbf{Q}_\lambda\mathbf{Q}'_\lambda)(\mathbf{I} - \lambda\mathbf{W})\mathbf{y}\right]$$

where $\mathbf{Q}_\lambda$ is obtained by decomposing $(\mathbf{X} - \lambda\mathbf{WX}) = \mathbf{Q}_\lambda\mathbf{R}_\lambda$.

The relationship between the log determinant term and the sum of squares term in the log likelihood function in the spatial error model is analogous to that in the spatial lag model, but the sum of squares term involves more computation in the case of the spatial error model. In all cases, a simple line search may be used to find ρ or λ, and other coefficients may be calculated using an ancilliary regression once this has been done.

The general model is more demanding, and requires that ρ and λ be found by constrained numerical optimization in two dimensions by searching for the maximum on the surface of the log likelihood function, which is like that of the spatial error model with additional terms in $\mathbf{I} - \rho\mathbf{W}$:

$$\ell(\rho, \lambda, \sigma^2) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma^2 + \ln |\mathbf{I} - \rho\mathbf{W}| + \ln |\mathbf{I} - \lambda\mathbf{W}|$$
$$-\frac{1}{2\sigma^2} \left[\mathbf{y}'(\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \lambda\mathbf{W})'(\mathbf{I} - \mathbf{Q}_\lambda\mathbf{Q}'_\lambda)(\mathbf{I} - \lambda\mathbf{W})(\mathbf{I} - \rho\mathbf{W})\mathbf{y} \right]$$

This suggests that the tuning of the constrained numerical optimization function, including the provision of starting values, reasonable stopping criteria, and also the choice of algorithm may all affect the results achieved. The Stata implementation uses a grid search for initial values of (ρ, λ) (Drukker et al., 2011), the Spatial Econometrics toolbox uses the generalized spatial two-stage least squares estimates, with the option of the user providing initial values, and the **spdep** implementation for row-standardised spatial weights matrices, uses either four candidate pairs of initial values at $(-0.8, 0.8)$, $(0,0)$, $(0.8, 0.8)$ and $(0.8, -0.8)$, a full grid of nine points at the same settings, or user provided initial values (which permits the use of weights matrices that are not row standardised); optimizers may be chosen by the user.

Detailed reviews of methods for computing the Jacobian may be found in LeSage and Pace (2009); Smirnov and Anselin (2009); Bivand (2010b), and interested readers are refered to these. The methods used for computing the Jacobian in **spdep** are presented in full in Bivand (2010b); here we use the dense matrix eigenvalue method `eigen` (Ord, 1975, p. 121) for the English garbage data set, and the updating Cholesky decomposition method `Matrix`, using sparse matrix functions in the R **Matrix** package (Bates and Maechler, 2011), and based on Pace and Barry (1997), for the two larger data sets.

When sparse matrix methods or approximations are used, motivated by the size of N, no analytical asymptotic standard errors for the coefficients in spatial lag, Durbin or general SARAR models will be available, nor will the standard error of λ be available in the spatial error case. This may be addressed by computing a numerical Hessian for an augmented function fitting both ρ and/or λ and β starting at the maximum likelihood optimum. The covariance matrix of coefficient estimates is required for the Monte Carlo testing of measures of the impacts of explanatory variables, as we will see below.

With some data sets, models, and variable scaling — fortunately not those used in these examples, one meets difficulties in inverting the numerical Hessian returned from finite difference computation. This unfortunate problem may be worked around by replacing most of the matrix with analytical values, termed the analytical-numerical mixed Hessian by LeSage and Pace (2009, pp. 54–60). The awkward trace term for the interaction between λ and σ^2 — $tr(\mathbf{W}(\mathbf{I} - \tilde{\lambda}\mathbf{W})^{-1})$ — may be approximated by a series of traces of the powered weights matrix, either computed using sparse matrix or Monte Carlo techniques. The analytical-numerical

mixed Hessian is available in **spdep** for the spatial lag, Durbin, and error models, but not yet for the SAC Durbin model.

4 Fitting models using maximum likelihood for the English data set

The `sacsarlm` function has been added to **spdep** to permit the fitting of the SAC model, and it takes a `type=` argument to add the spatially lagged right hand side variables to make a SAC Durbin specification. We fit both of these model forms to the augmented pre-CCT and post-CCT models, which take political control into account. Table 1 does not report the coefficient values, because we should more properly report the impacts (emanating effects) of the right hand variables

Table 1: Summary of SAC and SAC Durbin model output (asymptotic standard errors in parentheses).

	pre-CCT	pre-CCT	post-CCT	post-CCT
Model	SAC	SAC Durbin	SAC	SAC Durbin
ρ	0.1006	0.4866	0.05381	0.3634
	(0.05271)	(0.1314)	(0.05521)	(0.2075)
λ	0.1672	-0.4418	0.1524	-0.2728
	(0.0931)	(0.187)	(0.09458)	(0.2639)
Log likelihood	22.78	41.06	37.62	52.45
σ^2	0.05044	0.04116	0.04615	0.04043
AIC	-21.57	-42.12	-51.24	-64.9

As we can see, the SAC Durbin model outperforms the SAC model in both cases, suggesting that the inclusion of the spatially lagged right hand side variables was justified. Before testing against other alternatives, let us examine the log-likelihood function surfaces shown in Figure 1. The values of the spatial coefficients and the optimal log likelihood function values fitted using `sacsarlm` in **spdep** and using `spreg ml` in Stata were identical: pre-CCT SAC $\rho = 0.1006, \lambda = 0.1672$, pre-CCT SAC Durbin $\rho = 0.4866, \lambda = -0.4418$, post-CCT SAC $\rho = 0.0538, \lambda = 0.1524$, post-CCT SAC Durbin $\rho = 0.3634, \lambda = -0.2728$. Since both use eigenvalues to compute the Jacobian, this is as expected; there are small differences in coefficient standard errors.

The surfaces shown in Figure 1 thus represent the optimization as computed using R and Stata. We see that while the SAC surfaces, shown in grey, are quite
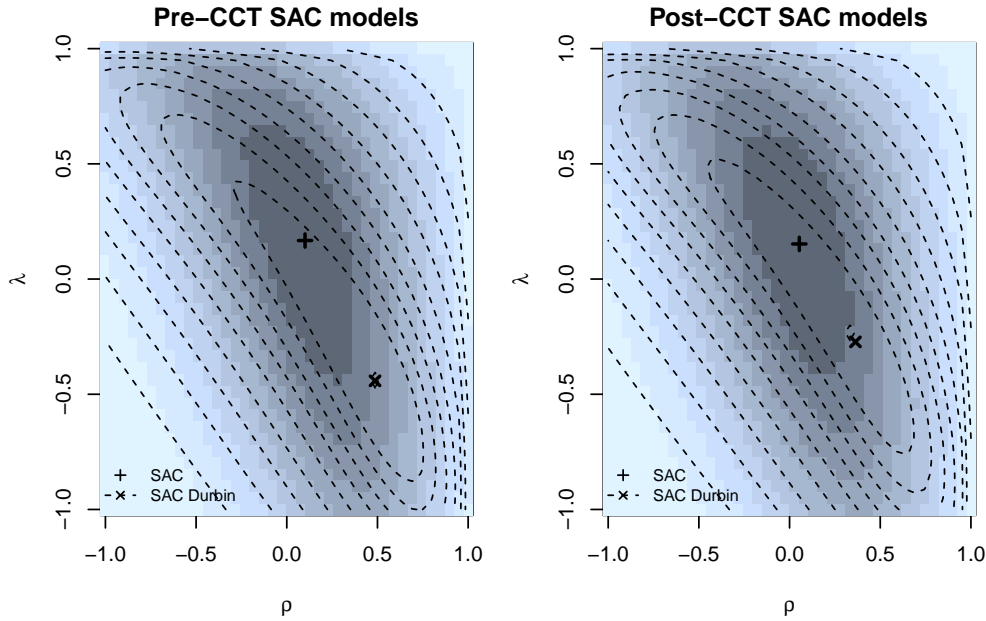
Figure 1: Log-likelihood function surfaces for the pre-CCT and post-CCT models, for SAC and SAC Durbin specifications.

amenable to numerical optimization, the SAC Durbin surfaces, shown with contours, both show a "banana" ridge running from low ρ, high λ, through moderate/high ρ, moderate/high λ, to high ρ, low λ. This appears to be the visual expression of the difficulty of indentifiction between the two coefficients noted by Elhorst (2010) among others. Moving on to test fitted model specifications, we also fit the spatial Durbin, spatial lag, spatial error Durbin, and spatial error models.

Table 2 shows that the results of the likelihood ratio tests between the SAC Durbin models against the SAC models for both the pre-CCT and post-CCT clearly favour of the model including the spatially lagged right hand side variables. We also find that the LR tests between the SAC Durbin and spatial Durbin, and the SAC Durbin and error Durbin variants are marginally significant for the pre-CCT data, but not significant for the post-CCT data. Consequently, we choose to proceed with the pre-CCT SAC Durbin model and the post-CCT Durbin model. Testing the spatial Durbin against the spatial lag model, we see that the inclusion of the spatially lagged right hand side variables appears justified in both pre-CCT and post-CCT cases; the likelihood ratio test against the spatial error model rejects the Common Factor hypothesis again in both cases.

Table 3 reports on the values of ρ for the pre-CCT and post-CCT spatial Durbin and lag models, together with their log likelihood values and σ^2. It also reports the Lagrange Multiplier test for residual spatial autocorrelation p-values; the pre-

Table 2: Likelihood ratio test results.

Model 1	Model 2	Likelihood ratio	p-value
Pre-CCT SAC Durbin	Pre-CCT SAC	36.56	1.387e-05
Pre-CCT SAC Durbin	Pre-CCT Durbin	4.067	0.04374
Pre-CCT SAC Durbin	Pre-CCT error Durbin	6.302	0.01206
Pre-CCT Durbin	Pre-CCT lag	34.88	2.816e-05
Pre-CCT Durbin	Pre-CCT error	35.64	2.045e-05
Post-CCT SAC Durbin	Post-CCT SAC	29.66	0.0002424
Post-CCT SAC Durbin	Post-CCT Durbin	1.057	0.3039
Post-CCT SAC Durbin	Post-CCT error Durbin	2.147	0.1428
Post-CCT Durbin	Post-CCT lag	30.82	0.0001511
Post-CCT Durbin	Post-CCT error	29.53	0.0002559

Table 3: Summary of spatial Durbin and lag model output (asymptotic standard errors in parentheses).

	pre-CCT Durbin	pre-CCT lag	post-CCT Durbin	post-CCT lag
ρ	0.1701	0.1495	0.1477	0.09712
	(0.07608)	(0.04233)	(0.07733)	(0.04432)
Log likelihood	39.03	21.59	51.92	36.51
σ^2	0.04572	0.05099	0.04229	0.04664
AIC	-40.06	-21.18	-65.84	-51.02
LM p-value	0.007471	0.1704	0.1283	0.1572

CCT spatial Durbin appears to induce residual spatial autocorrelation through the included spatially lagged right hand side variables, but this is alleviated post-CCT. The residual spatial autocorrelation detected in the pre-CCT spatial Durbin estimates corresponds to the significant λ coefficient estimate for the pre-CCT SAC Durbin model reported in Table 1.

Using the spatial Durbin specification, the significances of the pre-CCT and post-CCT estimates of ρ are: pre-CCT: 0.0254, post-CCT: 0.05617. Compared to the tabulations in Bivand and Szymanski (2000, p. 215), using only the spatial lag specification (computed using SpaceStat, and which agree with the output for the lag models calculated here), we see that the significance of the lag pre-CCT ρ is 0.0004138, and of the post-CCT ρ is 0.02844. The change in ρ remains that hypothesised in the earlier paper, but is much smaller in size. However, as we can see from Table 1, the aggregate spatial "signal" is strongly reduced from the pre-CCT SAC Durbin to the post-CCT spatial Durbin estimates; the two spatial coefficients in th pre-CCT SAC Durbin model are both significant.

It is now necessary to revisit the interpretation of the coefficients on the cost-sharpening variables of log density of units and log real wages, as this should now

be done through impact measures.In fitting spatial lag and spatial Durbin models, it has emerged over time that, unlike the spatial error model, the spatial dependence in the parameter ρ feeds back, obliging analysts to base interpretation not on the fitted parameters β, and γ where appropriate, but rather on correctly formulated impact measures (LeSage and Pace, 2009).

This feedback comes from the fact that the elements of the variance-covariance matrix of the coefficients for the maximum likelihood spatial error model linking λ and β are zero, $\partial^2 \ell/(\partial\beta\partial\lambda) = \mathbf{0}$, while in the spatial lag model (and by extension in the spatial Durbin model): $\partial^2 \ell/(\partial\beta\partial\rho) \neq \mathbf{0}$. In the spatial error model, for right hand side variable r, $\partial y_i/\partial x_{ir} = \beta_r$ and $\partial y_i/\partial x_{jr} = 0$ for $i \neq j$; in the spatial lag model, $\partial y_i/\partial x_{jr} = ((\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{I}\beta_r)_{ij}$, where $\mathbf{I}$ is the $N \times N$ identity matrix, and $(\mathbf{I} - \rho\mathbf{W})^{-1}$ is known to be dense (LeSage and Pace, 2009, p. 33–42).

The variance-covariance matrix of the coefficients and the series of traces of the powered weights matrix are the key ingredients needed to compute impact measures for spatial lag and spatial Durbin models; both of these are based on the representation of weights matrices as sparse matrices. We can also compute the measures analytically for smaller data sets; here we will contrast the 1980 US election and Lucas (OH) data sets, where the former is small enough to permit all the output values to be compared.

An estimate of the coefficient variance-covariance matrix is needed for Monte Carlo simulation of the impact measures, although the measures themselves may be computed without an estimate of this matrix. LeSage and Pace (2009, pp. 33–42, 114–115) and LeSage and Fischer (2008) provide the background and implementation details for impact measures.

The awkward $S_r(\mathbf{W}) = ((\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{I}\beta_r)$ matrix term needed to calculate impact measures for the lag model, and $S_r(\mathbf{W}) = ((\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{I}\beta_r - \mathbf{W}\gamma_r))$ for the spatial Durbin model, may be approximated using traces of powers of the spatial weights matrix as well as analytically. The average direct impacts are represented by the sum of the diagonal elements of the matrix divided by N for each exogenous variable, the average total impacts are the sum of all matrix elements divided by N for each exogenous variable, while the average indirect impacts are the differences between these two impact vectors.

In **spdep**, `impacts` methods are available for ML spatial lag, spatial Durbin, SAC and SAC Durbin fitted model objects. The methods use truncated series of traces using different ways of computing the traces, here powering a sparse matrix, which goes dense, to get exact traces.

Figure 2 shows that the conclusions in Bivand and Szymanski (2000) with respect to the sharpening of the impact of the cost shaping log real wage variable are sustained when the interpretation is recast in the form of direct and total impact measures. The distributions of the Monte Carlo simulations move away from
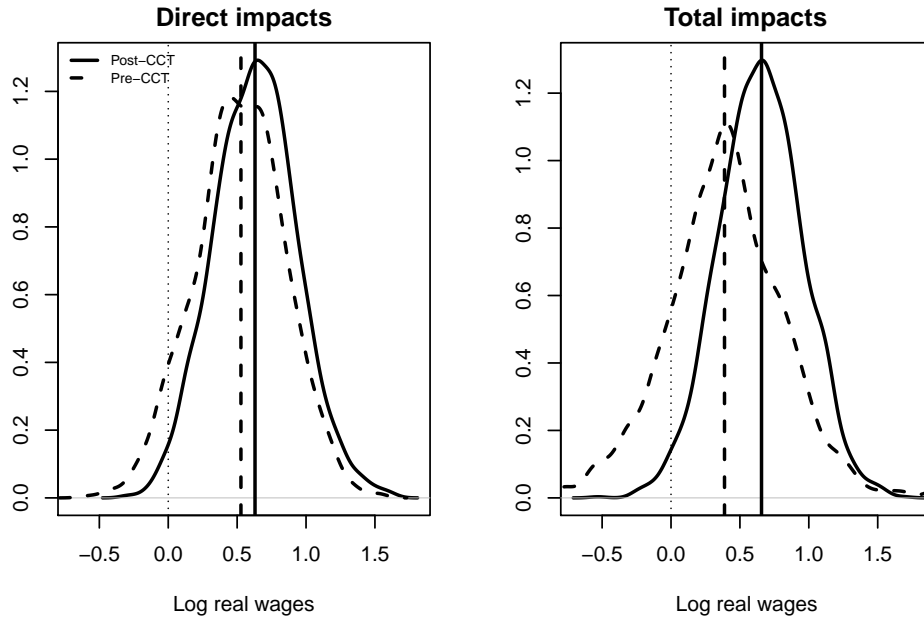
11

Figure 2: Direct and total impacts for log real wages, pre-CCT SAC Durbin and post-CCT Durbin estimates, Monte Carlo tests with 2000 simulations each.

zero, with a direct post-CCT spatial Durbin p-value of 0.0351 and a total post-CCT spatial Durbin p-value of 0.0318.

This does not, however, hold for the sharpening on another cost shaping variable, the log density of units, which is expected to be significant and negative after the introduction of CCT, as we see in Figure 3. While the direct post-CCT spatial Durbin p-value is highly significant and the sign correct (5.55e-06), the indirect impact has a different sign, and so the total impact has a disappointing p-value of 0.869. If we only fit a spatial lag model, the impacts mirror the interpretation in the original paper, and the total impacts of both cost-shaping variables have the expected signs and are both significant.

The conclusions drawn in Bivand and Szymanski (2000) need to be revised in the light of developments in spatial econometrics. "Raising the bar" changes those results in two respects, first in enriching the spatial lag models used for both pre-CCT and post-CCT data sets to pre-CCT SAC Durbin, and to post-CCT spatial Durbin specifications. The inclusion of the spatially lagged explanatory variables induced spatial error autocorrelation in the pre-CCT spatial Durbin model, indicating that the spatial process is not being fully captured by the spatial lag of the dependent variable when the spatially lagged explanatory variables are included. The pre-CCT spatial Durbin model fits the data better than the pre-CCT spatial lag
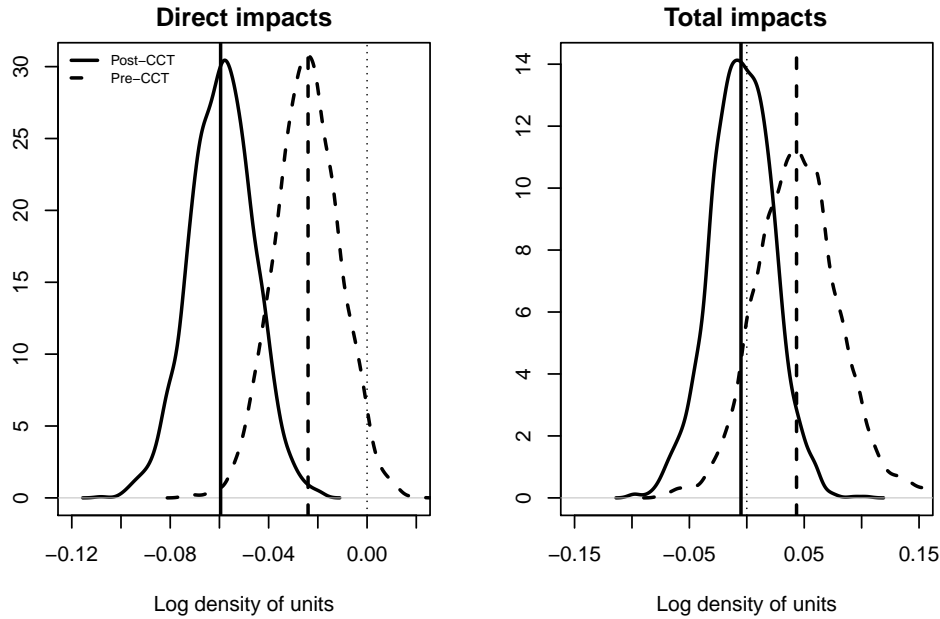
12

Figure 3: Direct and total impacts for log density of units, pre-CCT SAC Durbin and post-CCT Durbin estimates, Monte Carlo tests with 2000 simulations each.

model, but marginally worse than the pre-CCT SAC Durbin model, which has two significant spatial coefficients. If we consider the strength of the spatial processes between the pre-CCT SAC Durbin model and the post-CCT spatial Durbin model, we can comfortably sustain our former conclusion that the introduction of compulsory competitive tendering reduces spatial relationships between local authorities in garbage collection costs.

The second change is that when we move from attempting to interpret the regression coefficients of the explanatory variables to a proper analysis of variable impacts (emanating effects), using the pre-CCT SAC Durbin and post-CCT spatial Durbin models, we find that the total impact of the collection unit density variable is not significant. The impacts shift in the correct direction (higher density should lead to lower costs), and the direct impacts are significant, but the total impacts are not. This is perhaps to be expected, given the geography of the local authorities, where authorities with higher and lower densities are adjacent in some parts of the country. The conclusion with respect to the other cost-shaping variable, real wages, is unchanged, and shows that its significance increased markedly following the introduction of compulsory competitive tendering.

13

5 Larger data sets

Returning to the two larger data sets used in Bivand (2010a), we will be more concerned with the implications of fitting spatial autoregressive models where the model is very possibly misspecified with respect to omitted explanatory variables. Neither of these data sets have clearly motivated or complete modelling contexts. Presidential election turnout is influenced by other explanatory variables, some of which may be related to cultural background for broader regions than the observed counties, than those included. House selling prices are typically closely related to neighbourhood qualities, which here are unobserved. These omissions may lead to a bundle of spatial signals that are approximated by the included autoregressive term or terms.

We will fit the SAC and SAC Durbin models for the two data sets, followed by the spatial Durbin and error Durbin variants, and test using the likelihood ratio. Because of the large number of observations, the fast updating sparse Cholesky method is used for computing the Jacobian in optimising the log likelihood function, and in the gridded profiling calculations reported in Figure 4.

Table 4: Summary of large data set model output (numerical Hessian standard errors in parentheses).

	Election	Election	Lucas County	Lucas County
Model	SAC	SAC Durbin	SAC	SAC Durbin
ρ	-0.3933	-0.5113	0.6898	0.805
	(0.03626)	(0.04729)	(0.005414)	(0.003315)
λ	0.8703	0.8901	-0.3871	-0.581
	(0.01231)	(0.01323)	(0.01274)	(0.008077)
Log likelihood	4099	4115	-7336	-6184
σ^2	0.003312	0.003171	0.07731	0.05701
AIC	-8184	-8210	14704	12425

Table 4 shows summary results for the two data sets for SAC and SAC Durbin fitted models. In both cases, the values of ρ and λ take more extreme values in the SAC Durbin case, with negative ρ and positive λ in the Election data case, and with signs reversed in the Lucas County housing data case. In this table, and in Table 5, we see that the SAC model is rejected in favour of the SAC Durbin model in both cases. In addition, as Table 5 shows, likelihood ratio tests comparing the SAC Durbin model with the spatial Durbin and error Durbin models for both data sets all point to the better fit of the SAC Durbin specification.

It is very possible that both large data set models are wrongly specified. The election data set only includes three variables, and in addition may be strongly affected by inhomogeneous observational units, because the counties used differ very

greatly in population size. The Lucas County house price data set is certainly affected by the omission of contextual variables reflecting neighbourhood qualities, and this must engender a range of spatial processes, which are not fully captured by the spatially lagged dependent variable. It may also be the case that the very sparse spatial weights used are insufficiently dense to mop up the autocorrelation present.

Table 5: Likelihood ratio test results for large data sets.

Model 1	Model 2	Likelihood ratio	p-value
Election SAC Durbin	Election SAC	32.23	4.686e-07
Election SAC Durbin	Election Durbin	141.9	< 2.2e-16
Election SAC Durbin	Election error Durbin	54.67	1.424e-13
Lucas County SAC Durbin	Lucas County SAC	2303	< 2.2e-16
Lucas County SAC Durbin	Lucas County Durbin	2246	< 2.2e-16
Lucas County SAC Durbin	Lucas County error Durbin	2944	< 2.2e-16

Figure 4, like the results for the English garbage data reported in Figure 1, shows again the hallmark "banana" ridge shape of the surface of the log-likelihood function for SAC, and especially SAC Durbin models. It is a matter of concern that in the Lucas County case, the SAC Durbin surface has (at least) one local optimum at the low ρ, high λ end of the banana, and the global optimum at the low λ, high ρ end. Use of a finer grid may show whether there are more than two optima, but demonstrating more than one is already worrying.

6   Conclusions

In examining some of the practical consequences of "raising the bar", it has been shown that the conclusions of work published earlier have required modification. Had the earlier study used a spatial Durbin model, rather than a spatial lag model, it is possible that the need for revision would have been seen, but in practice the interpretation of the impacts (emanating effects) of explanatory variables has only been undertaken since 2008. It is the introduction of the interpretation of impact measures that changes the inference in this case, rather than the insertion of an additional spatially autocorrelated error.

In the English garbage case, the model is adequately specified, with a model of the suggested causes of the pre-CCT spatial dependence, and a hypothesis that the dependence will be attenuated following the introduction of compulsory competitive tendering. This hypothesis is also sustained using newer methods.

In the two larger data sets, we have no behavioural model for observed spatial autocorrelation, and in addition we have reason to believe that the models suffer
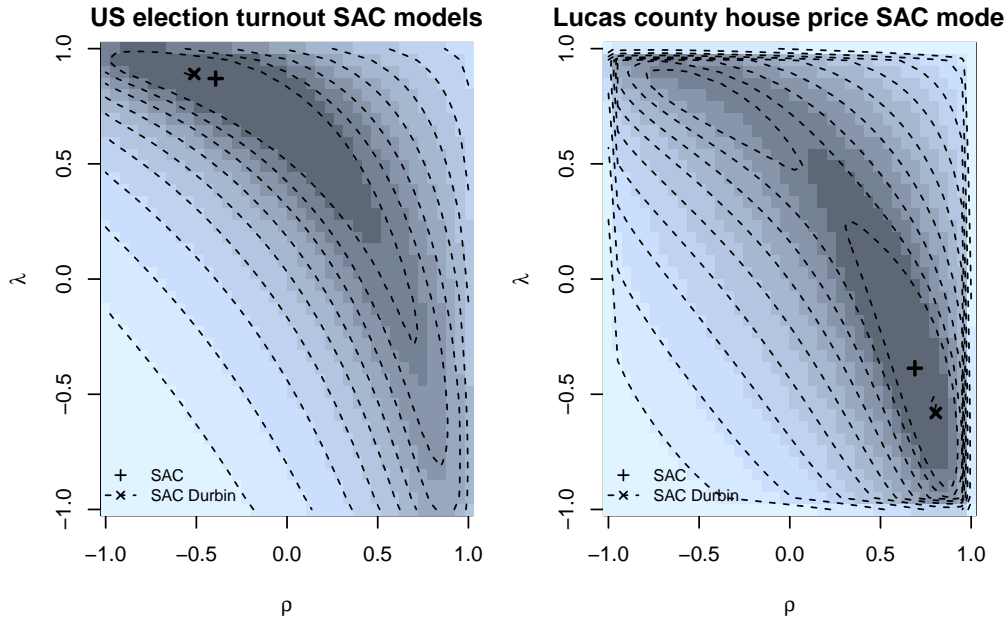
15

Figure 4: Log-likelihood function surfaces for the US election turnout and Lucas county house price models, for SAC and SAC Durbin specifications.

from omitted variables. Finally, the election turnout data set is observed for counties, which are very heterogeneous aggregations of voter turnout, and additionally the dependent variable arguably should be bounded between 0 and 1 (with the exeptional observation exceeding a 100% turnout).

Consequently, we are in a potentially difficult situation anyway, a situation that perhaps leads to the observed significant spatial autoregressive coefficients with opposed signs. They appear to be "picking up" spatial signals that are coming from the omitted explanatory variables, rather than to be expressing behavioural dependencies in space. The fitted models with two spatial parameters do, however, fit better than single parameter models, especially when the spatially lagged explanatory variables are included, but this arguably does not suggest that they are in any sense capturing "real" spatial relationships.

Since the debate on "raising the bar" is only now beginning, it seems sensible to conclude with questions rather than assertions. The following seem to be among the salient open issues. Given the numerical issues involved in fitting the SAC Durbin model, how should one interpret the output? Is it reasonable to feel that the underlying problem is that ρ and λ are insufficiently identified, and how might one test this possibility? Is this a situation in which other sources of misspecifiaction, for example heteroskedasticity, are feeding through into an apparent second, negative

16

spatial autoregressive process? Is this related to the insights given in Griffith (2006) with repect to hidden negative spatial autocorrelation? Clearly, there is substantial need for further research in order to be able to provide practitioners with adequate guidelines for model fitting.

References

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

Bates, D. and Maechler, M. (2011). *Matrix: Sparse and Dense Matrix Classes and Methods*. R package version 1.0-1.

Bivand, R. S. (1984). Regression modeling with spatial dependence: an application of some class selection and estimation methods. *Geographical Analysis*, 16:25–37.

Bivand, R. S. (2006). Implementing spatial data analysis software tools in R. *Geographical Analysis*, 38(1):23–40.

Bivand, R. S. (2010a). Comparing estimation methods for spatial econometrics techniques using R. Discussion paper 2010:26, Department of Economics, Norwegian School of Economics and Business Administration.

Bivand, R. S. (2010b). Computing the Jacobian in spatial models: an applied survey. Discussion paper 2010:20, Department of Economics, Norwegian School of Economics and Business Administration.

Bivand, R. S., Pebesma, E. J., and Gómez-Rubio, V. (2008). *Applied Spatial Data Analysis with R*. Springer, New York.

Bivand, R. S. and Portnov, B. A. (2004). Exploring spatial data analysis techniques using R: the case of observations with no neighbours. In Anselin, L., Florax, R. J. G. M., and Rey, S. J., editors, *Advances in Spatial Econometrics: Methodology, Tools, Applications*, pages 121–142. Springer, Berlin.

Bivand, R. S. and Szymanski, S. (1997). Spatial dependence through local yardstick competition: theory and testing. *Economics Letters*, 55:257–265.

Bivand, R. S. and Szymanski, S. (2000). Modelling the spatial impact of the introduction of compulsory competitive tendering. *Regional Science and Urban Economics*, 30:203–219.

Cliff, A. D. and Ord, J. K. (1973). *Spatial Autocorrelation*. Pion, London.

Drukker, D. M., Prucha, I., and Raciborski, R. (2011). Maximum-likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances. Working paper, Department of Economics, University of Maryland.

Elhorst, J. P. (2010). Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis*, 5:9–28.

Griffith, D. (2006). Hidden negative spatial autocorrelation. *Journal of Geographical Systems*, 8:335–355.

Kelejian, H. H. (2008). A spatial j-test for model specification against a single or a set of non-nested alternatives. *Letters in Spatial and Resource Sciences*, 1:3–11.

Kelejian, H. H. and Piras, G. (2011). An extension of Kelejian's J-test for non-nested spatial models. *Regional Sscience and Urban Economics*, 41(3):281–292.

Kelejian, H. H. and Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40:509–533.

LeSage, J. and Fischer, M. (2008). Spatial growth regression: Model specification, estimation and interpretation. *Spatial Economic Analysis*, 3:275–304.

LeSage, J. and Pace, R. (2009). *Introduction to Spatial Econometrics*. CRC Press, Boca Raton, FL.

Manski, C. F. (1993). Identification of endogenous social effects: the reflection problem. *Review of Economic Studies*, 60:531–542.

Ord, J. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70(349):120–126.

Pace, R. and Barry, R. (1997). Fast CARs. *Journal of Statistical Computation and Simulation*, 59(2):123–145.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ripley, B. D. (1981). *Spatial Statistics*. Wiley, New York.

18

Smirnov, O. and Anselin, L. (2009). An O(N) parallel method of computing the Log-Jacobian of the variable transformation for models with spatial interaction on a lattice. *Computational Statistics & Data Analysis*, 53(8):2980 – 2988.

Szymanski, S. (1996). The impact of compulsory competitive tendering on refuse collection services. *Fiscal Studies*, 17:1–19.

Issued in the series Discussion Papers 2010

2010

01/10 January, Øystein Foros, **Hans Jarle Kind**, and **Greg Shaffer**, "Mergers and Partial Ownership"

02/10 January, **Astrid Kunze** and Kenneth R. Troske, "Life-cycle patterns in male/female differences in job search".

03/10 January, **Øystein Daljord** and **Lars Sørgard**, "Single-Product versus Uniform SSNIPs".

04/10 January, **Alexander W. Cappelen**, James Konow, **Erik Ø. Sørensen**, and **Bertil Tungodden**, "Just luck: an experimental study of risk taking and fairness".

05/10 February, **Laurence Jacquet**, "Optimal labor income taxation under maximin: an upper bound".

06/10 February, **Ingvild Almås**, Tarjei Havnes, and Magne Mogstad, "Baby booming inequality? Demographic change and inequality in Norway, 1967-2004".

07/10 February, **Laurence Jacquet**, Etienne Lehmann, and Bruno van der Linden, "Optimal redistributive taxation with both extensive and intensive responses".

08/10 February, **Fred Schroyen**, "Income risk aversion with quantity constraints".

09/10 March, **Ingvild Almås** and Magne Mogstad, "Older or Wealthier? The impact of age adjustment on cross-sectional inequality measures".

10/10 March, Ari Hyytinen, **Frode Steen**, and Otto Toivanen, "Cartels Uncovered".

11/10 April, **Karl Ove Aarbu**, "Demand patterns for treatment insurance in Norway".

12/10 May, **Sandra E. Black**, Paul J. Devereux, and **Kjell G. Salvanes**, "Under pressure? The effect of peers on outcomes of young adults".

13/10 May, **Ola Honningdal Grytten** and Arngrim Hunnes, "A chronology of financial crises for Norway".

14/10 May, Anders Bjørklund and **Kjell G. Salvanes**, "Education and family background: Mechanisms and policies".

15/10 July, **Eva Benedicte D. Norman** and **Victor D. Norman**, "Agglomeration, tax competition and local public goods supply".

16/10 July, **Eva Benedicte D. Norman**, "The price of decentralization".

17/10 July, **Eva Benedicte D. Norman**, "Public goods production and private sector productivity".

18/10 July, **Kurt Richard Brekke**, Tor Helge Holmås, and Odd Rune Straume, "Margins and Market Shares: Pharmacy Incentives for Generic Substitution".

19/10 August, **Karl Ove Aarbu**, "Asymmetric information – evidence from the home insurance market".

20/10 August. **Roger Bivand**, "Computing the Jacobian in spatial models: an applied survey".

21/10 August, **Sturla Furunes Kvamsdal**, "An overview of Empirical Analysis of behavior of fishermen facing new regulations.

22/10 September, Torbjørn Hægeland, Lars Johannessen Kirkebøen, Odbjørn Raaum, and **Kjell G. Salvanes**, " Why children of college graduates outperform their schoolmates: A study of cousins and adoptees".

23/10 September**, Agnar Sandmo**, "Atmospheric Externalities and Environmental Taxation".

24/10 October, **Kjell G. Salvanes,** Katrine Løken, and Pedro Carneiro**, "**A flying start? Long term consequences of maternal time investments in children during their first year of life".

25/10 September, **Roger Bivand, "**Exploiting Parallelization in Spatial Statistics: an Applied Survey using R".

26/10 September, **Roger Bivand**, "Comparing estimation methods for spatial econometrics techniques using R".

27/10 October. **Lars Mathiesen**, **Øivind Anti Nilsen**, and **Lars Sørgard**, "Merger simulations with observed diversion ratios."

28/10 November, **Alexander W. Cappelen, Knut Nygaard, Erik Ø. Sørensen**, and **Bertil Tungodden**, "Efficiency, equality and reciprocity in social preferences: A comparison of students and a representative population".

29/10 December, **Magne Krogstad Asphjell**, Wilko Letterie, **Øivind A. Nilsen**, and Gerard A. Pfann, "Sequentiality versus Simultaneity: Interrelated Factor Demand".

2011

01/11 January, **Lars Ivar Oppedal Berge**, **Kjetil Bjorvatn**, and **Bertil Tungodden**, "Human and financial capital for microenterprise development: Evidence from a field and lab experiment."

02/11 February, **Kurt R. Brekke**, Luigi Siciliani, and Odd Rune Straume, "Quality competition with profit constraints: do non-profit firms provide higher quality than for-profit firms?"

03/11 February, **Gernot Doppelhofer** and Melvyn Weeks, "Robust Growth Determinants".

04/11 February, Manudeep Bhuller, Magne Mogstad, and **Kjell G. Salvanes**, "Life-Cycle Bias and the Returns to Schooling in Current and Lifetime Earnings".

05/11 March, **Knut Nygaard**, "Forced board changes: Evidence from Norway".

06/11 March, **Sigbjørn Birkeland d.y.**, "Negotiation under possible third party settlement".

07/11 April, **Fred Schroyen**, "Attitudes towards income risk in the presence of quantity constraints".

08/11 April, Craig Brett and **Laurence Jacquet**, "Workforce or Workfare?"

09/11 May, **Bjørn Basberg**, "A Crisis that Never Came. The Decline of the European Antarctic Whaling Industry in the 1950s and -60s".

10/11 June, Joseph A. Clougherty, Klaus Gugler, and **Lars Sørgard**, "Cross-Border Mergers and Domestic Wages: Integrating Positive 'Spillover' Effects and Negative 'Bargaining' Effects".

11/11 July, **Øivind A. Nilsen**, Arvid Raknerud, and Terje Skjerpen, "Using the Helmert-transformation to reduce dimensionality in a mixed model: Application to a wage equation with worker and …rm heterogeneity".

12/11 July, Karin Monstad, Carol Propper, and **Kjell G. Salvanes**, "Is teenage motherhood contagious? Evidence from a Natural Experiment".

13/11 August, **Kurt R. Brekke**, Rosella Levaggi, Luigi Siciliani, and Odd Rune Straume, "Patient Mobility, Health Care Quality and Welfare".

14/11 July, **Sigbjørn Birkeland d.y.**, "Fairness motivation in bargaining".

15/11 September, **Sigbjørn Birkeland d.y**, **Alexander Cappelen**, **Erik Ø. Sørensen**, and **Bertil Tungodden**, "Immoral criminals? An experimental study of social preferences among prisoners".

16/11 September, **Hans Jarle Kind**, Guttorm Schjelderup, and Frank Stähler, "Newspaper Differentiation and Investments in Journalism: The Role of Tax Policy".

17/11 **Gregory Corcos**, Massimo Del Gatto, Giordano Mion, and Gianmarco I.P. Ottaviano, "Productivity and Firm Selection: Quantifying the "New" Gains from Trade".

18/11 **Grant R. McDermott** and **Øivind Anti Nilsen**, "Electricity Prices, River Temperatures and Cooling Water Scarcity".

19/11 Pau Olivella and **Fred Schroyen**, "Multidimensional screening in a monopolistic insurance market".

20/11 **Liam Brunt**, "Property rights and economic growth: evidence from a natural experiment".

21/11 Pau Olivella and **Fred Schroyen**, "Multidimensional screening in a monopolistic insurance market: proofs".

22/11 **Roger Bivand**, "After "Raising the Bar": applied maximum likelihood estimation of families of models in spatial econometrics".

NHH

Norges
Handelshøyskole

Norwegian School of Economics

NHH
Helleveien 30
NO-5045 Bergen
Norway

Tlf/Tel: +47 55 95 90 00
Faks/Fax: +47 55 95 91 00
nhh.postmottak@nhh.no
www.nhh.no

Norges
Handelshøyskole

Norwegian School of Economics

NHH
Helleveien 30
NO-5045 Bergen
Norway

Tlf/Tel: +47 55 95 90 00
Faks/Fax: +47 55 95 91 00
nhh.postmottak@nhh.no
www.nhh.no