Spatial Econometric models: specification, estimation and impact analysis

Cem Ertur*

*Laboratoire d'Economie d'Orléans
UMR CNRS 6221 - Université d'Orléans - France
E-mail : cem.ertur@univ-orleans.fr
web site : http://certur.free.fr/

Motivation 1

- The baseline cross-section regression model

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i \quad \varepsilon_i \sim i.i.d.(0, \sigma^2) \qquad i = 1, \ldots, N$$

- Well known problems: heterogeneity / endogeneity / omitted variables etc.

  $\Rightarrow$ Violation of the ***exchangeability hypothesis***
  "different patterns of realized errors are equally likely to occur if the realizations are permuted across countries." (Durlauf et al. 2005, p. 581)
- Neglected problem: ***interaction*** between individuals
  $\Rightarrow$ spatial autocorrelation
  $\Rightarrow$ spatial spillovers: feedback effects and indirect effects

Motivation 1

- The baseline cross-section regression model

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i \quad \varepsilon_i \sim i.i.d.(0, \sigma^2) \qquad i = 1, \ldots, N$$

- Well known problems: heterogeneity / endogeneity / omitted variables etc.

  $\Rightarrow$ Violation of the ***exchangeability hypothesis***
  "different patterns of realized errors are equally likely to occur if the realizations are permuted across countries." (Durlauf et al. 2005, p. 581)

- Neglected problem: ***interaction*** between individuals
  $\Rightarrow$ spatial autocorrelation
  $\Rightarrow$ spatial spillovers: feedback effects and indirect effects

Motivation 1

- The baseline cross-section regression model

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i \quad \varepsilon_i \sim i.i.d.(0, \sigma^2) \qquad i = 1, \ldots, N$$

- Well known problems: heterogeneity / endogeneity / omitted variables etc.

  $\Rightarrow$ Violation of the ***exchangeability hypothesis***
  "different patterns of realized errors are equally likely to occur if the realizations are permuted across countries." (Durlauf et al. 2005, p. 581)

- Neglected problem: ***interaction*** between individuals
  $\Rightarrow$ spatial autocorrelation
  $\Rightarrow$ spatial spillovers: feedback effects and indirect effects

Motivation 1

- The baseline cross-section regression model

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i \quad \varepsilon_i \sim i.i.d.(0, \sigma^2) \qquad i = 1, \ldots, N$$

- Well known problems: heterogeneity / endogeneity / omitted variables etc.

  $\Rightarrow$ Violation of the ***exchangeability hypothesis***
  "different patterns of realized errors are equally likely to occur if the realizations are permuted across countries." (Durlauf et al. 2005, p. 581)
- Neglected problem: ***interaction*** between individuals
  $\Rightarrow$ spatial autocorrelation
  $\Rightarrow$ spatial spillovers: feedback effects and indirect effects

Motivation 2

- Great progress has been made in the past decade on the theoretical aspects of spatial econometrics, but
- Little has been proposed from the economic theory perspective to justify the use of spatial econometric models
- Little attention has been paid to the interpretation of the models in terms of impacts in the applied literature
- The aim of this presentation is to bridge the gap by proposing interpretations of various spatial econometric (cross-section and panel data) models when it comes to impact analysis
- We believe that this could be useful for applied researchers and policy makers

Motivation 2

- Great progress has been made in the past decade on the theoretical aspects of spatial econometrics, but
- Little has been proposed from the economic theory perspective to justify the use of spatial econometric models
- Little attention has been paid to the interpretation of the models in terms of impacts in the applied literature
- The aim of this presentation is to bridge the gap by proposing interpretations of various spatial econometric (cross-section and panel data) models when it comes to impact analysis
- We believe that this could be useful for applied researchers and policy makers

Motivation 2

- Great progress has been made in the past decade on the theoretical aspects of spatial econometrics, but
- Little has been proposed from the economic theory perspective to justify the use of spatial econometric models
- Little attention has been paid to the interpretation of the models in terms of impacts in the applied literature
- The aim of this presentation is to bridge the gap by proposing interpretations of various spatial econometric (cross-section and panel data) models when it comes to impact analysis
- We believe that this could be useful for applied researchers and policy makers

Motivation 2

- Great progress has been made in the past decade on the theoretical aspects of spatial econometrics, but
- Little has been proposed from the economic theory perspective to justify the use of spatial econometric models
- Little attention has been paid to the interpretation of the models in terms of impacts in the applied literature
- The aim of this presentation is to bridge the gap by proposing interpretations of various spatial econometric (cross-section and panel data) models when it comes to impact analysis
- We believe that this could be useful for applied researchers and policy makers

Motivation 2

- Great progress has been made in the past decade on the theoretical aspects of spatial econometrics, but
- Little has been proposed from the economic theory perspective to justify the use of spatial econometric models
- Little attention has been paid to the interpretation of the models in terms of impacts in the applied literature
- The aim of this presentation is to bridge the gap by proposing interpretations of various spatial econometric (cross-section and panel data) models when it comes to impact analysis
- We believe that this could be useful for applied researchers and policy makers

Outline

- Cross-Section Spatial Econometric Models
  - SAR, SDM and SEM specifications
    - Basic results
    - Hypotheses
    - Parameter Space
    - Estimation
  - Interaction multipliers
    - Impact analysis
    - Impact of random shocks
  - Applications
    - Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence, Journal of Applied Econometrics, 22, 1033-1062, 2007
    - A Contribution to the Theory and Empirics of Schumpeterian Growth with Worldwide Interactions, Journal of Economic Growth, 16:3, 215-255, 2011
- Spatial Panel Data Models

Outline

- Cross-Section Spatial Econometric Models
  - SAR, SDM and SEM specifications
    - Basic results
    - Hypotheses
    - Parameter Space
    - Estimation
  - Interaction multipliers
    - Impact analysis
    - Impact of random shocks
  - Applications
    - Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence, Journal of Applied Econometrics, 22, 1033-1062, 2007
    - A Contribution to the Theory and Empirics of Schumpeterian Growth with Worldwide Interactions, Journal of Economic Growth, 16:3, 215-255, 2011
- Spatial Panel Data Models

Outline

- Cross-Section Spatial Econometric Models
  - SAR, SDM and SEM specifications
    - Basic results
    - Hypotheses
    - Parameter Space
    - Estimation
  - Interaction multipliers
    - Impact analysis
    - Impact of random shocks
  - Applications
    - Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence, Journal of Applied Econometrics, 22, 1033-1062, 2007
    - A Contribution to the Theory and Empirics of Schumpeterian Growth with Worldwide Interactions, Journal of Economic Growth, 16:3, 215-255, 2011
- Spatial Panel Data Models

Outline

- Cross-Section Spatial Econometric Models
  - SAR, SDM and SEM specifications
    - Basic results
    - Hypotheses
    - Parameter Space
    - Estimation
  - Interaction multipliers
    - Impact analysis
    - Impact of random shocks
  - Applications
    - Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence, Journal of Applied Econometrics, 22, 1033-1062, 2007
    - A Contribution to the Theory and Empirics of Schumpeterian Growth with Worldwide Interactions, Journal of Economic Growth, 16:3, 215-255, 2011
- Spatial Panel Data Models

Outline

- Cross-Section Spatial Econometric Models
  - SAR, SDM and SEM specifications
    - Basic results
    - Hypotheses
    - Parameter Space
    - Estimation
  - Interaction multipliers
    - Impact analysis
    - Impact of random shocks
  - Applications
    - Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence, Journal of Applied Econometrics, 22, 1033-1062, 2007
    - A Contribution to the Theory and Empirics of Schumpeterian Growth with Worldwide Interactions, Journal of Economic Growth, 16:3, 215-255, 2011
- Spatial Panel Data Models

Outline

- Cross-Section Spatial Econometric Models
  - SAR, SDM and SEM specifications
    - Basic results
    - Hypotheses
    - Parameter Space
    - Estimation
  - Interaction multipliers
    - Impact analysis
    - Impact of random shocks
  - Applications
    - Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence, Journal of Applied Econometrics, 22, 1033-1062, 2007
    - A Contribution to the Theory and Empirics of Schumpeterian Growth with Worldwide Interactions, Journal of Economic Growth, 16:3, 215-255, 2011
- Spatial Panel Data Models

Outline

- Cross-Section Spatial Econometric Models
  - SAR, SDM and SEM specifications
    - Basic results
    - Hypotheses
    - Parameter Space
    - Estimation
  - Interaction multipliers
    - Impact analysis
    - Impact of random shocks
  - Applications
    - Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence, Journal of Applied Econometrics, 22, 1033-1062, 2007
    - A Contribution to the Theory and Empirics of Schumpeterian Growth with Worldwide Interactions, Journal of Economic Growth, 16:3, 215-255, 2011
- Spatial Panel Data Models

Outline

- Cross-Section Spatial Econometric Models
  - SAR, SDM and SEM specifications
    - Basic results
    - Hypotheses
    - Parameter Space
    - Estimation
  - Interaction multipliers
    - Impact analysis
    - Impact of random shocks
  - Applications
    - Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence, Journal of Applied Econometrics, 22, 1033-1062, 2007
    - A Contribution to the Theory and Empirics of Schumpeterian Growth with Worldwide Interactions, Journal of Economic Growth, 16:3, 215-255, 2011
- Spatial Panel Data Models

Outline

- Cross-Section Spatial Econometric Models
  - SAR, SDM and SEM specifications
    - Basic results
    - Hypotheses
    - Parameter Space
    - Estimation
  - Interaction multipliers
    - Impact analysis
    - Impact of random shocks
  - Applications
    - Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence, Journal of Applied Econometrics, 22, 1033-1062, 2007
    - A Contribution to the Theory and Empirics of Schumpeterian Growth with Worldwide Interactions, Journal of Economic Growth, 16:3, 215-255, 2011
- Spatial Panel Data Models

Outline

- Cross-Section Spatial Econometric Models
  - SAR, SDM and SEM specifications
    - Basic results
    - Hypotheses
    - Parameter Space
    - Estimation
  - Interaction multipliers
    - Impact analysis
    - Impact of random shocks
  - Applications
    - Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence, Journal of Applied Econometrics, 22, 1033-1062, 2007
    - A Contribution to the Theory and Empirics of Schumpeterian Growth with Worldwide Interactions, Journal of Economic Growth, 16:3, 215-255, 2011
- Spatial Panel Data Models

Outline

- Cross-Section Spatial Econometric Models
  - SAR, SDM and SEM specifications
    - Basic results
    - Hypotheses
    - Parameter Space
    - Estimation
  - Interaction multipliers
    - Impact analysis
    - Impact of random shocks
  - Applications
    - Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence, Journal of Applied Econometrics, 22, 1033-1062, 2007
    - A Contribution to the Theory and Empirics of Schumpeterian Growth with Worldwide Interactions, Journal of Economic Growth, 16:3, 215-255, 2011
- Spatial Panel Data Models

Outline

- Cross-Section Spatial Econometric Models
  - SAR, SDM and SEM specifications
    - Basic results
    - Hypotheses
    - Parameter Space
    - Estimation
  - Interaction multipliers
    - Impact analysis
    - Impact of random shocks
  - Applications
    - Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence, Journal of Applied Econometrics, 22, 1033-1062, 2007
    - A Contribution to the Theory and Empirics of Schumpeterian Growth with Worldwide Interactions, Journal of Economic Growth, 16:3, 215-255, 2011
- Spatial Panel Data Models

Outline

- Cross-Section Spatial Econometric Models
  - SAR, SDM and SEM specifications
    - Basic results
    - Hypotheses
    - Parameter Space
    - Estimation
  - Interaction multipliers
    - Impact analysis
    - Impact of random shocks
  - Applications
    - Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence, Journal of Applied Econometrics, 22, 1033-1062, 2007
    - A Contribution to the Theory and Empirics of Schumpeterian Growth with Worldwide Interactions, Journal of Economic Growth, 16:3, 215-255, 2011
- Spatial Panel Data Models

SAR specification: The spatial autoregressive model

- The SAR specification adds the spatially lagged endogenous variable to the standard explanatory variables on the right hand side of the classical cross section regression model, for $i = 1, ..., N$ we have then:

$$y_{i,N} = \beta_0 + \rho_N \sum_{j=1}^{N} w_{ij,N} y_{j,N} + \sum_{k=1}^{K} x_{ik,N} \beta_{k,N} + \varepsilon_{i,N} \tag{1}$$

or in more compact matrix form:

$$y_N = \beta_0 \iota_N + \rho_N W_N y_N + X_N \beta_N + \varepsilon_N \tag{2}$$

- If the interaction matrix is normalized to have row sums of unity, with weights $w^*_{ij,N} = w_{ij,N} / \sum_j w_{ij,N}$, then the i^{th} row of the vector $W^*_N y_N$, can be interpreted as the weighted average of the neighboring values of the spatial unit i using the spatial lag operator, i.e. $[W^*_N y_N]_i = \sum_j w^*_{ij,N} y_{j,N}$.
- Note that, in matrix form, $W^*_N = D_N W_N$, where $D_N = \text{diag}(1 / \sum_j w_{ij,N})$ is a diagonal matrix containing the inverse of the row sums of W_N.

SAR specification: The spatial autoregressive model

- The SAR specification adds the spatially lagged endogenous variable to the standard explanatory variables on the right hand side of the classical cross section regression model, for $i = 1, ..., N$ we have then:

$$y_{i,N} = \beta_0 + \rho_N \sum_{j=1}^{N} w_{ij,N} y_{j,N} + \sum_{k=1}^{K} x_{ik,N} \beta_{k,N} + \varepsilon_{i,N} \tag{1}$$

  or in more compact matrix form:

$$y_N = \beta_0 \iota_N + \rho_N W_N y_N + X_N \beta_N + \varepsilon_N \tag{2}$$

- If the interaction matrix is normalized to have row sums of unity, with weights $w^*_{ij,N} = w_{ij,N} / \sum_j w_{ij,N}$, then the i^{th} row of the vector $W^*_N y_N$, can be interpreted as the weighted average of the neighboring values of the spatial unit i using the spatial lag operator, i.e. $[W^*_N y_N]_i = \sum_j w^*_{ij,N} y_{j,N}$.
- Note that, in matrix form, $W^*_N = D_N W_N$, where $D_N = \text{diag}(1 / \sum_j w_{ij,N})$ is a diagonal matrix containing the inverse of the row sums of W_N.

SAR specification: The spatial autoregressive model

- The SAR specification adds the spatially lagged endogenous variable to the standard explanatory variables on the right hand side of the classical cross section regression model, for $i = 1, ..., N$ we have then:

$$y_{i,N} = \beta_0 + \rho_N \sum_{j=1}^{N} w_{ij,N} y_{j,N} + \sum_{k=1}^{K} x_{ik,N} \beta_{k,N} + \varepsilon_{i,N} \tag{1}$$

  or in more compact matrix form:

$$y_N = \beta_0 \iota_N + \rho_N W_N y_N + X_N \beta_N + \varepsilon_N \tag{2}$$

- If the interaction matrix is normalized to have row sums of unity, with weights $w^*_{ij,N} = w_{ij,N} / \sum_j w_{ij,N}$, then the i^{th} row of the vector $W^*_N y_N$, can be interpreted as the weighted average of the neighboring values of the spatial unit i using the spatial lag operator, i.e. $[W^*_N y_N]_i = \sum_j w^*_{ij,N} y_{j,N}$.
- Note that, in matrix form, $W^*_N = D_N W_N$, where $D_N = \text{diag}(1 / \sum_j w_{ij,N})$ is a diagonal matrix containing the inverse of the row sums of W_N.

SDM specification: The spatial Durbin model

- A straightforward generalization of the SAR model consists to add the spatially lagged exogenous variables to the specification to obtain the so-called Spatial Durbin Model, for $i = 1, ..., N$ we have then:

$$y_{i,N} = \beta_0 + \rho_N \sum_{j=1}^{N} w_{ij,N} y_{j,N} + \sum_{k=1}^{K} x_{ik,N} \beta_{k,N} + \sum_{k=1}^{K} w_{ij,N} z_{ik,N} \gamma_{k,N} + \varepsilon_{i,N} \tag{3}$$

or in more compact matrix form:

$$y_N = \rho_N W_N y_N + X_N \beta_N + W_N Z_N \gamma_N + \varepsilon_N \tag{4}$$

- For ease of exposition, let us write the SDM as a SAR model as follows:

$$y_N = \rho_N W_N y_N + \widetilde{X}_N b_N + \varepsilon_N \tag{5}$$

where $\widetilde{X}_N = [\iota \;\; X_N \;\; W_N Z_N]$ and $b_N = [\beta_0, \beta_N', \gamma_N']'$ with Z_N being either X_N or $Z_N = [X_N \; \widetilde{Z}_N]$.

SDM specification: The spatial Durbin model

- A straightforward generalization of the SAR model consists to add the spatially lagged exogenous variables to the specification to obtain the so-called Spatial Durbin Model, for $i = 1, ..., N$ we have then:

$$y_{i,N} = \beta_0 + \rho_N \sum_{j=1}^{N} w_{ij,N} y_{j,N} + \sum_{k=1}^{K} x_{ik,N} \beta_{k,N} + \sum_{k=1}^{K} w_{ij,N} z_{ik,N} \gamma_{k,N} + \varepsilon_{i,N} \tag{3}$$

or in more compact matrix form:

$$y_N = \rho_N W_N y_N + X_N \beta_N + W_N Z_N \gamma_N + \varepsilon_N \tag{4}$$

- For ease of exposition, let us write the SDM as a SAR model as follows:

$$y_N = \rho_N W_N y_N + \widetilde{X}_N b_N + \varepsilon_N \tag{5}$$

where $\widetilde{X}_N = [\iota \;\; X_N \;\; W_N Z_N]$ and $b_N = [\beta_0, \beta_N', \gamma_N']'$ with Z_N being either X_N or $Z_N = [X_N \; \widetilde{Z}_N]$.

Basic results

Spatial filter

- For simplicity and without loss of generality, consider the SAR model as defined in equation (2), with X_N including now the constant term:

$$y_N = \rho_N W_N y_N + X_N \beta_N + \varepsilon_N \tag{6}$$

- The spatial filter is defined as $(I_N - \rho_N W_N)$. Note that if ρ_N was known, we could filter out spatial autocorrelation from y_N and then use the classical regression model as follows:

$$(I_N - \rho_N W_N) y_N = X_N \beta_N + \varepsilon_N \tag{7}$$

Of course ρ_N is usually unknown and has to be estimated as well as β_N and $\sigma^2_{\varepsilon,N}$.

Basic results

Spatial filter

- For simplicity and without loss of generality, consider the SAR model as defined in equation (2), with X_N including now the constant term:

$$y_N = \rho_N W_N y_N + X_N \beta_N + \varepsilon_N \tag{6}$$

- The spatial filter is defined as $(I_N - \rho_N W_N)$. Note that if ρ_N was known, we could filter out spatial autocorrelation from y_N and then use the classical regression model as follows:

$$(I_N - \rho_N W_N) y_N = X_N \beta_N + \varepsilon_N \tag{7}$$

Of course ρ_N is usually unknown and has to be estimated as well as β_N and $\sigma^2_{\varepsilon,N}$.

Basic results

Reduced form

- If $(I_N - \rho_N W_N)$ is invertible, the reduced form of the SAR specification is obtained as:

$$y_N = (I_N - \rho_N W_N)^{-1} X_N \beta_N + (I_N - \rho_N W_N)^{-1} \varepsilon_N \tag{8}$$

- Therefore, we need to precisely define the invertibility condition for $(I_N - \rho W)$ which is needed to write the reduced form: $(I_N - \rho_N W_N)$ is invertible if $\det(I_N - \rho_N W_N) \neq 0$. Note first that if $\rho_N = 0$, $(I_N - \rho_N W_N)$ is non singular. Now consider $\rho_N \neq 0$, we have:

$$\begin{aligned} \det(I_N - \rho_N W_N) &= \det[(-\rho_N)(W_N - \frac{1}{\rho_N} I_N)] \\ &= (-\rho_N)^N \det(W_N - \frac{1}{\rho_N} I_N) \end{aligned} \tag{9}$$

- Therefore $\det(I_N - \rho_N W_N) \neq 0$ and $(I_N - \rho_N W_N)$ is non singular if $\rho_N^{-1} \notin \{\nu_{1,N}, ..., \nu_{N,N}\}$, where $\nu_{1,N}, ..., \nu_{N,N}$ denote the eigenvalues of W_N, i.e. if ρ_N^{-1} is not an eigenvalue of W_N.
- The parameter space for ρ_N will be specified and discussed in more details below.

Basic results

Reduced form

- If $(I_N - \rho_N W_N)$ is invertible, the reduced form of the SAR specification is obtained as:

$$y_N = (I_N - \rho_N W_N)^{-1} X_N \beta_N + (I_N - \rho_N W_N)^{-1} \varepsilon_N \tag{8}$$

- Therefore, we need to precisely define the invertibility condition for $(I_N - \rho W)$ which is needed to write the reduced form: $(I_N - \rho_N W_N)$ is invertible if $\det(I_N - \rho_N W_N) \neq 0$. Note first that if $\rho_N = 0$, $(I_N - \rho_N W_N)$ is non singular. Now consider $\rho_N \neq 0$, we have:

$$\begin{aligned} \det(I_N - \rho_N W_N) &= \det[(-\rho_N)(W_N - \frac{1}{\rho_N} I_N)] \\ &= (-\rho_N)^N \det(W_N - \frac{1}{\rho_N} I_N) \end{aligned} \tag{9}$$

- Therefore $\det(I_N - \rho_N W_N) \neq 0$ and $(I_N - \rho_N W_N)$ is non singular if $\rho_N^{-1} \notin \{\nu_{1,N}, ..., \nu_{N,N}\}$, where $\nu_{1,N}, ..., \nu_{N,N}$ denote the eigenvalues of W_N, i.e. if ρ_N^{-1} is not an eigenvalue of W_N.
- The parameter space for ρ_N will be specified and discussed in more details below.

Basic results

Reduced form

- If $(I_N - \rho_N W_N)$ is invertible, the reduced form of the SAR specification is obtained as:

$$y_N = (I_N - \rho_N W_N)^{-1} X_N \beta_N + (I_N - \rho_N W_N)^{-1} \varepsilon_N \tag{8}$$

- Therefore, we need to precisely define the invertibility condition for $(I_N - \rho W)$ which is needed to write the reduced form: $(I_N - \rho_N W_N)$ is invertible if $\det(I_N - \rho_N W_N) \neq 0$. Note first that if $\rho_N = 0$, $(I_N - \rho_N W_N)$ is non singular. Now consider $\rho_N \neq 0$, we have:

$$\begin{aligned} \det(I_N - \rho_N W_N) &= \det[(-\rho_N)(W_N - \frac{1}{\rho_N} I_N)] \\ &= (-\rho_N)^N \det(W_N - \frac{1}{\rho_N} I_N) \end{aligned} \tag{9}$$

- Therefore $\det(I_N - \rho_N W_N) \neq 0$ and $(I_N - \rho_N W_N)$ is non singular if $\rho_N^{-1} \notin \{\nu_{1,N}, ..., \nu_{N,N}\}$, where $\nu_{1,N}, ..., \nu_{N,N}$ denote the eigenvalues of W_N, i.e. if ρ_N^{-1} is not an eigenvalue of W_N.
- The parameter space for ρ_N will be specified and discussed in more details below.

Basic results

Reduced form

- If $(I_N - \rho_N W_N)$ is invertible, the reduced form of the SAR specification is obtained as:

$$y_N = (I_N - \rho_N W_N)^{-1} X_N \beta_N + (I_N - \rho_N W_N)^{-1} \varepsilon_N \tag{8}$$

- Therefore, we need to precisely define the invertibility condition for $(I_N - \rho W)$ which is needed to write the reduced form: $(I_N - \rho_N W_N)$ is invertible if $\det(I_N - \rho_N W_N) \neq 0$. Note first that if $\rho_N = 0$, $(I_N - \rho_N W_N)$ is non singular. Now consider $\rho_N \neq 0$, we have:

$$\begin{aligned} \det(I_N - \rho_N W_N) &= \det[(-\rho_N)(W_N - \frac{1}{\rho_N} I_N)] \\ &= (-\rho_N)^N \det(W_N - \frac{1}{\rho_N} I_N) \end{aligned} \tag{9}$$

- Therefore $\det(I_N - \rho_N W_N) \neq 0$ and $(I_N - \rho_N W_N)$ is non singular if $\rho_N^{-1} \notin \{\nu_{1,N}, ..., \nu_{N,N}\}$, where $\nu_{1,N}, ..., \nu_{N,N}$ denote the eigenvalues of W_N, i.e. if ρ_N^{-1} is not an eigenvalue of W_N.
- The parameter space for ρ_N will be specified and discussed in more details below.

Basic results

Reduced form implications

The reduced form has two important implications:

- First, in (conditional) mean, the value of the dependent variable in the cross-sectional unit i will be affected not only by the values taken by the explanatory variable in the cross-sectional unit i, but also by those in all other cross-sectional units through the inverse spatial transformation $(I - \rho_N W_N)^{-1}$, which is a full matrix. This is the so-called *global spatial multiplier effect* or *global interaction effect*.
- Second, a random shock in a specific cross-sectional unit i does not only affect the value of the dependent variable in the same unit, but also has an impact on the values of dependent variable in all other cross-sectional units through the same inverse spatial transformation. This is the so-called *spatial "diffusion" process of random shocks*.

Basic results

Reduced form implications

The reduced form has two important implications:

- First, in (conditional) mean, the value of the dependent variable in the cross-sectional unit *i* will be affected not only by the values taken by the explanatory variable in the cross-sectional unit *i*, but also by those in all other cross-sectional units through the inverse spatial transformation $(I - \rho_N W_N)^{-1}$, which is a full matrix. This is the so-called *global spatial multiplier effect* or *global interaction effect*.
- Second, a random shock in a specific cross-sectional unit *i* does not only affect the value of the dependent variable in the same unit, but also has an impact on the values of dependent variable in all other cross-sectional units through the same inverse spatial transformation. This is the so-called *spatial "diffusion" process of random shocks*.

Basic results

Variance-covariance matrix

- Assuming that X_N is nonstochastic, the mathematical expectation of y_N is $\mathrm{E}(y_N) = (I_N - \rho_N W_N)^{-1} X_N \beta_N$.
- The variance-covariance matrix of y_N is in turn obtained as:

$$\mathrm{V}(y_N) = \sigma^2_{\varepsilon,N} (I_N - \rho_N W_N)^{-1} (I_N - \rho_N W'_N)^{-1} \tag{10}$$

The structure of this variance-covariance matrix is such that every location is correlated with every other location in the system, but closer locations more so. It is also interesting to note that the variances at each location, are related to the neighborhood structure and are therefore not constant, inducing heteroskedasticity even though the initial process is not heteroskedastic.

Basic results

Variance-covariance matrix

- Assuming that X_N is nonstochastic, the mathematical expectation of y_N is $\mathrm{E}(y_N) = (I_N - \rho_N W_N)^{-1} X_N \beta_N$.
- The variance-covariance matrix of y_N is in turn obtained as:

$$\mathrm{V}(y_N) = \sigma^2_{\varepsilon,N} (I_N - \rho_N W_N)^{-1} (I_N - \rho_N W'_N)^{-1} \tag{10}$$

The structure of this variance-covariance matrix is such that every location is correlated with every other location in the system, but closer locations more so. It is also interesting to note that the variances at each location, are related to the neighborhood structure and are therefore not constant, inducing heteroskedasticity even though the initial process is not heteroskedastic.

Basic results

Variance-covariance matrix

- Moreover, note that:

$$W_N y_N = W_N(I_N - \rho_N W_N)^{-1} X_N \beta_N + W_N(I_N - \rho_N W_N)^{-1} \varepsilon_N \tag{11}$$

and that the spatially lagged endogenous variable $W_N y_N$ is correlated with the error term ε_N, because, in general:

$$\begin{aligned}
\mathrm{E}[(W_N y_N)' \varepsilon_N] &= \mathrm{E}[(W_N(I_N - \rho_N W_N)^{-1} X_N \beta_N + W_N(I_N - \rho_N W_N)^{-1} \varepsilon_N)' \varepsilon_N] \\
&= \mathrm{E}[\beta_N' X_N'(I_N - \rho_N W_N')^{-1} W_N' \varepsilon_N + \varepsilon_N'(I_N - \rho_N W_N')^{-1} W_N' \varepsilon_N] \\
&= \beta_N' X_N'(I_N - \rho_N W_N')^{-1} W_N' \,\mathrm{E}[\varepsilon_N] + \mathrm{E}[\varepsilon_N'(I_N - \rho_N W_N')^{-1} W_N' \varepsilon_N] \\
&= \mathrm{E}[\varepsilon_N'(I_N - \rho_N W_N')^{-1} W_N' \varepsilon_N] \qquad (12) \\
&= \mathrm{E}[\mathrm{tr}(\varepsilon_N'(I_N - \rho_N W_N')^{-1} W_N' \varepsilon_N)] \\
&= \mathrm{E}[\mathrm{tr}(I_N - \rho_N W_N')^{-1} W_N' \varepsilon_N \varepsilon_N'] \\
&= \mathrm{tr}(I_N - \rho_N W_N')^{-1} W_N' \,\mathrm{E}(\varepsilon_N \varepsilon_N') \\
&= \sigma^2_{\varepsilon,N} \,\mathrm{tr}(I_N - \rho_N W_N')^{-1} W_N' \neq 0
\end{aligned}$$

- In general tr $W_N(I_N - \rho_N W_N)^{-1}$ will not be equal to zero. A direct implication of this result is that the parameters of equation (1) cannot be consitently estimated by OLS.

Basic results

Variance-covariance matrix

- Moreover, note that:

$$W_N y_N = W_N(I_N - \rho_N W_N)^{-1} X_N \beta_N + W_N(I_N - \rho_N W_N)^{-1} \varepsilon_N \tag{11}$$

and that the spatially lagged endogenous variable $W_N y_N$ is correlated with the error term ε_N, because, in general:

$$\begin{aligned}
\mathrm{E}[(W_N y_N)' \varepsilon_N] &= \mathrm{E}[(W_N(I_N - \rho_N W_N)^{-1} X_N \beta_N + W_N(I_N - \rho_N W_N)^{-1} \varepsilon_N)' \varepsilon_N] \\
&= \mathrm{E}[\beta_N' X_N'(I_N - \rho_N W_N')^{-1} W_N' \varepsilon_N + \varepsilon_N'(I_N - \rho_N W_N')^{-1} W_N' \varepsilon_N] \\
&= \beta_N' X_N'(I_N - \rho_N W_N')^{-1} W_N' \,\mathrm{E}[\varepsilon_N] + \mathrm{E}[\varepsilon_N'(I_N - \rho_N W_N')^{-1} W_N' \varepsilon_N] \\
&= \mathrm{E}[\varepsilon_N'(I_N - \rho_N W_N')^{-1} W_N' \varepsilon_N] \qquad (12) \\
&= \mathrm{E}[\mathrm{tr}(\varepsilon_N'(I_N - \rho_N W_N')^{-1} W_N' \varepsilon_N)] \\
&= \mathrm{E}[\mathrm{tr}(I_N - \rho_N W_N')^{-1} W_N' \varepsilon_N \varepsilon_N'] \\
&= \mathrm{tr}(I_N - \rho_N W_N')^{-1} W_N' \,\mathrm{E}(\varepsilon_N \varepsilon_N') \\
&= \sigma^2_{\varepsilon,N} \,\mathrm{tr}(I_N - \rho_N W_N')^{-1} W_N' \neq 0
\end{aligned}$$

- In general $\mathrm{tr}\, W_N(I_N - \rho_N W_N)^{-1}$ will not be equal to zero. A direct implication of this result is that the parameters of equation (1) cannot be consitently estimated by OLS.

Hypotheses 1

More specifically, the interaction matrix and the autoregressive parameter are generally assumed to satisfy the following assumptions (Lee, 2004).

- **Assumption 1** (a) The disturbances $\{\varepsilon_{i,N} : 1 \leq i \leq N, N \geq 1\}$ are identically distributed. Moreover, for each sample size N, they are jointly independently distributed with $\mathrm{E}(\varepsilon_{i,N}) = 0$ and $\mathrm{E}(\varepsilon_{i,N}^2) = \sigma_{\varepsilon,N}^2$, where $0 < \sigma_{\varepsilon,N}^2 < b$ with $b < \infty$. (b) Finally, $\mathrm{E}(|\varepsilon_{i,N}|^{4+\eta})$ for some $\eta > 0$ exists, that is, a moment higher than the fourth exists.
- **Assumption 2** The elements of X_N are uniformly bounded constants, X_N has the full rank k, and $\lim_{N\to\infty}(1/N)X_N'X_N$ exists and is non singular.
- **Assumption 3** (a) All diagonal elements of W_N are zero. (b) $\rho_N \in (-\underline{a}_N^{\rho}, \bar{a}_N^{\rho})$ with $0 < \underline{a}_N^{\rho}, \bar{a}_N^{\rho} \leq a_N^{\rho} < \infty$. (c) The matrix $I_N - \rho_N W$ is non singular for all $\rho_N \in (-\underline{a}_N^{\rho}, \bar{a}_N^{\rho})$.
- **Assumption 4** The row and column sums of the sequences of matrices W_N and $(I_N - \rho W_N)^{-1}$ are bounded uniformly in absolute value.

Hypotheses 1

More specifically, the interaction matrix and the autoregressive parameter are generally assumed to satisfy the following assumptions (Lee, 2004).

- **Assumption 1** (a) The disturbances $\{\varepsilon_{i,N} : 1 \leq i \leq N, N \geq 1\}$ are identically distributed. Moreover, for each sample size N, they are jointly independently distributed with $\mathrm{E}(\varepsilon_{i,N}) = 0$ and $\mathrm{E}(\varepsilon_{i,N}^2) = \sigma_{\varepsilon,N}^2$, where $0 < \sigma_{\varepsilon,N}^2 < b$ with $b < \infty$. (b) Finally, $\mathrm{E}(|\varepsilon_{i,N}|^{4+\eta})$ for some $\eta > 0$ exists, that is, a moment higher than the fourth exists.
- **Assumption 2** The elements of X_N are uniformly bounded constants, X_N has the full rank k, and $\lim_{N\to\infty}(1/N)X_N'X_N$ exists and is non singular.
- **Assumption 3** (a) All diagonal elements of W_N are zero. (b) $\rho_N \in (-\underline{a}_N^{\rho}, \bar{a}_N^{\rho})$ with $0 < \underline{a}_N^{\rho}, \bar{a}_N^{\rho} \leq a_N^{\rho} < \infty$. (c) The matrix $I_N - \rho_N W$ is non singular for all $\rho_N \in (-\underline{a}_N^{\rho}, \bar{a}_N^{\rho})$.
- **Assumption 4** The row and column sums of the sequences of matrices W_N and $(I_N - \rho W_N)^{-1}$ are bounded uniformly in absolute value.

Hypotheses 1

More specifically, the interaction matrix and the autoregressive parameter are generally assumed to satisfy the following assumptions (Lee, 2004).

- **Assumption 1** (a) The disturbances $\{\varepsilon_{i,N} : 1 \leq i \leq N, N \geq 1\}$ are identically distributed. Moreover, for each sample size N, they are jointly independently distributed with $\mathrm{E}(\varepsilon_{i,N}) = 0$ and $\mathrm{E}(\varepsilon_{i,N}^2) = \sigma_{\varepsilon,N}^2$, where $0 < \sigma_{\varepsilon,N}^2 < b$ with $b < \infty$. (b) Finally, $\mathrm{E}(|\varepsilon_{i,N}|^{4+\eta})$ for some $\eta > 0$ exists, that is, a moment higher than the fourth exists.
- **Assumption 2** The elements of X_N are uniformly bounded constants, X_N has the full rank k, and $\lim_{N\to\infty}(1/N)X_N'X_N$ exists and is non singular.
- **Assumption 3** (a) All diagonal elements of W_N are zero. (b) $\rho_N \in (-\underline{a}_N^{\rho}, \bar{a}_N^{\rho})$ with $0 < \underline{a}_N^{\rho}, \bar{a}_N^{\rho} \leq a_N^{\rho} < \infty$. (c) The matrix $I_N - \rho_N W$ is non singular for all $\rho_N \in (-\underline{a}_N^{\rho}, \bar{a}_N^{\rho})$.
- **Assumption 4** The row and column sums of the sequences of matrices W_N and $(I_N - \rho W_N)^{-1}$ are bounded uniformly in absolute value.

Hypotheses 1

More specifically, the interaction matrix and the autoregressive parameter are generally assumed to satisfy the following assumptions (Lee, 2004).

- **Assumption 1** (a) The disturbances $\{\varepsilon_{i,N} : 1 \leq i \leq N, N \geq 1\}$ are identically distributed. Moreover, for each sample size N, they are jointly independently distributed with $\mathrm{E}(\varepsilon_{i,N}) = 0$ and $\mathrm{E}(\varepsilon_{i,N}^2) = \sigma_{\varepsilon,N}^2$, where $0 < \sigma_{\varepsilon,N}^2 < b$ with $b < \infty$. (b) Finally, $\mathrm{E}(|\varepsilon_{i,N}|^{4+\eta})$ for some $\eta > 0$ exists, that is, a moment higher than the fourth exists.
- **Assumption 2** The elements of X_N are uniformly bounded constants, X_N has the full rank k, and $\lim_{N\to\infty}(1/N)X_N'X_N$ exists and is non singular.
- **Assumption 3** (a) All diagonal elements of W_N are zero. (b) $\rho_N \in (-\underline{a}_N^{\rho}, \bar{a}_N^{\rho})$ with $0 < \underline{a}_N^{\rho}, \bar{a}_N^{\rho} \leq a_N^{\rho} < \infty$. (c) The matrix $I_N - \rho_N W$ is non singular for all $\rho_N \in (-\underline{a}_N^{\rho}, \bar{a}_N^{\rho})$.
- **Assumption 4** The row and column sums of the sequences of matrices W_N and $(I_N - \rho W_N)^{-1}$ are bounded uniformly in absolute value.

Hypotheses 2

- Let A_N be a square matrix, we say that the row and column sums of the sequences of matrices A_N is bounded uniformly in absolute value if there exists a constant $c < \infty$ that does not depend on N such that:

$$\|A_N\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{N} |a_{ij,N}| < c, \quad \|A_N\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{N} |a_{ij,N}| < c, \quad \text{for all } N$$

  Note that this condition is identical to the condition that the sequences of the maximum row sum matrix norms $\|A_N\|_\infty$ and the maximum column sum matrix norms $\|A_N\|_1$ are bounded Horn and Johnson (1985, p. 294-295).

- Under this set of assumptions, Lee (2004) proofs consistency, asymptotic normality and efficiency of the Quasi Maximum Likelihood estimators.

Hypotheses 2

- Let A_N be a square matrix, we say that the row and column sums of the sequences of matrices A_N is bounded uniformly in absolute value if there exists a constant $c < \infty$ that does not depend on N such that:

$$\|A_N\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{N} |a_{ij,N}| < c, \quad \|A_N\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{N} |a_{ij,N}| < c, \quad \text{for all } N$$

  Note that this condition is identical to the condition that the sequences of the maximum row sum matrix norms $\|A_N\|_\infty$ and the maximum column sum matrix norms $\|A_N\|_1$ are bounded Horn and Johnson (1985, p. 294-295).
- Under this set of assumptions, Lee (2004) proofs consistency, asymptotic normality and efficiency of the Quasi Maximum Likelihood estimators.

Discussion 1: Triangular arrays

- **Assumption 1** allows the error term to depend on the sample size N, i.e. to form *triangular arrays*. Note that even if the error term does not depend on N, the elements of y_N would still depend on N since the elements of the inverse of $I_N - \rho_N W_N$ would generally depend on N. A *triangular array* (tableau triangulaire) of random variables is a doubly indexed sequence in which each row (column) is only as long as the row's (column's) index. For example, the first element of the vector y will be different if $N = 10$ and $N = 15$. This implies that these elements and the vector y should be indexed by N:

$$y_N = (y_{1N}, y_{2N}, \ldots, y_{NN})$$

Our samples for y for $N = 1, 2, 3, \ldots$ are then (in rows):

$$\begin{array}{l|lll} N=1 & y_{11} & & \\ N=2 & y_{12} & y_{22} & \\ N=3 & y_{13} & y_{23} & y_{33} \\ \ldots & & \ldots & \end{array}$$

where $y_{11} \neq y_{12} \neq y_{13}$, $y_{22} \neq y_{23}$ etc.
The triangular nature of the variables, which leads to certain statistical problems, especially concerning the asymptotic properties of the maximum likelihood estimators, has, as far as we know, only been recognized by (Kelejian and Prucha 1998) or (Lee 2002).

Discussion 2

- **Assumption 2:** The nonstochastic nature of X_N and its uniform boundedness conditions are for simplicity. They can be relaxed without any problem: if the elements of X_N are stochastic and have unbounded ranges, conditions in Assumption 2 can be replaced by some finite moment conditions.
- **Assumption 3 and 4 deserve some particular attention as they are properly specific to spatial regression models.**
  - **Assumption 3(a)** is clearly a normalization rule: it implies that no unit is viewed as its own neighbor.
  - **Assumption 3(b)** underlines that the autoregressive parameter ρ depends on the sample size N as underlined by Kelejian and Prucha (2010). It also defines the parameter space for ρ_N as an interval around zero such that $(I_N - \rho_N W_N)$ is non-singular for values of ρ_N in that interval.
  - **Assumption 3(c)** ensures that y_N is uniquely defined in the reduced form equation.

Discussion 3: Variance-covariance matrix

- **Assumptions 3 and 4** imply that the row and column sums of the variance-covariance matrix of y_N in equation (10) are uniformly bounded in absolute value, thus limiting the degree of correlation between the elements of y_N.
- Indeed if A_N and B_N are two matrices conformable for multiplication and whose row and column sums are uniformly bounded in absolute value. Then the row and column sums of $A_N B_N$ are also uniformly bounded in absolute value (Kelejian and Prucha p. 516, 1999).
- The extent of correlation is limited in virtually all large sample analysis. Making an analogy to the time series literature, these assumptions ensure that the process for the dependent variable exhibit a fading memory.
- **Illustration:** let $\{X_i\},\ i = 1, ..., N$ be a random sample, where $\mathrm{E}(X_i) = \mu$, $\mathrm{V}(X_i) = \sigma^2$ for all i and $\mathrm{cov}(X_i, X_j) = a\sigma^2$ with $0 < a < 1$ for $i \neq j$. Consider now the sample mean: $\bar{X}_N$, then clearly $\mathrm{E}(\bar{X}_N) = \mu$, but $\mathrm{V}(\bar{X}_N) = \frac{\sigma^2}{N^2}[N^2 a + N(1-a)]$ and $\lim_{N\to\infty} \mathrm{V}(\bar{X}_N) = a\sigma^2 \neq 0$. Therefore, $\bar{X}_N$ is not consistent in mean square. For $\bar{X}_N$ to be consistent in mean square, the extent of correlation must be limited so as its variance goes to zero when the sample size goes to infinity, which implies that the covariances go to zero for large $|i - j|$.

Discussion 3: Variance-covariance matrix

- **Assumptions 3 and 4** imply that the row and column sums of the variance-covariance matrix of y_N in equation (10) are uniformly bounded in absolute value, thus limiting the degree of correlation between the elements of y_N.
- Indeed if A_N and B_N are two matrices conformable for multiplication and whose row and column sums are uniformly bounded in absolute value. Then the row and column sums of $A_N B_N$ are also uniformly bounded in absolute value (Kelejian and Prucha p. 516, 1999).
- The extent of correlation is limited in virtually all large sample analysis. Making an analogy to the time series literature, these assumptions ensure that the process for the dependent variable exhibit a fading memory.
- **Illustration:** let $\{X_i\},\ i = 1, ..., N$ be a random sample, where $\mathrm{E}(X_i) = \mu$, $\mathrm{V}(X_i) = \sigma^2$ for all i and $\mathrm{cov}(X_i, X_j) = a\sigma^2$ with $0 < a < 1$ for $i \neq j$. Consider now the sample mean: $\bar{X}_N$, then clearly $\mathrm{E}(\bar{X}_N) = \mu$, but $\mathrm{V}(\bar{X}_N) = \frac{\sigma^2}{N^2}[N^2 a + N(1-a)]$ and $\lim_{N\to\infty} \mathrm{V}(\bar{X}_N) = a\sigma^2 \neq 0$. Therefore, $\bar{X}_N$ is not consistent in mean square. For $\bar{X}_N$ to be consistent in mean square, the extent of correlation must be limited so as its variance goes to zero when the sample size goes to infinity, which implies that the covariances go to zero for large $|i - j|$.

Discussion 3: Variance-covariance matrix

- **Assumptions 3 and 4** imply that the row and column sums of the variance-covariance matrix of y_N in equation (10) are uniformly bounded in absolute value, thus limiting the degree of correlation between the elements of y_N.
- Indeed if A_N and B_N are two matrices conformable for multiplication and whose row and column sums are uniformly bounded in absolute value. Then the row and column sums of $A_N B_N$ are also uniformly bounded in absolute value (Kelejian and Prucha p. 516, 1999).
- The extent of correlation is limited in virtually all large sample analysis. Making an analogy to the time series literature, these assumptions ensure that the process for the dependent variable exhibit a fading memory.
- **Illustration:** let $\{X_i\},\ i = 1, ..., N$ be a random sample, where $\mathrm{E}(X_i) = \mu$, $\mathrm{V}(X_i) = \sigma^2$ for all i and $\mathrm{cov}(X_i, X_j) = a\sigma^2$ with $0 < a < 1$ for $i \neq j$. Consider now the sample mean: $\bar{X}_N$, then clearly $\mathrm{E}(\bar{X}_N) = \mu$, but $\mathrm{V}(\bar{X}_N) = \frac{\sigma^2}{N^2}[N^2 a + N(1-a)]$ and $\lim_{N\to\infty} \mathrm{V}(\bar{X}_N) = a\sigma^2 \neq 0$. Therefore, $\bar{X}_N$ is not consistent in mean square. For $\bar{X}_N$ to be consistent in mean square, the extent of correlation must be limited so as its variance goes to zero when the sample size goes to infinity, which implies that the covariances go to zero for large $|i - j|$.

Discussion 3: Variance-covariance matrix

- **Assumptions 3 and 4** imply that the row and column sums of the variance-covariance matrix of y_N in equation (10) are uniformly bounded in absolute value, thus limiting the degree of correlation between the elements of y_N.
- Indeed if A_N and B_N are two matrices conformable for multiplication and whose row and column sums are uniformly bounded in absolute value. Then the row and column sums of $A_N B_N$ are also uniformly bounded in absolute value (Kelejian and Prucha p. 516, 1999).
- The extent of correlation is limited in virtually all large sample analysis. Making an analogy to the time series literature, these assumptions ensure that the process for the dependent variable exhibit a fading memory.
- **Illustration:** let $\{X_i\},\ i = 1, ..., N$ be a random sample, where $\mathrm{E}(X_i) = \mu$, $\mathrm{V}(X_i) = \sigma^2$ for all i and $\mathrm{cov}(X_i, X_j) = a\sigma^2$ with $0 < a < 1$ for $i \neq j$. Consider now the sample mean: $\bar{X}_N$, then clearly $\mathrm{E}(\bar{X}_N) = \mu$, but $\mathrm{V}(\bar{X}_N) = \frac{\sigma^2}{N^2}[N^2 a + N(1-a)]$ and $\lim_{N\to\infty} \mathrm{V}(\bar{X}_N) = a\sigma^2 \neq 0$. Therefore, $\bar{X}_N$ is not consistent in mean square. For $\bar{X}_N$ to be consistent in mean square, the extent of correlation must be limited so as its variance goes to zero when the sample size goes to infinity, which implies that the covariances go to zero for large $|i - j|$.

Discussion 4: Interaction matrices

- In practice, interactions matrices, specially spatial weight matrices, are often specified to be row normalized.
- In many of these cases, no spatial unit is assumed to be a neighbor to more than a given number q of other spatial units. That is, for every j the number of $w_{ij} \neq 0$ is less than or equal to q.
- In other words, each spatial unit has a *limited* number of neighbors regardless of the sample size N. Clearly, in such cases, the spatial weight matrix W_N is *sparse* for large N and Assumption 3 is satisfied.
- Also, in other cases, the spatial weight matrix does not contain zeros, but the weights are formulated such that they decline rapidly as a function of some measure of distance between neighbors. Again, in such cases Assumption 3 will typically be satisfied for W_N.

Discussion 4: Interaction matrices

- In practice, interactions matrices, specially spatial weight matrices, are often specified to be row normalized.
- In many of these cases, no spatial unit is assumed to be a neighbor to more than a given number q of other spatial units. That is, for every j the number of $w_{ij} \neq 0$ is less than or equal to q.
- In other words, each spatial unit has a *limited* number of neighbors regardless of the sample size N. Clearly, in such cases, the spatial weight matrix W_N is *sparse* for large N and Assumption 3 is satisfied.
- Also, in other cases, the spatial weight matrix does not contain zeros, but the weights are formulated such that they decline rapidly as a function of some measure of distance between neighbors. Again, in such cases Assumption 3 will typically be satisfied for W_N.

Discussion 4: Interaction matrices

- In practice, interactions matrices, specially spatial weight matrices, are often specified to be row normalized.
- In many of these cases, no spatial unit is assumed to be a neighbor to more than a given number q of other spatial units. That is, for every j the number of $w_{ij} \neq 0$ is less than or equal to q.
- In other words, each spatial unit has a *limited* number of neighbors regardless of the sample size N. Clearly, in such cases, the spatial weight matrix W_N is *sparse* for large N and Assumption 3 is satisfied.
- Also, in other cases, the spatial weight matrix does not contain zeros, but the weights are formulated such that they decline rapidly as a function of some measure of distance between neighbors. Again, in such cases Assumption 3 will typically be satisfied for W_N.

Discussion 4: Interaction matrices

- In practice, interactions matrices, specially spatial weight matrices, are often specified to be row normalized.
- In many of these cases, no spatial unit is assumed to be a neighbor to more than a given number q of other spatial units. That is, for every j the number of $w_{ij} \neq 0$ is less than or equal to q.
- In other words, each spatial unit has a *limited* number of neighbors regardless of the sample size N. Clearly, in such cases, the spatial weight matrix W_N is *sparse* for large N and Assumption 3 is satisfied.
- Also, in other cases, the spatial weight matrix does not contain zeros, but the weights are formulated such that they decline rapidly as a function of some measure of distance between neighbors. Again, in such cases Assumption 3 will typically be satisfied for W_N.

Parameter space: Motivations 1

- **Assumption 3(b)** defines the parameter space for the autoregressive parameter.
- In the existing literature the parameter space for the autoregressive parameter is typically taken to be the interval $(-1, 1)$, or a subset thereof, and the autoregressive parameter is assumed not to depend on the sample size.
- However, in applications it is typically found that for un-normalized interaction matrices, $I_N - \rho_N W_N$ is singular for some values of $\rho \in (-1, 1)$.
- To avoid this situation, many applied researchers normalize each row of their interaction matrices in such a way that $I_N - \rho_N W_N$ is non-singular for all $\rho \in (-1, 1)$.
- Let us now discuss the implications of various normalizations of the interaction matrix.

Parameter space: Motivations 1

- **Assumption 3(b)** defines the parameter space for the autoregressive parameter.
- In the existing literature the parameter space for the autoregressive parameter is typically taken to be the interval $(-1, 1)$, or a subset thereof, and the autoregressive parameter is assumed not to depend on the sample size.
- However, in applications it is typically found that for un-normalized interaction matrices, $I_N - \rho_N W_N$ is singular for some values of $\rho \in (-1, 1)$.
- To avoid this situation, many applied researchers normalize each row of their interaction matrices in such a way that $I_N - \rho_N W_N$ is non-singular for all $\rho \in (-1, 1)$.
- Let us now discuss the implications of various normalizations of the interaction matrix.

Parameter space: Motivations 1

- **Assumption 3(b)** defines the parameter space for the autoregressive parameter.
- In the existing literature the parameter space for the autoregressive parameter is typically taken to be the interval $(-1, 1)$, or a subset thereof, and the autoregressive parameter is assumed not to depend on the sample size.
- However, in applications it is typically found that for un-normalized interaction matrices, $I_N - \rho_N W_N$ is singular for some values of $\rho \in (-1, 1)$.
- To avoid this situation, many applied researchers normalize each row of their interaction matrices in such a way that $I_N - \rho_N W_N$ is non-singular for all $\rho \in (-1, 1)$.
- Let us now discuss the implications of various normalizations of the interaction matrix.

Parameter space: Motivations 1

- **Assumption 3(b)** defines the parameter space for the autoregressive parameter.
- In the existing literature the parameter space for the autoregressive parameter is typically taken to be the interval $(-1, 1)$, or a subset thereof, and the autoregressive parameter is assumed not to depend on the sample size.
- However, in applications it is typically found that for un-normalized interaction matrices, $I_N - \rho_N W_N$ is singular for some values of $\rho \in (-1, 1)$.
- To avoid this situation, many applied researchers normalize each row of their interaction matrices in such a way that $I_N - \rho_N W_N$ is non-singular for all $\rho \in (-1, 1)$.
- Let us now discuss the implications of various normalizations of the interaction matrix.

Parameter space: Motivations 1

- **Assumption 3(b)** defines the parameter space for the autoregressive parameter.
- In the existing literature the parameter space for the autoregressive parameter is typically taken to be the interval $(-1, 1)$, or a subset thereof, and the autoregressive parameter is assumed not to depend on the sample size.
- However, in applications it is typically found that for un-normalized interaction matrices, $I_N - \rho_N W_N$ is singular for some values of $\rho \in (-1, 1)$.
- To avoid this situation, many applied researchers normalize each row of their interaction matrices in such a way that $I_N - \rho_N W_N$ is non-singular for all $\rho \in (-1, 1)$.
- Let us now discuss the implications of various normalizations of the interaction matrix.

Parameter space: Motivations 2

- Suppose c_N denotes a scalar normalization factor. Clearly, this normalization factor may depend on the sample size N.
- For example, some of our results below relate to the case in which c_N corresponds to the maximal row or column sum of the absolute values of the elements of W_N.
- Given such a normalizing factor, an equivalent specification for y_N is obtained if $\rho_N W_N$ is replaced by $\rho_N^* W_N^*$ where $\rho_N^* = c_N \rho_N$ and $W_N^* = W_N / c_N$.
- It is important to observe that even if ρ_N and its corresponding parameter space do not depend on N, ρ_N^* and its implied parameter space *will* depend on the sample size as a result of the normalization of the interaction matrix.
- It is for this reason that we allow in Assumption 3 for the elements of the interaction matrices, and the autoregressive parameters and the corresponding parameter spaces to depend on N. Of course, Assumption 3 also covers the case where the true data generating process corresponds to a model where autoregressive parameters do not depend on N.

Parameter space: Motivations 2

- Suppose c_N denotes a scalar normalization factor. Clearly, this normalization factor may depend on the sample size N.
- For example, some of our results below relate to the case in which c_N corresponds to the maximal row or column sum of the absolute values of the elements of W_N.
- Given such a normalizing factor, an equivalent specification for y_N is obtained if $\rho_N W_N$ is replaced by $\rho_N^* W_N^*$ where $\rho_N^* = c_N \rho_N$ and $W_N^* = W_N / c_N$.
- It is important to observe that even if ρ_N and its corresponding parameter space do not depend on N, ρ_N^* and its implied parameter space *will* depend on the sample size as a result of the normalization of the interaction matrix.
- It is for this reason that we allow in Assumption 3 for the elements of the interaction matrices, and the autoregressive parameters and the corresponding parameter spaces to depend on N. Of course, Assumption 3 also covers the case where the true data generating process corresponds to a model where autoregressive parameters do not depend on N.

Parameter space: Motivations 2

- Suppose c_N denotes a scalar normalization factor. Clearly, this normalization factor may depend on the sample size N.
- For example, some of our results below relate to the case in which c_N corresponds to the maximal row or column sum of the absolute values of the elements of W_N.
- Given such a normalizing factor, an equivalent specification for y_N is obtained if $\rho_N W_N$ is replaced by $\rho_N^* W_N^*$ where $\rho_N^* = c_N \rho_N$ and $W_N^* = W_N / c_N$.
- It is important to observe that even if ρ_N and its corresponding parameter space do not depend on N, ρ_N^* and its implied parameter space *will* depend on the sample size as a result of the normalization of the interaction matrix.
- It is for this reason that we allow in Assumption 3 for the elements of the interaction matrices, and the autoregressive parameters and the corresponding parameter spaces to depend on N. Of course, Assumption 3 also covers the case where the true data generating process corresponds to a model where autoregressive parameters do not depend on N.

Parameter space: Motivations 2

- Suppose c_N denotes a scalar normalization factor. Clearly, this normalization factor may depend on the sample size N.
- For example, some of our results below relate to the case in which c_N corresponds to the maximal row or column sum of the absolute values of the elements of W_N.
- Given such a normalizing factor, an equivalent specification for y_N is obtained if $\rho_N W_N$ is replaced by $\rho_N^* W_N^*$ where $\rho_N^* = c_N \rho_N$ and $W_N^* = W_N / c_N$.
- It is important to observe that even if ρ_N and its corresponding parameter space do not depend on N, ρ_N^* and its implied parameter space *will* depend on the sample size as a result of the normalization of the interaction matrix.
- It is for this reason that we allow in Assumption 3 for the elements of the interaction matrices, and the autoregressive parameters and the corresponding parameter spaces to depend on N. Of course, Assumption 3 also covers the case where the true data generating process corresponds to a model where autoregressive parameters do not depend on N.

Parameter space: Motivations 2

- Suppose c_N denotes a scalar normalization factor. Clearly, this normalization factor may depend on the sample size N.
- For example, some of our results below relate to the case in which c_N corresponds to the maximal row or column sum of the absolute values of the elements of W_N.
- Given such a normalizing factor, an equivalent specification for y_N is obtained if $\rho_N W_N$ is replaced by $\rho_N^* W_N^*$ where $\rho_N^* = c_N \rho_N$ and $W_N^* = W_N / c_N$.
- It is important to observe that even if ρ_N and its corresponding parameter space do not depend on N, ρ_N^* and its implied parameter space *will* depend on the sample size as a result of the normalization of the interaction matrix.
- It is for this reason that we allow in Assumption 3 for the elements of the interaction matrices, and the autoregressive parameters and the corresponding parameter spaces to depend on N. Of course, Assumption 3 also covers the case where the true data generating process corresponds to a model where autoregressive parameters do not depend on N.

Parameter space: Theorem 1

- Having defined in Assumption 3, the parameter space for ρ as an interval around zero such that $(I_N - \rho W)$ is non-singular for values of ρ in that interval, the following Theorem gives the bounds for that interval.
- **Theorem 1** (Kelejian and Prucha, 2010, p. 56)
  Let τ denote the spectral radius of W, i.e.,

$$\tau = \max\{|\nu_1|, ..., |\nu_N|\}$$

  where $|\nu_1|, ..., |\nu_N|$ denote the modulus of the eigenvalues of W. Then $(I_N - \rho W)$ is nonsingular for all values of ρ in the interval $(-1/\tau, 1/\tau)$.
- **Proof**
  Consider that for $\rho \neq 0$, $\det(I_N - \rho W) = \det[(-\rho)(W - \frac{1}{\rho} I_N)] = (-\rho)^N \det(W - \frac{1}{\rho} I_N)$.
  Consequently $(I_N - \rho W)$ is non singular for values of $\rho^{-1} \notin \{\nu_1, ..., \nu_N\}$, i.e. if ρ^{-1} is not an eigenvalue of W. In particular $(I_N - \rho W)$ is nonsingular for $|\rho^{-1}| > \tau$. Rewriting the last inequality as $|\rho| < \tau^{-1}$ completes the proof.

Parameter space: Theorem 1

- Having defined in Assumption 3, the parameter space for ρ as an interval around zero such that $(I_N - \rho W)$ is non-singular for values of ρ in that interval, the following Theorem gives the bounds for that interval.
- **Theorem 1** (Kelejian and Prucha, 2010, p. 56)
  Let τ denote the spectral radius of W, i.e.,

$$\tau = \max\{|\nu_1|, ..., |\nu_N|\}$$

  where $|\nu_1|, ..., |\nu_N|$ denote the modulus of the eigenvalues of W. Then $(I_N - \rho W)$ is nonsingular for all values of ρ in the interval $(-1/\tau, 1/\tau)$.
- **Proof**
  Consider that for $\rho \neq 0$, $\det(I_N - \rho W) = \det[(-\rho)(W - \frac{1}{\rho} I_N)] = (-\rho)^N \det(W - \frac{1}{\rho} I_N)$.
  Consequently $(I_N - \rho W)$ is non singular for values of $\rho^{-1} \notin \{\nu_1, ..., \nu_N\}$, i.e. if ρ^{-1} is not an eigenvalue of W. In particular $(I_N - \rho W)$ is nonsingular for $|\rho^{-1}| > \tau$. Rewriting the last inequality as $|\rho| < \tau^{-1}$ completes the proof.

Parameter space: Theorem 1

- Having defined in Assumption 3, the parameter space for ρ as an interval around zero such that $(I_N - \rho W)$ is non-singular for values of ρ in that interval, the following Theorem gives the bounds for that interval.
- **Theorem 1** (Kelejian and Prucha, 2010, p. 56)
  Let τ denote the spectral radius of W, i.e.,

$$\tau = \max\{|\nu_1|, ..., |\nu_N|\}$$

  where $|\nu_1|, ..., |\nu_N|$ denote the modulus of the eigenvalues of W. Then $(I_N - \rho W)$ is nonsingular for all values of ρ in the interval $(-1/\tau, 1/\tau)$.
- **Proof**
  Consider that for $\rho \neq 0$, $\det(I_N - \rho W) = \det[(-\rho)(W - \frac{1}{\rho} I_N)] = (-\rho)^N \det(W - \frac{1}{\rho} I_N)$.
  Consequently $(I_N - \rho W)$ is non singular for values of $\rho^{-1} \notin \{\nu_1, ..., \nu_N\}$, i.e. if ρ^{-1} is not an eigenvalue of W. In particular $(I_N - \rho W)$ is nonsingular for $|\rho^{-1}| > \tau$. Rewriting the last inequality as $|\rho| < \tau^{-1}$ completes the proof.

Parameter space: Theorem 2

- **Theorem 2** (Kelejian and Prucha, 2010, p.56)
  Let

$$r = \max_{1 \leq i \leq N} \sum_{j=1}^{N} |w_{ij}|, \qquad c = \max_{1 \leq j \leq N} \sum_{i=1}^{N} |w_{ij}|$$

  and let

$$\tau^* = \min(r, c)$$

  Then $\tau \leq \tau^*$ and consequently $I_N - \rho W$ is non-singular for all values of ρ in the interval $(-1/\tau^*, 1/\tau^*)$.

- **Proof**
  Note first that r is the maximum row sum matrix norm and c is the maximum column sum matrix norm of W. As an immediate consequence of Gershgorin's Theorem (Horn and Johnson, 1985, p. 344-346): the spectral radius is the greatest lower bound for the values of all matrix norms of W. We then have $\tau = \max\{|\nu_1|, ..., |\nu_N|\} \leq \tau^*$. The result now follows from Theorem 1.

Parameter space: Theorem 2

- **Theorem 2** (Kelejian and Prucha, 2010, p.56)
  Let

$$r = \max_{1 \leq i \leq N} \sum_{j=1}^{N} |w_{ij}|, \qquad c = \max_{1 \leq j \leq N} \sum_{i=1}^{N} |w_{ij}|$$

  and let

$$\tau^* = \min(r, c)$$

  Then $\tau \leq \tau^*$ and consequently $I_N - \rho W$ is non-singular for all values of ρ in the interval $(-1/\tau^*, 1/\tau^*)$.

- **Proof**
  Note first that r is the maximum row sum matrix norm and c is the maximum column sum matrix norm of W. As an immediate consequence of Gershgorin's Theorem (Horn and Johnson, 1985, p. 344-346): the spectral radius is the greatest lower bound for the values of all matrix norms of W. We then have $\tau = \max\{|\nu_1|, ..., |\nu_N|\} \leq \tau^*$. The result now follows from Theorem 1.

Parameter space: Remarks

- Note however that the specification of the parameter space is here somewhat more restrictive than the previous one, i.e. $(-1/\tau^*, 1/\tau^*) \subset (-1/\tau, 1/\tau)$.
- Note that if $|\rho| < \frac{1}{\tau^*}$ or $\frac{1}{\tau}$, $(I_N - \rho W)^{-1}$ can be expanded into an infinite series as:

$$(I_N - \rho W)^{-1} = I_N + \rho W + \rho^2 W^2 + ... + \rho^r W^r + ... = \sum_{r=0}^{\infty} \rho^r W^r \tag{13}$$

- It is possible to consider the specification of the parameter space for ρ in some special cases such as when W is a row normalized interaction matrix or when W is a symmetric, not row normalized interaction matrix.

Parameter space: Remarks

- Note however that the specification of the parameter space is here somewhat more restrictive than the previous one, i.e. $(-1/\tau^*, 1/\tau^*) \subset (-1/\tau, 1/\tau)$.
- Note that if $|\rho| < \frac{1}{\tau^*}$ or $\frac{1}{\tau}$, $(I_N - \rho W)^{-1}$ can be expanded into an infinite series as:

$$(I_N - \rho W)^{-1} = I_N + \rho W + \rho^2 W^2 + ... + \rho^r W^r + ... = \sum_{r=0}^{\infty} \rho^r W^r \tag{13}$$

- It is possible to consider the specification of the parameter space for ρ in some special cases such as when W is a row normalized interaction matrix or when W is a symmetric, not row normalized interaction matrix.

Parameter space: Remarks

- Note however that the specification of the parameter space is here somewhat more restrictive than the previous one, i.e. $(-1/\tau^*, 1/\tau^*) \subset (-1/\tau, 1/\tau)$.
- Note that if $|\rho| < \frac{1}{\tau^*}$ or $\frac{1}{\tau}$, $(I_N - \rho W)^{-1}$ can be expanded into an infinite series as:

$$(I_N - \rho W)^{-1} = I_N + \rho W + \rho^2 W^2 + ... + \rho^r W^r + ... = \sum_{r=0}^{\infty} \rho^r W^r \tag{13}$$

- It is possible to consider the specification of the parameter space for ρ in some special cases such as when W is a row normalized interaction matrix or when W is a symmetric, not row normalized interaction matrix.

Parameter space

Row normalized W matrix 1

- **Theorem**
  If W is a row normalized interaction matrix, then $(I_N - \rho W)^{-1}$ exists for all $|\rho| < 1$.
- **Proof 1**
  Consider the case where W is normalized to have row sums of unity. All the eigenvalues of the row normalized W matrix are then less than 1: indeed, in this case $r = 1$, then $|\nu_i| \leq 1$ for all i. Using Theorem 2, $(I_N - \rho W)$ is nonsingular for all values of ρ in the interval $(-1, 1)$.
- **Proof 2**
  Alternatively, let us prove this result using Gershgorin's Theorem.

Parameter space

Row normalized W matrix 1

- **Theorem**
  If W is a row normalized interaction matrix, then $(I_N - \rho W)^{-1}$ exists for all $|\rho| < 1$.
- **Proof 1**
  Consider the case where W is normalized to have row sums of unity. All the eigenvalues of the row normalized W matrix are then less than 1: indeed, in this case $r = 1$, then $|\nu_i| \leq 1$ for all i. Using Theorem 2, $(I_N - \rho W)$ is nonsingular for all values of ρ in the interval $(-1, 1)$.
- **Proof 2**
  Alternatively, let us prove this result using Gershgorin's Theorem.

Parameter space

Row normalized W matrix 1

- **Theorem**
  If W is a row normalized interaction matrix, then $(I_N - \rho W)^{-1}$ exists for all $|\rho| < 1$.
- **Proof 1**
  Consider the case where W is normalized to have row sums of unity. All the eigenvalues of the row normalized W matrix are then less than 1: indeed, in this case $r = 1$, then $|\nu_i| \leq 1$ for all i. Using Theorem 2, $(I_N - \rho W)$ is nonsingular for all values of ρ in the interval $(-1, 1)$.
- **Proof 2**
  Alternatively, let us prove this result using Gershgorin's Theorem.

Parameter space

Row normalized W matrix 2

- **Gershgorin's Theorem**
  Let A be a $(N \times N)$ square matrix with elements a_{ij}. Let

$$R_i = \sum_{j=1, j \neq i}^{N} |a_{ij}|, \qquad C_j = \sum_{i=1, i \neq j}^{N} |a_{ij}|$$

  Then each eigenvalue of A lies in at least one of the N circles

$$|\nu - a_{ii}| \leq R_i, \ i = 1, ..., N$$

  and hence in the union of these circles. Also each eigenvalue of A lies in at least one of the N circles, and hence in their union

$$|\nu - a_{jj}| \leq C_j, \ j = 1, ..., N$$

- **Proof 2**
  Consider now W which has $w_{ij} \geq 0$ and $w_{ii} = 0$ and let

$$r = \max_i \sum_{j=1}^{N} w_{ij} = \max_i R_i, \qquad c = \max_j \sum_{i=1}^{N} w_{ij} = \max_j C_j,$$

  Then the eigenvalues of W satisfy for $i = 1, ..., N$: $|\nu_i| \leq r, |\nu_i| \leq c$. If W is row normalized, $r = 1$ and so $|\nu_i| \leq 1$.

Parameter space

Row normalized W matrix 2

- **Gershgorin's Theorem**
  Let A be a $(N \times N)$ square matrix with elements a_{ij}. Let

$$R_i = \sum_{j=1, j \neq i}^{N} |a_{ij}|, \qquad C_j = \sum_{i=1, i \neq j}^{N} |a_{ij}|$$

  Then each eigenvalue of A lies in at least one of the N circles

$$|\nu - a_{ii}| \leq R_i, \ i = 1, ..., N$$

  and hence in the union of these circles. Also each eigenvalue of A lies in at least one of the N circles, and hence in their union

$$|\nu - a_{jj}| \leq C_j, \ j = 1, ..., N$$

- **Proof 2**
  Consider now W which has $w_{ij} \geq 0$ and $w_{ii} = 0$ and let

$$r = \max_i \sum_{j=1}^{N} w_{ij} = \max_i R_i, \qquad c = \max_j \sum_{i=1}^{N} w_{ij} = \max_j C_j,$$

  Then the eigenvalues of W satisfy for $i = 1, ..., N$: $|\nu_i| \leq r, |\nu_i| \leq c$. If W is row normalized, $r = 1$ and so $|\nu_i| \leq 1$.

Parameter space

Row normalized W matrix 3

Next, let Q be a nonsingular matrix that triangularize W as follows:

$$QWQ^{-1} = G_\nu, \qquad G_\nu = \begin{pmatrix} \nu_1 & \dots & * \\ \vdots & \ddots & \vdots \\ 0 & \dots & \nu_N \end{pmatrix}$$

This is always possible, e.g. G_ν may represent the Jordan normal form (Horn and Johnson 1985, p.119-128). Then

$$\begin{aligned} \det(I_N - \rho W) &= \det(Q^{-1}Q(I_N - \rho W)) = \det(Q(I_N - \rho W)Q^{-1}) \\ &= \det(QQ^{-1} - \rho QWQ^{-1}) = \det(Q^{-1}Q - \rho G_\nu) = \det(I_N - \rho G_\nu) \\ &= (1 - \rho\nu_1)(1 - \rho\nu_2)\dots(1 - \rho\nu_N) = \Pi_{i=1}^{N}(1 - \rho\nu_i) \end{aligned}$$

and $\det(I_N - \rho W) \neq 0$ for any $|\rho| < 1$, since $|\rho\nu_i| \leq |\rho| < 1$.

- Moreover it is clear that $\det(I_N - \rho W) > 0$ for any $|\rho| < 1$, meaning that $I_N - \rho W$ is definite positive with all eigenvalues strictly positive.

Parameter space

Row normalized W matrix 3

Next, let Q be a nonsingular matrix that triangularize W as follows:

$$QWQ^{-1} = G_\nu, \qquad G_\nu = \begin{pmatrix} \nu_1 & \dots & * \\ \vdots & \ddots & \vdots \\ 0 & \dots & \nu_N \end{pmatrix}$$

This is always possible, e.g. G_ν may represent the Jordan normal form (Horn and Johnson 1985, p.119-128). Then

$$\begin{aligned} \det(I_N - \rho W) &= \det(Q^{-1}Q(I_N - \rho W)) = \det(Q(I_N - \rho W)Q^{-1}) \\ &= \det(QQ^{-1} - \rho QWQ^{-1}) = \det(Q^{-1}Q - \rho G_\nu) = \det(I_N - \rho G_\nu) \\ &= (1 - \rho\nu_1)(1 - \rho\nu_2)\dots(1 - \rho\nu_N) = \Pi_{i=1}^{N}(1 - \rho\nu_i) \end{aligned}$$

and $\det(I_N - \rho W) \neq 0$ for any $|\rho| < 1$, since $|\rho\nu_i| \leq |\rho| < 1$.

- Moreover it is clear that $\det(I_N - \rho W) > 0$ for any $|\rho| < 1$, meaning that $I_N - \rho W$ is definite positive with all eigenvalues strictly positive.

Parameter space

Not row normalized W matrix

- If W is not row normalized $I_N - \rho W$ will generally be singular for certain values of $|\rho| < 1$. Then using Theorems 1 or 2 it is always possible to normalize the interaction matrix in such a way that the inverse of $I_N - \rho W$ will exist in an easily established region.
- Consider for exemple the following model where W is not row normalized:

$$y = \rho W y + X\beta + \epsilon = \rho^* W^* y + X\beta + \epsilon \tag{14}$$

where $\rho^* = \tau^* \rho$, $W^* = \frac{W}{\tau^*}$ and $\tau^* = \min(r, c)$ defined in Theorem 2. Note that $|I_N - \rho^* W^*| \neq 0$ for:

$$|\rho^*| < \frac{1}{\min\left(\frac{r}{\tau^*}, \frac{c}{\tau^*}\right)} = \frac{1}{\frac{1}{\tau^*}\min(r, c)} = 1 \tag{15}$$

So if the model is normalized using τ^* and ρ^* is taken to be the parameter, $(I_N - \rho^* W^*)^{-1}$ exists for all $|\rho^*| < 1$. One would then estimate ρ^* as a parameter, and since $\rho^* = \tau^* \rho$, one would estimate ρ as $\hat{\rho} = \hat{\rho}^* / \tau^*$.

- We could also use τ, the spectral radius of W, defined in Theorem 1, for an alternative normalization. Note however that τ^* is much easier to compute than τ, especially for large sample sizes.

Parameter space

Not row normalized W matrix

- If W is not row normalized $I_N - \rho W$ will generally be singular for certain values of $|\rho| < 1$. Then using Theorems 1 or 2 it is always possible to normalize the interaction matrix in such a way that the inverse of $I_N - \rho W$ will exist in an easily established region.
- Consider for exemple the following model where W is not row normalized:

$$y = \rho W y + X\beta + \epsilon = \rho^* W^* y + X\beta + \epsilon \tag{14}$$

where $\rho^* = \tau^* \rho$, $W^* = \frac{W}{\tau^*}$ and $\tau^* = \min(r, c)$ defined in Theorem 2. Note that $|I_N - \rho^* W^*| \neq 0$ for:

$$|\rho^*| < \frac{1}{\min\left(\frac{r}{\tau^*}, \frac{c}{\tau^*}\right)} = \frac{1}{\frac{1}{\tau^*}\min(r, c)} = 1 \tag{15}$$

So if the model is normalized using τ^* and ρ^* is taken to be the parameter, $(I_N - \rho^* W^*)^{-1}$ exists for all $|\rho^*| < 1$. One would then estimate ρ^* as a parameter, and since $\rho^* = \tau^* \rho$, one would estimate ρ as $\hat{\rho} = \hat{\rho}^* / \tau^*$.

- We could also use τ, the spectral radius of W, defined in Theorem 1, for an alternative normalization. Note however that τ^* is much easier to compute than τ, especially for large sample sizes.

Parameter space

Not row normalized W matrix

- If W is not row normalized $I_N - \rho W$ will generally be singular for certain values of $|\rho| < 1$. Then using Theorems 1 or 2 it is always possible to normalize the interaction matrix in such a way that the inverse of $I_N - \rho W$ will exist in an easily established region.
- Consider for exemple the following model where W is not row normalized:

$$y = \rho W y + X\beta + \epsilon = \rho^* W^* y + X\beta + \epsilon \tag{14}$$

where $\rho^* = \tau^* \rho$, $W^* = \frac{W}{\tau^*}$ and $\tau^* = \min(r, c)$ defined in Theorem 2. Note that $|I_N - \rho^* W^*| \neq 0$ for:

$$|\rho^*| < \frac{1}{\min\left(\frac{r}{\tau^*}, \frac{c}{\tau^*}\right)} = \frac{1}{\frac{1}{\tau^*}\min(r, c)} = 1 \tag{15}$$

So if the model is normalized using τ^* and ρ^* is taken to be the parameter, $(I_N - \rho^* W^*)^{-1}$ exists for all $|\rho^*| < 1$. One would then estimate ρ^* as a parameter, and since $\rho^* = \tau^* \rho$, one would estimate ρ as $\hat{\rho} = \hat{\rho}^* / \tau^*$.

- We could also use τ, the spectral radius of W, defined in Theorem 1, for an alternative normalization. Note however that τ^* is much easier to compute than τ, especially for large sample sizes.

Parameter space

Symmetric not row normalized W matrix

- **Theorem**
  Consider the case where W is symmetric, all the eigenvalues of W are then real. Assume that W is not row normalized. Let ν_{max} and ν_{min} be the largest and the smallest eigenvalues of W. Assume as will typically be case if all the eigenvalues of W are real, that $\nu_{max} > 0$ and $\nu_{min} < 0$. Then $(I_N - \rho W)$ is nonsingular for all values of ρ in the interval $(\nu_{min}^{-1}, \nu_{max}^{-1})$.
- **Proof**
  If $\rho = 0$, $(I_N - \rho W)$ is nonsingular. If $\rho \neq 0$ we have:

$$|I_N - \rho W| = (1 - \rho\nu_1)(1 - \rho\nu_2)\dots(1 - \rho\nu_N) = \Pi_{i=1}^{N}(1 - \rho\nu_i)$$

  so $(I_N - \rho W)$ is nonsingular unless ρ is equal to the inverse of an eigenvalue $\nu_1^{-1}, \nu_2^{-1} \dots \nu_N^{-1}$, i.e ρ^{-1} is equal to an eigenvalue. Thus $(I_N - \rho W)$ is nonsingular if $\rho^{-1} < \nu_{min}$ or $\rho > \nu_{min}^{-1}$ and $\rho^{-1} > \nu_{max}$ or $\rho < \nu_{max}^{-1}$ and therefore if $\rho \in (\nu_{min}^{-1}, \nu_{max}^{-1})$

Parameter space

Symmetric not row normalized W matrix

- **Theorem**
  Consider the case where W is symmetric, all the eigenvalues of W are then real. Assume that W is not row normalized. Let ν_{max} and ν_{min} be the largest and the smallest eigenvalues of W. Assume as will typically be case if all the eigenvalues of W are real, that $\nu_{max} > 0$ and $\nu_{min} < 0$. Then $(I_N - \rho W)$ is nonsingular for all values of ρ in the interval $(\nu_{min}^{-1}, \nu_{max}^{-1})$.
- **Proof**
  If $\rho = 0$, $(I_N - \rho W)$ is nonsingular. If $\rho \neq 0$ we have:

$$|I_N - \rho W| = (1 - \rho\nu_1)(1 - \rho\nu_2)\ldots(1 - \rho\nu_N) = \Pi_{i=1}^{N}(1 - \rho\nu_i)$$

  so $(I_N - \rho W)$ is nonsingular unless ρ is equal to the inverse of an eigenvalue $\nu_1^{-1}, \nu_2^{-1} \ldots \nu_N^{-1}$, i.e ρ^{-1} is equal to an eigenvalue. Thus $(I_N - \rho W)$ is nonsingular if $\rho^{-1} < \nu_{min}$ or $\rho > \nu_{min}^{-1}$ and $\rho^{-1} > \nu_{max}$ or $\rho < \nu_{max}^{-1}$ and therefore if $\rho \in (\nu_{min}^{-1}, \nu_{max}^{-1})$

Parameter space

Symmetric not row normalized W matrix

- Consider now the normalization by τ of a symmetric not row-normalized matrix W. As all the eigenvalues of this matrix are real so that $\nu_{max} > 0$ and $\nu_{min} < 0$. We know that $(I_N - \rho W)$ is nonsingular for all values of ρ in the interval $(1/\nu_{min}, 1/\nu_{max})$.
- Consider $\rho^* = \tau\rho$ and $W^* = W/\tau$, eigenvalues of the normalized matrix W^* are given by $\nu^* = \nu/\tau$ and $(I_N - \rho^* W^*)$ is nonsingular for all values of ρ^* in the interval $(\tau/\nu_{min}, \tau/\nu_{max})$.
- Suppose that $|\nu_{max}| > |\nu_{min}|$, then the previous interval becomes $(\nu_{max}/\nu_{min}, 1)$. Suppose now that $|\nu_{max}| < |\nu_{min}|$, then $\tau = |\nu_{min}|$ and the previous interval becomes $(-1, |\nu_{min}|/\nu_{max})$. Note that those two intervals are less restrictive than the interval $(-1, 1)$: $(-1, 1) \subset (\nu_{max}/\nu_{min}, 1)$ and $(-1, 1) \subset (-1, |\nu_{min}|/\nu_{max})$.

Parameter space

Symmetric not row normalized W matrix

- Consider now the normalization by τ of a symmetric not row-normalized matrix W. As all the eigenvalues of this matrix are real so that $\nu_{max} > 0$ and $\nu_{min} < 0$. We know that $(I_N - \rho W)$ is nonsingular for all values of ρ in the interval $(1/\nu_{min}, 1/\nu_{max})$.
- Consider $\rho^* = \tau\rho$ and $W^* = W/\tau$, eigenvalues of the normalized matrix W^* are given by $\nu^* = \nu/\tau$ and $(I_N - \rho^* W^*)$ is nonsingular for all values of ρ^* in the interval $(\tau/\nu_{min}, \tau/\nu_{max})$.
- Suppose that $|\nu_{max}| > |\nu_{min}|$, then the previous interval becomes $(\nu_{max}/\nu_{min}, 1)$. Suppose now that $|\nu_{max}| < |\nu_{min}|$, then $\tau = |\nu_{min}|$ and the previous interval becomes $(-1, |\nu_{min}|/\nu_{max})$. Note that those two intervals are less restrictive than the interval $(-1, 1)$: $(-1, 1) \subset (\nu_{max}/\nu_{min}, 1)$ and $(-1, 1) \subset (-1, |\nu_{min}|/\nu_{max})$.

Parameter space

Symmetric not row normalized W matrix

- Consider now the normalization by τ of a symmetric not row-normalized matrix W. As all the eigenvalues of this matrix are real so that $\nu_{max} > 0$ and $\nu_{min} < 0$. We know that $(I_N - \rho W)$ is nonsingular for all values of ρ in the interval $(1/\nu_{min}, 1/\nu_{max})$.
- Consider $\rho^* = \tau\rho$ and $W^* = W/\tau$, eigenvalues of the normalized matrix W^* are given by $\nu^* = \nu/\tau$ and $(I_N - \rho^* W^*)$ is nonsingular for all values of ρ^* in the interval $(\tau/\nu_{min}, \tau/\nu_{max})$.
- Suppose that $|\nu_{max}| > |\nu_{min}|$, then the previous interval becomes $(\nu_{max}/\nu_{min}, 1)$. Suppose now that $|\nu_{max}| < |\nu_{min}|$, then $\tau = |\nu_{min}|$ and the previous interval becomes $(-1, |\nu_{min}|/\nu_{max})$. Note that those two intervals are less restrictive than the interval $(-1, 1)$: $(-1, 1) \subset (\nu_{max}/\nu_{min}, 1)$ and $(-1, 1) \subset (-1, |\nu_{min}|/\nu_{max})$.

Parameter space: Special Cases 1

- **Similarity**
  Recall that a Matrix B is said to be similar to a matrix A if there exist a nonsingular matrix S such that $B = S^{-1}AS$. Note also that if A and B are similar, then they have the same eigenvalues, counting multiplicity (Horn and Johnson 1985, p.44-45)
- Consider the case where the row normalized matrix is computed from a symmetric matrix, the row normalized matrix is no more symmetric and may have complex eigenvalues. However in this case, it will have the same real eigenvalues and determinant as a similar symmetric matrix.
- **Theorem**
  For symmetric matrices similar to row normalized matrices where $\nu_{max} = 1$, $(I_N - \rho W)$ is nonsingular for all values of ρ in the interval $(\nu_{min}^{-1}, 1)$.
- **Proof**
  Consider $R = I_N - \rho DW$, where DW is row normalized and non-symmetric. However, in this case DW has the same real eigenvalues as the symmetric but not row normalized matrix $D^{1/2}WD^{1/2}$. Consider the similarity transformation:

$$D^{-1/2}RD^{1/2} = D^{-1/2}(I_N - \rho DW)D^{1/2} = I_N - \rho D^{-1/2}DWD^{1/2} = I_N - \rho D^{1/2}WD^{1/2}$$

  Then $I_N - \rho DW$ and $I_N - \rho D^{1/2}WD^{1/2}$ have then the same real eigenvalues and determinant. As DW is row normalized $\nu_{max} = 1$ and $I_N - \rho D^{1/2}WD^{1/2}$ is nonsingular for all values of ρ in the interval $(\nu_{min}^{-1}, 1)$.

Parameter space: Special Cases 1

- **Similarity**
  Recall that a Matrix B is said to be similar to a matrix A if there exist a nonsingular matrix S such that $B = S^{-1}AS$. Note also that if A and B are similar, then they have the same eigenvalues, counting multiplicity (Horn and Johnson 1985, p.44-45)
- Consider the case where the row normalized matrix is computed from a symmetric matrix, the row normalized matrix is no more symmetric and may have complex eigenvalues. However in this case, it will have the same real eigenvalues and determinant as a similar symmetric matrix.
- **Theorem**
  For symmetric matrices similar to row normalized matrices where $\nu_{max} = 1$, $(I_N - \rho W)$ is nonsingular for all values of ρ in the interval $(\nu_{min}^{-1}, 1)$.
- **Proof**
  Consider $R = I_N - \rho DW$, where DW is row normalized and non-symmetric. However, in this case DW has the same real eigenvalues as the symmetric but not row normalized matrix $D^{1/2}WD^{1/2}$. Consider the similarity transformation:

$$D^{-1/2}RD^{1/2} = D^{-1/2}(I_N - \rho DW)D^{1/2} = I_N - \rho D^{-1/2}DWD^{1/2} = I_N - \rho D^{1/2}WD^{1/2}$$

  Then $I_N - \rho DW$ and $I_N - \rho D^{1/2}WD^{1/2}$ have then the same real eigenvalues and determinant. As DW is row normalized $\nu_{max} = 1$ and $I_N - \rho D^{1/2}WD^{1/2}$ is nonsingular for all values of ρ in the interval $(\nu_{min}^{-1}, 1)$.

Parameter space: Special Cases 1

- **Similarity**
  Recall that a Matrix B is said to be similar to a matrix A if there exist a nonsingular matrix S such that $B = S^{-1}AS$. Note also that if A and B are similar, then they have the same eigenvalues, counting multiplicity (Horn and Johnson 1985, p.44-45)
- Consider the case where the row normalized matrix is computed from a symmetric matrix, the row normalized matrix is no more symmetric and may have complex eigenvalues. However in this case, it will have the same real eigenvalues and determinant as a similar symmetric matrix.
- **Theorem**
  For symmetric matrices similar to row normalized matrices where $\nu_{max} = 1$, $(I_N - \rho W)$ is nonsingular for all values of ρ in the interval $(\nu_{min}^{-1}, 1)$.
- **Proof**
  Consider $R = I_N - \rho DW$, where DW is row normalized and non-symmetric. However, in this case DW has the same real eigenvalues as the symmetric but not row normalized matrix $D^{1/2}WD^{1/2}$. Consider the similarity transformation:

$$D^{-1/2}RD^{1/2} = D^{-1/2}(I_N - \rho DW)D^{1/2} = I_N - \rho D^{-1/2}DWD^{1/2} = I_N - \rho D^{1/2}WD^{1/2}$$

  Then $I_N - \rho DW$ and $I_N - \rho D^{1/2}WD^{1/2}$ have then the same real eigenvalues and determinant. As DW is row normalized $\nu_{max} = 1$ and $I_N - \rho D^{1/2}WD^{1/2}$ is nonsingular for all values of ρ in the interval $(\nu_{min}^{-1}, 1)$.

Parameter space: Special Cases 1

- **Similarity**
  Recall that a Matrix B is said to be similar to a matrix A if there exist a nonsingular matrix S such that $B = S^{-1}AS$. Note also that if A and B are similar, then they have the same eigenvalues, counting multiplicity (Horn and Johnson 1985, p.44-45)
- Consider the case where the row normalized matrix is computed from a symmetric matrix, the row normalized matrix is no more symmetric and may have complex eigenvalues. However in this case, it will have the same real eigenvalues and determinant as a similar symmetric matrix.
- **Theorem**
  For symmetric matrices similar to row normalized matrices where $\nu_{max} = 1$, $(I_N - \rho W)$ is nonsingular for all values of ρ in the interval $(\nu_{min}^{-1}, 1)$.
- **Proof**
  Consider $R = I_N - \rho DW$, where DW is row normalized and non-symmetric. However, in this case DW has the same real eigenvalues as the symmetric but not row normalized matrix $D^{1/2}WD^{1/2}$. Consider the similarity transformation:

$$D^{-1/2}RD^{1/2} = D^{-1/2}(I_N - \rho DW)D^{1/2} = I_N - \rho D^{-1/2}DWD^{1/2} = I_N - \rho D^{1/2}WD^{1/2}$$

  Then $I_N - \rho DW$ and $I_N - \rho D^{1/2}WD^{1/2}$ have then the same real eigenvalues and determinant. As DW is row normalized $\nu_{max} = 1$ and $I_N - \rho D^{1/2}WD^{1/2}$ is nonsingular for all values of ρ in the interval $(\nu_{min}^{-1}, 1)$.

Parameter space: Special Cases 2

- **Strictly diagonally dominant matrix**
  Consider again $R = I_N - \rho DW$ where DW is row normalized and non-symmetric and $|\rho| < 1$, then R is strictly diagonally dominant. This means that the diagonal element of R (which equals 1) strictly exceeds the sum of the other elements in the row (which equals ρ since W is row normalized). Strictly diagonally dominant matrices are invertible, therefore R is invertible.
- **Consider the case where W has complex eigenvalues:**
  W is row normalized and is not similar to a symmetric matrix. If a real matrix has complex eigenvalues, these come in complex conjugate pairs. Consider the determinant of $I_N - \rho W$:

$$\det(I_N - \rho W) = \Pi_{i=1}^{N}(1 - \rho\nu_i) = [\Pi_{i=3}^{N}(1 - \rho\nu_i)](1 - \rho\nu_1)(1 - \rho\nu_2)$$

  where, without loss of generality, one of the complex conjugate pairs of eigenvalues appears in ν_1 and the other in ν_2. If the product $(1 - \rho\nu_1)(1 - \rho\nu_2)$ equals 0, this would lead to a zero determinant which would imply that $(I_N - \rho W)$ is singular. The question is then what values of ρ could lead to a singular $(I_N - \rho W)$? Let $\nu_1 = r + ic$ and $\nu_2 = r - ic$ where r is the real part of ν_1 and ν_2, c is the imaginary part of ν_1 and ν_2 and i is the square root of -1, so that $i^2 = -1$. Assume that $c \neq 0$.

Parameter space: Special Cases 2

- **Strictly diagonally dominant matrix**
  Consider again $R = I_N - \rho DW$ where DW is row normalized and non-symmetric and $|\rho| < 1$, then R is strictly diagonally dominant. This means that the diagonal element of R (which equals 1) strictly exceeds the sum of the other elements in the row (which equals ρ since W is row normalized). Strictly diagonally dominant matrices are invertible, therefore R is invertible.
- **Consider the case where W has complex eigenvalues:**
  W is row normalized and is not similar to a symmetric matrix. If a real matrix has complex eigenvalues, these come in complex conjugate pairs. Consider the determinant of $I_N - \rho W$:

$$\det(I_N - \rho W) = \Pi_{i=1}^{N}(1 - \rho\nu_i) = [\Pi_{i=3}^{N}(1 - \rho\nu_i)](1 - \rho\nu_1)(1 - \rho\nu_2)$$

  where, without loss of generality, one of the complex conjugate pairs of eigenvalues appears in ν_1 and the other in ν_2. If the product $(1 - \rho\nu_1)(1 - \rho\nu_2)$ equals 0, this would lead to a zero determinant which would imply that $(I_N - \rho W)$ is singular. The question is then what values of ρ could lead to a singular $(I_N - \rho W)$? Let $\nu_1 = r + ic$ and $\nu_2 = r - ic$ where r is the real part of ν_1 and ν_2, c is the imaginary part of ν_1 and ν_2 and i is the square root of -1, so that $i^2 = -1$. Assume that $c \neq 0$.

Parameter space: Special Cases 3

- Consider now:

$$\begin{aligned}(1-\rho\nu_1)(1-\rho\nu_2) &= 0\\ (1-\rho r-\rho ic)(1-\rho r+\rho ic) &= 0\\ 1-2\rho r+\rho^2 r^2-\rho^2 i^2 c^2 &= 0\\ 1-2\rho r+\rho^2(r^2+c^2) &= 0\end{aligned} \tag{16}$$

  The discriminant of this quadratic equation in ρ is $d = -4c^2 < 0$. Therefore the quadratic equation will have two complex roots. This means that a real ρ cannot result as a root of this quadratic equation. In other words, complex conjugate eigenvalues do not affect whether $(I_N - \rho W)$ is singular. Only pure real eigenvalues can affect the singularity of $(I_N - \rho W)$.

- Consequently, for W with complex eigenvalues, the interval for ρ which guarantees the non singularity of $(I_N - \rho W)$ is $(1/\nu_{min}, 1)$ where ν_{min} is here the most negative purely real eigenvalue of W (LeSage and Pace, 2009, p.88-89).

Parameter space: Special Cases 3

- Consider now:

$$\begin{aligned}(1-\rho\nu_1)(1-\rho\nu_2) &= 0\\ (1-\rho r-\rho ic)(1-\rho r+\rho ic) &= 0\\ 1-2\rho r+\rho^2 r^2-\rho^2 i^2 c^2 &= 0\\ 1-2\rho r+\rho^2(r^2+c^2) &= 0\end{aligned} \tag{16}$$

  The discriminant of this quadratic equation in ρ is $d = -4c^2 < 0$. Therefore the quadratic equation will have two complex roots. This means that a real ρ cannot result as a root of this quadratic equation. In other words, complex conjugate eigenvalues do not affect whether $(I_N - \rho W)$ is singular. Only pure real eigenvalues can affect the singularity of $(I_N - \rho W)$.

- Consequently, for W with complex eigenvalues, the interval for ρ which guarantees the non singularity of $(I_N - \rho W)$ is $(1/\nu_{min}, 1)$ where ν_{min} is here the most negative purely real eigenvalue of W (LeSage and Pace, 2009, p.88-89).

Parameter space

Definite positiveness of $(I - \rho W)$

- We know that:

$$\det(I_N - \rho W) = \Pi_{i=1}^{N}(1 - \rho\nu_i)$$

- It is straightforward to see that $\det(I_N - \rho W) > 0$ if $|\rho| < 1/\tau$, where $\tau = \max\{|\nu_1|, ..., |\nu_N|\}$ as defined in Theorem 1, as well as if $|\rho| < 1/\tau^*$ where $\tau^* = \min(r, c)$ with $r = \max_{1 \leq i \leq N} \sum_{j=1}^{N} |w_{ij}|$ and $c = \max_{1 \leq j \leq N} \sum_{i=1}^{N} |w_{ij}|$ as defined in Theorem 2. It will also be the case if W is a row normalized matrix with $|\rho| < 1$ as in Theorem 3.

Parameter space

Definite positiveness of $(I - \rho W)$

- We know that:
$$\det(I_N - \rho W) = \Pi_{i=1}^{N}(1 - \rho\nu_i)$$
- It is straightforward to see that $\det(I_N - \rho W) > 0$ if $|\rho| < 1/\tau$, where $\tau = \max\{|\nu_1|, ..., |\nu_N|\}$ as defined in Theorem 1, as well as if $|\rho| < 1/\tau^*$ where $\tau^* = \min(r, c)$ with $r = \max_{1 \leq i \leq N} \sum_{j=1}^{N} |w_{ij}|$ and $c = \max_{1 \leq j \leq N} \sum_{i=1}^{N} |w_{ij}|$ as defined in Theorem 2. It will also be the case if W is a row normalized matrix with $|\rho| < 1$ as in Theorem 3.

Estimation of the SAR model

- Let us consider again the first order Spatial Autoregressive Model SAR:

$$y = \rho W y + X\beta + \varepsilon$$

- The variance-covariance matrix for y is easily seen to be equal to:

$$\mathrm{V}(y) = \sigma^2 (I_N - \rho W)^{-1} (I_N - \rho W')^{-1} \tag{17}$$

Note that, in general, this matrix will be full and it's main diagonal will not be constant, inducing complete heteroscedasticity.

- It also follows from the reduced form (8) that the spatially lagged variable Wy is, in general, correlated with the error term since:

$$E[(Wy)'\varepsilon] = \sigma^2 \operatorname{tr} W(I_N - \rho W)^{-1} \neq 0 \tag{18}$$

Therefore OLS estimators will be inconsistent. The simultaneity embedded in the Wy term must be explicitly accounted for in a maximum likelihood estimation framework, as first outlined by Ord (1975).

Estimation of the SAR model

- Let us consider again the first order Spatial Autoregressive Model SAR:

$$y = \rho W y + X\beta + \varepsilon$$

- The variance-covariance matrix for y is easily seen to be equal to:

$$\mathrm{V}(y) = \sigma^2 (I_N - \rho W)^{-1} (I_N - \rho W')^{-1} \tag{17}$$

Note that, in general, this matrix will be full and it's main diagonal will not be constant, inducing complete heteroscedasticity.

- It also follows from the reduced form (8) that the spatially lagged variable Wy is, in general, correlated with the error term since:

$$E[(Wy)'\varepsilon] = \sigma^2 \operatorname{tr} W(I_N - \rho W)^{-1} \neq 0 \tag{18}$$

Therefore OLS estimators will be inconsistent. The simultaneity embedded in the Wy term must be explicitly accounted for in a maximum likelihood estimation framework, as first outlined by Ord (1975).

Estimation of the SAR model

- Let us consider again the first order Spatial Autoregressive Model SAR:

$$y = \rho W y + X\beta + \varepsilon$$

- The variance-covariance matrix for y is easily seen to be equal to:

$$\mathrm{V}(y) = \sigma^2 (I_N - \rho W)^{-1} (I_N - \rho W')^{-1} \tag{17}$$

  Note that, in general, this matrix will be full and it's main diagonal will not be constant, inducing complete heteroscedasticity.

- It also follows from the reduced form (8) that the spatially lagged variable Wy is, in general, correlated with the error term since:

$$E[(Wy)'\varepsilon] = \sigma^2 \operatorname{tr} W(I_N - \rho W)^{-1} \neq 0 \tag{18}$$

  Therefore OLS estimators will be inconsistent. The simultaneity embedded in the Wy term must be explicitly accounted for in a maximum likelihood estimation framework, as first outlined by Ord (1975).

Estimation of the SAR model

Log-likelihood function

- Under the additional hypothesis of normality of the error term, $\varepsilon \sim N(0, \sigma_\epsilon^2 I_N)$, the log-likelihood function for the SAR model is given by:

$$\begin{aligned} \ln L(\beta', \rho, \sigma_\epsilon^2) &= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma_\epsilon^2) + \ln|I - \rho W| \\ &\quad -\frac{1}{2\sigma_\epsilon^2}\left[(I - \rho W)y - X\beta\right]'\left[(I - \rho W)y - X\beta\right] \end{aligned} \tag{19}$$

- An important aspect of this log-likelihood function is the Jacobian of the transformation, which is the determinant of the $(N \times N)$ *full* matrix $(I_N - \rho W)$ for our model.
- Recall that moving from the likelihood for the Normal error vector to the likelihood for the observed dependent variable, the Jacobian of the transformation needs to be inserted. In the SAR model, this corresponds to:

$$\left|\frac{\partial \varepsilon}{\partial y}\right| = \left|\frac{\partial(y - \rho W y - X\beta)}{\partial y}\right| = |I_N - \rho W| \tag{20}$$

Estimation of the SAR model

Log-likelihood function

- Under the additional hypothesis of normality of the error term, $\varepsilon \sim N(0, \sigma_\epsilon^2 I_N)$, the log-likelihood function for the SAR model is given by:

$$\begin{aligned} \ln L(\beta', \rho, \sigma_\epsilon^2) &= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma_\epsilon^2) + \ln|I - \rho W| \\ &\quad -\frac{1}{2\sigma_\epsilon^2}\left[(I-\rho W)y - X\beta\right]'\left[(I-\rho W)y - X\beta\right] \end{aligned} \tag{19}$$

- An important aspect of this log-likelihood function is the Jacobian of the transformation, which is the determinant of the $(N \times N)$ *full* matrix $(I_N - \rho W)$ for our model.
- Recall that moving from the likelihood for the Normal error vector to the likelihood for the observed dependent variable, the Jacobian of the transformation needs to be inserted. In the SAR model, this corresponds to:

$$\left|\frac{\partial \varepsilon}{\partial y}\right| = \left|\frac{\partial(y - \rho W y - X\beta)}{\partial y}\right| = |I_N - \rho W| \tag{20}$$

Estimation of the SAR model

Log-likelihood function

- Under the additional hypothesis of normality of the error term, $\varepsilon \sim N(0, \sigma_\epsilon^2 I_N)$, the log-likelihood function for the SAR model is given by:

$$\begin{aligned} \ln L(\beta', \rho, \sigma_\epsilon^2) &= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma_\epsilon^2) + \ln|I - \rho W| \\ &\quad -\frac{1}{2\sigma_\epsilon^2}\left[(I-\rho W)y - X\beta\right]'\left[(I-\rho W)y - X\beta\right] \end{aligned} \tag{19}$$

- An important aspect of this log-likelihood function is the Jacobian of the transformation, which is the determinant of the $(N \times N)$ *full* matrix $(I_N - \rho W)$ for our model.
- Recall that moving from the likelihood for the Normal error vector to the likelihood for the observed dependent variable, the Jacobian of the transformation needs to be inserted. In the SAR model, this corresponds to:

$$\left|\frac{\partial \varepsilon}{\partial y}\right| = \left|\frac{\partial(y - \rho W y - X\beta)}{\partial y}\right| = |I_N - \rho W| \tag{20}$$

Estimation of the SAR model

First order conditions

- Note that this Jacobian reduces to a scalar 1 in the standard regression model, since $|\partial(y - X\beta)/\partial y| = |I_N| = 1$. Maximizing the log-likelihood function (19) is therefore not equivalent to minimizing weighted least squares, i.e. the last term in the log-likelihood function, as in the standard linear regression model since it ignores the Jacobian term.
- The log-Jacobian also implies constraints on the parameter space for ρ, which must be such that $|I_N - \rho W| > 0$.
- Let us write the usual first order conditions for the maximization of the log-likelihood function (19):

$$\begin{aligned}
\frac{\partial \ln L}{\partial \beta'} &= X'(I_N - \rho W)y - X'X\beta = 0 \qquad (21)\\
\frac{\partial \ln L}{\partial \rho} &= -\mathrm{tr}[W(I_N - \rho W)^{-1}] + \frac{1}{\sigma_\varepsilon^2}[(I_N - \rho W)y - X\beta]'Wy = 0\\
\frac{\partial \ln L}{\partial \sigma_\varepsilon^2} &= -\frac{N}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\varepsilon^4}[(I_N - \rho W)y - X\beta]'[(I_N - \rho W)y - X\beta] = 0
\end{aligned}$$

Estimation of the SAR model

First order conditions

- Note that this Jacobian reduces to a scalar 1 in the standard regression model, since $|\partial(y - X\beta)/\partial y| = |I_N| = 1$. Maximizing the log-likelihood function (19) is therefore not equivalent to minimizing weighted least squares, i.e. the last term in the log-likelihood function, as in the standard linear regression model since it ignores the Jacobian term.
- The log-Jacobian also implies constraints on the parameter space for ρ, which must be such that $|I_N - \rho W| > 0$.
- Let us write the usual first order conditions for the maximization of the log-likelihood function (19):

$$\begin{aligned}
\frac{\partial \ln L}{\partial \beta'} &= X'(I_N - \rho W)y - X'X\beta = 0 \qquad (21)\\
\frac{\partial \ln L}{\partial \rho} &= -\mathrm{tr}[W(I_N - \rho W)^{-1}] + \frac{1}{\sigma_\varepsilon^2}[(I_N - \rho W)y - X\beta]'Wy = 0\\
\frac{\partial \ln L}{\partial \sigma_\varepsilon^2} &= -\frac{N}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\varepsilon^4}[(I_N - \rho W)y - X\beta]'[(I_N - \rho W)y - X\beta] = 0
\end{aligned}$$

Estimation of the SAR model

First order conditions

- Note that this Jacobian reduces to a scalar 1 in the standard regression model, since $|\partial(y - X\beta)/\partial y| = |I_N| = 1$. Maximizing the log-likelihood function (19) is therefore not equivalent to minimizing weighted least squares, i.e. the last term in the log-likelihood function, as in the standard linear regression model since it ignores the Jacobian term.
- The log-Jacobian also implies constraints on the parameter space for ρ, which must be such that $|I_N - \rho W| > 0$.
- Let us write the usual first order conditions for the maximization of the log-likelihood function (19):

$$\begin{aligned}
\frac{\partial \ln L}{\partial \beta'} &= X'(I_N - \rho W)y - X'X\beta = 0 \qquad (21)\\
\frac{\partial \ln L}{\partial \rho} &= -\mathrm{tr}[W(I_N - \rho W)^{-1}] + \frac{1}{\sigma_\varepsilon^2}[(I_N - \rho W)y - X\beta]'Wy = 0\\
\frac{\partial \ln L}{\partial \sigma_\varepsilon^2} &= -\frac{N}{\sigma_\varepsilon^2} + \frac{1}{\sigma_\varepsilon^4}[(I_N - \rho W)y - X\beta]'[(I_N - \rho W)y - X\beta] = 0
\end{aligned}$$

Estimation of the SAR model

More on the Jacobian problem

- Note that here we have used the following result to compute the partial derivative of the determinant of $I - \rho W$ with respect to ρ:

$$\begin{aligned} \frac{\partial ln|I_N - \rho W|}{\partial \rho} &= \text{tr}(I_N - \rho W)^{-1} \frac{\partial (I_N - \rho W)}{\partial \rho} \\ &= \text{tr}(I_N - \rho W)^{-1}(-W) \\ &= -\text{tr}[W(I_N - \rho W)^{-1}] \end{aligned} \tag{22}$$

- The presence of the Jacobian term could complicate the computation of the maximum likelihood estimators which involves the repeated evaluation of this determinant. However Ord (1975) suggests that it can be expressed as a function of the eigenvalues ω_i of the spatial weights matrix W:

$$|I_N - \rho W| = \prod_{i=1}^{N}(1 - \rho\omega_i) \quad \Longrightarrow \quad \ln|I_N - \rho W| = \sum_{i=1}^{N} \ln(1 - \rho\omega_i) \tag{23}$$

This expression simplifies considerably the computations since the eigenvalues of W have to be computed only once at the outset of the numerical optimization procedure.

Estimation of the SAR model

More on the Jacobian problem

- Note that here we have used the following result to compute the partial derivative of the determinant of $I - \rho W$ with respect to ρ:

$$\begin{aligned} \frac{\partial ln|I_N - \rho W|}{\partial \rho} &= \text{tr}(I_N - \rho W)^{-1} \frac{\partial (I_N - \rho W)}{\partial \rho} \\ &= \text{tr}(I_N - \rho W)^{-1}(-W) \\ &= -\text{tr}[W(I_N - \rho W)^{-1}] \end{aligned} \tag{22}$$

- The presence of the Jacobian term could complicate the computation of the maximum likelihood estimators which involves the repeated evaluation of this determinant. However Ord (1975) suggests that it can be expressed as a function of the eigenvalues ω_i of the spatial weights matrix W:

$$|I_N - \rho W| = \prod_{i=1}^{N}(1 - \rho\omega_i) \quad \Longrightarrow \quad \ln|I_N - \rho W| = \sum_{i=1}^{N} \ln(1 - \rho\omega_i) \tag{23}$$

This expression simplifies considerably the computations since the eigenvalues of W have to be computed only once at the outset of the numerical optimization procedure.

Estimation of the SAR model

ML estimators given ρ

- From the usual first-order conditions, the maximum likelihood estimators of β and σ^2, given ρ, are obtained as:

$$\hat{\beta}_{ML}(\rho) = (X'X)^{-1}X'(I-\rho W)y \tag{24}$$

$$\hat{\sigma}^2_{ML}(\rho) = \frac{1}{N}\left[(I-\rho W)y - X\hat{\beta}_{ML}(\rho)\right]'\left[(I-\rho W)y - X\hat{\beta}_{ML}(\rho)\right] \tag{25}$$

- Note that, for convenience:

$$\hat{\beta}_{ML}(\rho) = \hat{\beta}_O - \rho\hat{\beta}_L \tag{26}$$

where $\hat{\beta}_O = (X'X)^{-1}X'y$ and $\hat{\beta}_L = (X'X)^{-1}X'Wy$. Define $\hat{e}_O = y - X\hat{\beta}_O$ and $\hat{e}_L = Wy - X\hat{\beta}_L$, it can be then easily seen that:

$$\hat{\sigma}^2_{ML}(\rho) = \left[\frac{(\hat{e}_O - \rho\hat{e}_L)'(\hat{e}_O - \rho\hat{e}_L)}{N}\right] \tag{27}$$

Estimation of the SAR model

ML estimators given ρ

- From the usual first-order conditions, the maximum likelihood estimators of β and σ^2, given ρ, are obtained as:

$$\hat{\beta}_{ML}(\rho) = (X'X)^{-1}X'(I-\rho W)y \tag{24}$$

$$\hat{\sigma}^2_{ML}(\rho) = \frac{1}{N}\left[(I-\rho W)y - X\hat{\beta}_{ML}(\rho)\right]'\left[(I-\rho W)y - X\hat{\beta}_{ML}(\rho)\right] \tag{25}$$

- Note that, for convenience:

$$\hat{\beta}_{ML}(\rho) = \hat{\beta}_O - \rho\hat{\beta}_L \tag{26}$$

where $\hat{\beta}_O = (X'X)^{-1}X'y$ and $\hat{\beta}_L = (X'X)^{-1}X'Wy$. Define $\hat{e}_O = y - X\hat{\beta}_O$ and $\hat{e}_L = Wy - X\hat{\beta}_L$, it can be then easily seen that:

$$\hat{\sigma}^2_{ML}(\rho) = \left[\frac{(\hat{e}_O - \rho\hat{e}_L)'(\hat{e}_O - \rho\hat{e}_L)}{N}\right] \tag{27}$$

Estimation of the SAR model

Concentrated log-likelihood function

- Substitution of (26) and (27) in the log-likelihood function (19) yields a concentrated log-likelihood as a non-linear function of a single parameter ρ:

$$\ln L(\rho) = -\frac{N}{2}[\ln(2\pi) + 1] + \sum_{i=1}^{N} \ln(1 - \rho\omega_i) - \frac{n}{2} \ln \left[\frac{(\hat{e}_O - \rho\hat{e}_L)'(\hat{e}_O - \rho\hat{e}_L)}{N} \right]$$

  where $\hat{e}_O$ and $\hat{e}_L$ are the estimated residuals in a regression of y on X and Wy on X, respectively.
- A maximum likelihood estimate for ρ is obtained from a numerical optimization of the concentrated log-likelihood function.
- The estimation procedure can then be described as follows:
  - We begin by first regressing by OLS y on X which yields $\hat{\beta}_O$ and Wy on X which yields $\hat{\beta}_L$.
  - We then compute the estimated residuals $\hat{e}_O$ and $\hat{e}_L$.
  - Given those, we maximize, using a numerical nonlinear optimization routine, the concentrated log-likelihood function to find $\hat{\rho}_{ML}$, until the numerical convergence criterion is met.
  - Given $\hat{\rho}_{ML}$ we can then compute $\hat{\beta}_{ML} = \hat{\beta}_O - \hat{\rho}\hat{\beta}_L$ and $\hat{\sigma}^2_{ML} = [(1/N)(\hat{e}_O - \hat{\rho}\hat{e}_L)'(\hat{e}_O - \hat{\rho}\hat{e}_L)]$.

Estimation of the SAR model

Concentrated log-likelihood function

- Substitution of (26) and (27) in the log-likelihood function (19) yields a concentrated log-likelihood as a non-linear function of a single parameter ρ:

$$\ln L(\rho) = -\frac{N}{2}[\ln(2\pi) + 1] + \sum_{i=1}^{N} \ln(1 - \rho\omega_i) - \frac{n}{2} \ln \left[\frac{(\hat{e}_O - \rho\hat{e}_L)'(\hat{e}_O - \rho\hat{e}_L)}{N} \right]$$

  where $\hat{e}_O$ and $\hat{e}_L$ are the estimated residuals in a regression of y on X and Wy on X, respectively.
- A maximum likelihood estimate for ρ is obtained from a numerical optimization of the concentrated log-likelihood function.
- The estimation procedure can then be described as follows:
  - We begin by first regressing by OLS y on X which yields $\hat{\beta}_O$ and Wy on X which yields $\hat{\beta}_L$.
  - We then compute the estimated residuals $\hat{e}_O$ and $\hat{e}_L$.
  - Given those, we maximize, using a numerical nonlinear optimization routine, the concentrated log-likelihood function to find $\hat{\rho}_{ML}$, until the numerical convergence criterion is met.
  - Given $\hat{\rho}_{ML}$ we can then compute $\hat{\beta}_{ML} = \hat{\beta}_O - \hat{\rho}\hat{\beta}_L$ and $\hat{\sigma}^2_{ML} = [(1/N)(\hat{e}_O - \hat{\rho}\hat{e}_L)'(\hat{e}_O - \hat{\rho}\hat{e}_L)]$.

Estimation of the SAR model

Concentrated log-likelihood function

- Substitution of (26) and (27) in the log-likelihood function (19) yields a concentrated log-likelihood as a non-linear function of a single parameter ρ:

$$\ln L(\rho) = -\frac{N}{2}[\ln(2\pi) + 1] + \sum_{i=1}^{N} \ln(1 - \rho\omega_i) - \frac{n}{2} \ln \left[\frac{(\hat{e}_O - \rho\hat{e}_L)'(\hat{e}_O - \rho\hat{e}_L)}{N} \right]$$

  where $\hat{e}_O$ and $\hat{e}_L$ are the estimated residuals in a regression of y on X and Wy on X, respectively.
- A maximum likelihood estimate for ρ is obtained from a numerical optimization of the concentrated log-likelihood function.
- The estimation procedure can then be described as follows:
  - We begin by first regressing by OLS y on X which yields $\hat{\beta}_O$ and Wy on X which yields $\hat{\beta}_L$.
  - We then compute the estimated residuals $\hat{e}_O$ and $\hat{e}_L$.
  - Given those, we maximize, using a numerical nonlinear optimization routine, the concentrated log-likelihood function to find $\hat{\rho}_{ML}$, until the numerical convergence criterion is met.
  - Given $\hat{\rho}_{ML}$ we can then compute $\hat{\beta}_{ML} = \hat{\beta}_O - \hat{\rho}\hat{\beta}_L$ and $\hat{\sigma}^2_{ML} = [(1/N)(\hat{e}_O - \hat{\rho}\hat{e}_L)'(\hat{e}_O - \hat{\rho}\hat{e}_L)]$.

Estimation of the SAR model

Concentrated log-likelihood function

- Substitution of (26) and (27) in the log-likelihood function (19) yields a concentrated log-likelihood as a non-linear function of a single parameter ρ:

$$\ln L(\rho) = -\frac{N}{2}[\ln(2\pi) + 1] + \sum_{i=1}^{N} \ln(1 - \rho\omega_i) - \frac{n}{2} \ln \left[\frac{(\hat{e}_O - \rho\hat{e}_L)'(\hat{e}_O - \rho\hat{e}_L)}{N} \right]$$

  where $\hat{e}_O$ and $\hat{e}_L$ are the estimated residuals in a regression of y on X and Wy on X, respectively.
- A maximum likelihood estimate for ρ is obtained from a numerical optimization of the concentrated log-likelihood function.
- The estimation procedure can then be described as follows:
  - We begin by first regressing by OLS y on X which yields $\hat{\beta}_O$ and Wy on X which yields $\hat{\beta}_L$.
  - We then compute the estimated residuals $\hat{e}_O$ and $\hat{e}_L$.
  - Given those, we maximize, using a numerical nonlinear optimization routine, the concentrated log-likelihood function to find $\hat{\rho}_{ML}$, until the numerical convergence criterion is met.
  - Given $\hat{\rho}_{ML}$ we can then compute $\hat{\beta}_{ML} = \hat{\beta}_O - \hat{\rho}\hat{\beta}_L$ and $\hat{\sigma}^2_{ML} = [(1/N)(\hat{e}_O - \hat{\rho}\hat{e}_L)'(\hat{e}_O - \hat{\rho}\hat{e}_L)]$.

Estimation of the SAR model

Concentrated log-likelihood function

- Substitution of (26) and (27) in the log-likelihood function (19) yields a concentrated log-likelihood as a non-linear function of a single parameter ρ:

$$\ln L(\rho) = -\frac{N}{2}[\ln(2\pi) + 1] + \sum_{i=1}^{N} \ln(1 - \rho\omega_i) - \frac{n}{2} \ln \left[\frac{(\hat{e}_O - \rho\hat{e}_L)'(\hat{e}_O - \rho\hat{e}_L)}{N} \right]$$

  where $\hat{e}_O$ and $\hat{e}_L$ are the estimated residuals in a regression of y on X and Wy on X, respectively.
- A maximum likelihood estimate for ρ is obtained from a numerical optimization of the concentrated log-likelihood function.
- The estimation procedure can then be described as follows:
  - We begin by first regressing by OLS y on X which yields $\hat{\beta}_O$ and Wy on X which yields $\hat{\beta}_L$.
  - We then compute the estimated residuals $\hat{e}_O$ and $\hat{e}_L$.
  - Given those, we maximize, using a numerical nonlinear optimization routine, the concentrated log-likelihood function to find $\hat{\rho}_{ML}$, until the numerical convergence criterion is met.
  - Given $\hat{\rho}_{ML}$ we can then compute $\hat{\beta}_{ML} = \hat{\beta}_O - \hat{\rho}\hat{\beta}_L$ and $\hat{\sigma}^2_{ML} = [(1/N)(\hat{e}_O - \hat{\rho}\hat{e}_L)'(\hat{e}_O - \hat{\rho}\hat{e}_L)]$.

Estimation of the SAR model

Concentrated log-likelihood function

- Substitution of (26) and (27) in the log-likelihood function (19) yields a concentrated log-likelihood as a non-linear function of a single parameter ρ:

$$\ln L(\rho) = -\frac{N}{2}[\ln(2\pi) + 1] + \sum_{i=1}^{N} \ln(1 - \rho\omega_i) - \frac{n}{2} \ln \left[\frac{(\hat{e}_O - \rho\hat{e}_L)'(\hat{e}_O - \rho\hat{e}_L)}{N} \right]$$

  where $\hat{e}_O$ and $\hat{e}_L$ are the estimated residuals in a regression of y on X and Wy on X, respectively.
- A maximum likelihood estimate for ρ is obtained from a numerical optimization of the concentrated log-likelihood function.
- The estimation procedure can then be described as follows:
  - We begin by first regressing by OLS y on X which yields $\hat{\beta}_O$ and Wy on X which yields $\hat{\beta}_L$.
  - We then compute the estimated residuals $\hat{e}_O$ and $\hat{e}_L$.
  - Given those, we maximize, using a numerical nonlinear optimization routine, the concentrated log-likelihood function to find $\hat{\rho}_{ML}$, until the numerical convergence criterion is met.
  - Given $\hat{\rho}_{ML}$ we can then compute $\hat{\beta}_{ML} = \hat{\beta}_O - \hat{\rho}\hat{\beta}_L$ and $\hat{\sigma}^2_{ML} = [(1/N)(\hat{e}_O - \hat{\rho}\hat{e}_L)'(\hat{e}_O - \hat{\rho}\hat{e}_L)]$.

Estimation of the SAR model

Concentrated log-likelihood function

- Substitution of (26) and (27) in the log-likelihood function (19) yields a concentrated log-likelihood as a non-linear function of a single parameter ρ:

$$\ln L(\rho) = -\frac{N}{2}[\ln(2\pi) + 1] + \sum_{i=1}^{N} \ln(1 - \rho\omega_i) - \frac{n}{2} \ln \left[\frac{(\hat{e}_O - \rho\hat{e}_L)'(\hat{e}_O - \rho\hat{e}_L)}{N} \right]$$

  where $\hat{e}_O$ and $\hat{e}_L$ are the estimated residuals in a regression of y on X and Wy on X, respectively.
- A maximum likelihood estimate for ρ is obtained from a numerical optimization of the concentrated log-likelihood function.
- The estimation procedure can then be described as follows:
  - We begin by first regressing by OLS y on X which yields $\hat{\beta}_O$ and Wy on X which yields $\hat{\beta}_L$.
  - We then compute the estimated residuals $\hat{e}_O$ and $\hat{e}_L$.
  - Given those, we maximize, using a numerical nonlinear optimization routine, the concentrated log-likelihood function to find $\hat{\rho}_{ML}$, until the numerical convergence criterion is met.
  - Given $\hat{\rho}_{ML}$ we can then compute $\hat{\beta}_{ML} = \hat{\beta}_O - \hat{\rho}\hat{\beta}_L$ and $\hat{\sigma}^2_{ML} = [(1/N)(\hat{e}_O - \hat{\rho}\hat{e}_L)'(\hat{e}_O - \hat{\rho}\hat{e}_L)]$.

Estimation of the SAR model

Asymptotic variance-covariance matrix

- Under the regularity conditions described for instance in Lee (2004, p.1902-1904), it can be shown that the maximum likelihood estimators have the usual asymptotic properties, including consistency, normality, and asymptotic efficiency.
- The asymptotic variance-covariance matrix follows as the inverse of the information matrix. Defining $W_A = W(I - \rho W)^{-1}$ to simplify notation, we have then:
$\text{AsyVar}[\beta', \rho, \sigma^2] =$

$$\begin{bmatrix} \frac{1}{\sigma^2} X'X & \frac{1}{\sigma^2}(X'W_A X\beta)' & 0 \\ \frac{1}{\sigma^2} X'W_A X\beta & tr\left[(W_A + W_A')W_A\right] + \frac{1}{\sigma^2}(W_A X\beta)'(W_A X\beta) & \frac{1}{\sigma^2} trW_A \\ 0 & \frac{1}{\sigma^2} trW_A & \frac{n}{2\sigma^4} \end{bmatrix}^{-1}$$

- The estimated asymptotic variance-covariance matrix will then be used for asymptotic inference on the parameters under the regularity conditions.

Estimation of the SAR model

Asymptotic variance-covariance matrix

- Under the regularity conditions described for instance in Lee (2004, p.1902-1904), it can be shown that the maximum likelihood estimators have the usual asymptotic properties, including consistency, normality, and asymptotic efficiency.
- The asymptotic variance-covariance matrix follows as the inverse of the information matrix. Defining $W_A = W(I - \rho W)^{-1}$ to simplify notation, we have then:
  $\text{AsyVar}[\beta', \rho, \sigma^2] =$

$$\begin{bmatrix} \frac{1}{\sigma^2} X'X & \frac{1}{\sigma^2}(X'W_A X\beta)' & 0 \\ \frac{1}{\sigma^2} X'W_A X\beta & tr\left[(W_A + W_A')W_A\right] + \frac{1}{\sigma^2}(W_A X\beta)'(W_A X\beta) & \frac{1}{\sigma^2} trW_A \\ 0 & \frac{1}{\sigma^2} trW_A & \frac{n}{2\sigma^4} \end{bmatrix}^{-1}$$

- The estimated asymptotic variance-covariance matrix will then be used for asymptotic inference on the parameters under the regularity conditions.

Estimation of the SAR model

Asymptotic variance-covariance matrix

- Under the regularity conditions described for instance in Lee (2004, p.1902-1904), it can be shown that the maximum likelihood estimators have the usual asymptotic properties, including consistency, normality, and asymptotic efficiency.
- The asymptotic variance-covariance matrix follows as the inverse of the information matrix. Defining $W_A = W(I - \rho W)^{-1}$ to simplify notation, we have then:
  $\text{AsyVar}[\beta', \rho, \sigma^2] =$

$$\begin{bmatrix} \frac{1}{\sigma^2} X'X & \frac{1}{\sigma^2}(X'W_A X\beta)' & 0 \\ \frac{1}{\sigma^2} X'W_A X\beta & tr\left[(W_A + W_A')W_A\right] + \frac{1}{\sigma^2}(W_A X\beta)'(W_A X\beta) & \frac{1}{\sigma^2} trW_A \\ 0 & \frac{1}{\sigma^2} trW_A & \frac{n}{2\sigma^4} \end{bmatrix}^{-1}$$

- The estimated asymptotic variance-covariance matrix will then be used for asymptotic inference on the parameters under the regularity conditions.

SEM specification: the spatial error model

Reduced form

- Let us consider the following cross section regression model with spatially autocorrelated errors:

$$y = X\beta + \varepsilon \qquad \varepsilon = \lambda W\varepsilon + u \tag{28}$$

- **Reduced form**

  If $I_N - \lambda W$ is invertible,noting that $\varepsilon = (I_N - \lambda W)^{-1}u$, the reduced form of the SEM specification is obtained as:

$$y = X\beta + (I_N - \lambda W)^{-1}u \tag{29}$$

  This reduced form implies that a random shock in a specific cross-sectional unit *i* does not only affect the value of the dependent variable in the same unit, but also has an impact on the values of dependent variable in all other cross-sectional units through the inverse spatial transformation $(I_N - \lambda W)^{-1}$. This is the so-called *spatial "diffusion" process of random shocks*. The SEM specification shares this property with the SAR model, but does not share the *global spatial multiplier effect* that we highlighted for the SAR process.

SEM specification: the spatial error model

Reduced form

- Let us consider the following cross section regression model with spatially autocorrelated errors:

$$y = X\beta + \varepsilon \qquad \varepsilon = \lambda W\varepsilon + u \tag{28}$$

- **Reduced form**
  If $I_N - \lambda W$ is invertible,noting that $\varepsilon = (I_N - \lambda W)^{-1}u$, the reduced form of the SEM specification is obtained as:

$$y = X\beta + (I_N - \lambda W)^{-1}u \tag{29}$$

  This reduced form implies that a random shock in a specific cross-sectional unit *i* does not only affect the value of the dependent variable in the same unit, but also has an impact on the values of dependent variable in all other cross-sectional units through the inverse spatial transformation $(I_N - \lambda W)^{-1}$. This is the so-called *spatial "diffusion" process of random shocks*. The SEM specification shares this property with the SAR model, but does not share the *global spatial multiplier effect* that we highlighted for the SAR process.

SEM specification: the spatial error model

Constrained Spatial Durbin Model

- Note that this model may also be interpreted as a constrained Spatial Durbin Model:

$$(I_N - \lambda W)y = (I_N - \lambda W)(\beta_0 \iota_N + X\beta) + u \tag{30}$$

$$y = (I_N - \lambda W)\beta_0 \iota_N + \lambda W y + X\beta - \lambda WX\beta + u \tag{31}$$

- Using a row-normalized interaction matrix W, one gets:

$$y = \frac{\beta_0}{1-\lambda}\iota_N + \lambda W y + X\beta - \lambda WX\beta + u \tag{32}$$

which is a Spatial Durbin Model:

$$y = \frac{\beta_0}{1-\lambda}\iota_N + \lambda W y + X\beta + WX\gamma + \varepsilon \tag{33}$$

with non linear constraints $\gamma = -\lambda\beta$.

SEM specification: the spatial error model

Constrained Spatial Durbin Model

- Note that this model may also be interpreted as a constrained Spatial Durbin Model:

$$(I_N - \lambda W)y = (I_N - \lambda W)(\beta_0 \iota_N + X\beta) + u \tag{30}$$

$$y = (I_N - \lambda W)\beta_0 \iota_N + \lambda W y + X\beta - \lambda WX\beta + u \tag{31}$$

- Using a row-normalized interaction matrix W, one gets:

$$y = \frac{\beta_0}{1-\lambda}\iota_N + \lambda W y + X\beta - \lambda WX\beta + u \tag{32}$$

which is a Spatial Durbin Model:

$$y = \frac{\beta_0}{1-\lambda}\iota_N + \lambda W y + X\beta + WX\gamma + \varepsilon \tag{33}$$

with non linear constraints $\gamma = -\lambda\beta$.

SEM specification: the spatial error model

Spatial filter and variance-covariance matrix

- **Spatial filter**
  If λ was known, we could filter out spatial autocorrelation from both of y and X and then use the classical regression model as follows:

$$(I_N - \lambda W)y = (I_N - \lambda W)X\beta + u \tag{34}$$

  Of course λ_N is usually unknown and has to be estimated as well as β_N and $\sigma^2_{u,N}$.

- **Variance-covariance matrix**
  Assuming that X is non stochastic, the mathematical expectation of y is trivially $\mathrm{E}(y) = X\beta$. Its variance-covariance matrix is in turn obtained as:

$$\mathrm{V}(y) = \mathrm{V}(\varepsilon) = \sigma^2_u (I_N - \lambda W)^{-1}(I_N - \lambda W)'^{-1} = \sigma^2_u \Omega_\varepsilon(\lambda) \tag{35}$$

  This is exactly the result we obtained for the SAR model. Note again that in general, this matrix will be full and it's main diagonal will not be constant. Thus spatial autocorrelation in the error term induces complete heteroscedasticity. This a case of non spherical error term, where OLS estimators are consistent but inefficient.

SEM specification: the spatial error model

Spatial filter and variance-covariance matrix

- **Spatial filter**
  If λ was known, we could filter out spatial autocorrelation from both of y and X and then use the classical regression model as follows:

$$(I_N - \lambda W)y = (I_N - \lambda W)X\beta + u \tag{34}$$

  Of course λ_N is usually unknown and has to be estimated as well as β_N and $\sigma^2_{u,N}$.

- **Variance-covariance matrix**
  Assuming that X is non stochastic, the mathematical expectation of y is trivially $\mathrm{E}(y) = X\beta$. Its variance-covariance matrix is in turn obtained as:

$$\mathrm{V}(y) = \mathrm{V}(\varepsilon) = \sigma^2_u(I_N - \lambda W)^{-1}(I_N - \lambda W)'^{-1} = \sigma^2_u\Omega_\varepsilon(\lambda) \tag{35}$$

  This is exactly the result we obtained for the SAR model. Note again that in general, this matrix will be full and it's main diagonal will not be constant. Thus spatial autocorrelation in the error term induces complete heteroscedasticity. This a case of non spherical error term, where OLS estimators are consistent but inefficient.

Estimation of the SEM

GLS and FGLS estimation

- The GLS estimator is given by:

$$\hat{\beta}_{GLS} = [X'\Omega_{\varepsilon}^{-1}(\lambda)X]^{-1}X'\Omega_{\varepsilon}^{-1}(\lambda)y \tag{36}$$

Clearly, this estimator would be BLUE if λ were known, we would have:

$$\mathrm{E}(\hat{\beta}_{GLS}) = \beta \text{ and } \mathrm{V}(\hat{\beta}_{GLS}) = \sigma_u^2[X'\Omega_{\varepsilon}^{-1}(\lambda)X]^{-1} \tag{37}$$

Of course as λ is unknown, the GLS estimator is not feasible.

- Nevertheless let's have a closer look at this estimator:

$$\begin{aligned}
\hat{\beta}_{GLS} &= [X'(I_N - \lambda W)'(I_N - \lambda W)X]^{-1}X'(I_N - \lambda W)'(I_N - \lambda W)y \\
\hat{\beta}_{GLS} &= [(X - \lambda WX)'(X - \lambda WX)]^{-1}(X - \lambda WX)'(y - \lambda Wy) \\
\hat{\beta}_{GLS} &= (X_L'X_L)^{-1}X_L'y_L
\end{aligned}$$

where $X_L = (I_N - \lambda W)X = X - \lambda WX$ and $y_L = (I_N - \lambda W)y = y - \lambda Wy$ may be interpreted as spatial counterparts to the Cochrane-Orcutt transformation in Times Series. The GLS estimator is thus the OLS estimator on spatially filtered variables, if λ were known.

Estimation of the SEM

GLS and FGLS estimation

- The GLS estimator is given by:

$$\hat{\beta}_{GLS} = [X'\Omega_{\varepsilon}^{-1}(\lambda)X]^{-1}X'\Omega_{\varepsilon}^{-1}(\lambda)y \tag{36}$$

Clearly, this estimator would be BLUE if λ were known, we would have:

$$\mathrm{E}(\hat{\beta}_{GLS}) = \beta \text{ and } \mathrm{V}(\hat{\beta}_{GLS}) = \sigma_u^2[X'\Omega_{\varepsilon}^{-1}(\lambda)X]^{-1} \tag{37}$$

Of course as λ is unknown, the GLS estimator is not feasible.

- Nevertheless let's have a closer look at this estimator:

$$\begin{aligned}
\hat{\beta}_{GLS} &= [X'(I_N - \lambda W)'(I_N - \lambda W)X]^{-1}X'(I_N - \lambda W)'(I_N - \lambda W)y \\
\hat{\beta}_{GLS} &= [(X - \lambda WX)'(X - \lambda WX)]^{-1}(X - \lambda WX)'(y - \lambda Wy) \\
\hat{\beta}_{GLS} &= (X_L'X_L)^{-1}X_L'y_L
\end{aligned}$$

where $X_L = (I_N - \lambda W)X = X - \lambda WX$ and $y_L = (I_N - \lambda W)y = y - \lambda Wy$ may be interpreted as spatial counterparts to the Cochrane-Orcutt transformation in Times Series. The GLS estimator is thus the OLS estimator on spatially filtered variables, if λ were known.

Estimation of the SEM

GLS and FGLS estimation

- Using the feasible GLS estimator requires to have a consistent estimator for λ that could be used to transform the model by the spatial Cochrane-Orcutt procedure and then estimate β by OLS:

$$\hat{\beta}_{FGLS} = [X'\Omega_{\varepsilon}^{-1}(\hat{\lambda})X]^{-1}X'\Omega_{\varepsilon}^{-1}(\hat{\lambda})y \quad (38)$$

$$\hat{\beta}_{FGLS} = (\tilde{X}_L'\tilde{X}_L)^{-1}\tilde{X}_L'\tilde{y}_L \quad (39)$$

  where $\tilde{X}_L = X - \hat{\lambda}WX$ and $\tilde{y}_L = y - \hat{\lambda}Wy$.

- The OLS estimator for β is consistent in the spatially autocorrelated error model, the errors can then be estimated consistently. However the OLS estimator for λ in the SAR model for the errors is not consistent as we previously showed, but Kelejian and Prucha (1998, 1999) propose a consistent generalized moments estimator for λ.

Estimation of the SEM

GLS and FGLS estimation

- Using the feasible GLS estimator requires to have a consistent estimator for λ that could be used to transform the model by the spatial Cochrane-Orcutt procedure and then estimate β by OLS:

$$\hat{\beta}_{FGLS} = [X'\Omega_{\varepsilon}^{-1}(\hat{\lambda})X]^{-1}X'\Omega_{\varepsilon}^{-1}(\hat{\lambda})y \quad (38)$$

$$\hat{\beta}_{FGLS} = (\tilde{X}_L'\tilde{X}_L)^{-1}\tilde{X}_L'\tilde{y}_L \quad (39)$$

  where $\tilde{X}_L = X - \hat{\lambda}WX$ and $\tilde{y}_L = y - \hat{\lambda}Wy$.

- The OLS estimator for β is consistent in the spatially autocorrelated error model, the errors can then be estimated consistently. However the OLS estimator for λ in the SAR model for the errors is not consistent as we previously showed, but Kelejian and Prucha (1998, 1999) propose a consistent generalized moments estimator for λ.

Maximum likelihood estimation of the SEM

Log-likelihood function

- Consider again the spatial error model (SEM) under Assumptions 1-4 and under the additional hypothesis of normality of the error term $u \sim N(0, \sigma_u^2 I_N)$:

$$y = X\beta + \varepsilon \qquad \varepsilon = \lambda W\varepsilon + u \tag{40}$$

- We then have:

$$\varepsilon \sim N(0, \sigma_u^2 \Omega_\varepsilon(\lambda)) \text{ and } y \sim N(X\beta, \sigma_u^2 \Omega_\varepsilon(\lambda)) \tag{41}$$

- Observing that $|\sigma_u^2 \Omega_\varepsilon(\lambda)| = |\sigma_u^2|^N |I - \lambda W|^{-2}$, the log-likelihood function for the SEM model is then obtained as:

$$\begin{aligned} \ln L(\beta', \lambda, \sigma_\epsilon^2) &= -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\sigma_u^2 \Omega_\varepsilon(\lambda)| - \frac{1}{2\sigma_u^2}(y - X\beta)'\Omega_\varepsilon(\lambda)^{-1}(y - X\beta) \\ \ln L(\beta', \lambda, \sigma_\epsilon^2) &= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma_u^2) + \ln|I - \lambda W| \\ &\quad - \frac{1}{2\sigma_u^2}(y - X\beta)'\Omega_\varepsilon(\lambda)^{-1}(y - X\beta) \end{aligned} \tag{42}$$

Maximum likelihood estimation of the SEM

Log-likelihood function

- Consider again the spatial error model (SEM) under Assumptions 1-4 and under the additional hypothesis of normality of the error term $u \sim N(0, \sigma_u^2 I_N)$:

$$y = X\beta + \varepsilon \qquad \varepsilon = \lambda W\varepsilon + u \tag{40}$$

- We then have:

$$\varepsilon \sim N(0, \sigma_u^2 \Omega_\varepsilon(\lambda)) \text{ and } y \sim N(X\beta, \sigma_u^2 \Omega_\varepsilon(\lambda)) \tag{41}$$

- Observing that $|\sigma_u^2 \Omega_\varepsilon(\lambda)| = |\sigma_u^2|^N |I - \lambda W|^{-2}$, the log-likelihood function for the SEM model is then obtained as:

$$\begin{aligned} \ln L(\beta', \lambda, \sigma_\epsilon^2) &= -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\sigma_u^2 \Omega_\varepsilon(\lambda)| - \frac{1}{2\sigma_u^2}(y - X\beta)'\Omega_\varepsilon(\lambda)^{-1}(y - X\beta) \\ \ln L(\beta', \lambda, \sigma_\epsilon^2) &= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma_u^2) + \ln|I - \lambda W| \\ &\quad -\frac{1}{2\sigma_u^2}(y - X\beta)'\Omega_\varepsilon(\lambda)^{-1}(y - X\beta) \end{aligned} \tag{42}$$

Maximum likelihood estimation of the SEM

Log-likelihood function

- Consider again the spatial error model (SEM) under Assumptions 1-4 and under the additional hypothesis of normality of the error term $u \sim N(0, \sigma_u^2 I_N)$:

$$y = X\beta + \varepsilon \qquad \varepsilon = \lambda W\varepsilon + u \tag{40}$$

- We then have:

$$\varepsilon \sim N(0, \sigma_u^2 \Omega_\varepsilon(\lambda)) \text{ and } y \sim N(X\beta, \sigma_u^2 \Omega_\varepsilon(\lambda)) \tag{41}$$

- Observing that $|\sigma_u^2 \Omega_\varepsilon(\lambda)| = |\sigma_u^2|^N |I - \lambda W|^{-2}$, the log-likelihood function for the SEM model is then obtained as:

$$\begin{aligned} \ln L(\beta', \lambda, \sigma_\epsilon^2) &= -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\sigma_u^2 \Omega_\varepsilon(\lambda)| - \frac{1}{2\sigma_u^2}(y - X\beta)'\Omega_\varepsilon(\lambda)^{-1}(y - X\beta) \\ \ln L(\beta', \lambda, \sigma_\epsilon^2) &= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma_u^2) + \ln|I - \lambda W| \\ &\quad -\frac{1}{2\sigma_u^2}(y - X\beta)'\Omega_\varepsilon(\lambda)^{-1}(y - X\beta) \end{aligned} \tag{42}$$

Maximum likelihood estimation of the SEM

Log-likelihood function

- Noting that $\Omega_\varepsilon(\lambda)^{-1} = (I - \lambda W)'(I - \lambda W)$, the log-likelihood may then be written as follows:

$$\begin{aligned}
\ln L(\beta', \lambda, \sigma_\epsilon^2) &= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma_u^2) + \ln|I - \lambda W| \\
&\quad -\frac{1}{2\sigma_u^2}\left[(I - \lambda W)(y - X\beta)\right]'\left[(I - \lambda W)(y - X\beta)\right] \\
\ln L(\beta', \lambda, \sigma_\epsilon^2) &= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma_u^2) + \ln|I - \lambda W| \\
&\quad -\frac{1}{2\sigma_u^2}\left[(I - \lambda W)y - (I - \lambda W)X\beta\right]'\left[(I - \lambda W)y - (I - \lambda W)X\beta\right] \\
\ln L(\beta', \lambda, \sigma_\epsilon^2) &= -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln(\sigma_u^2) + \ln|I - \lambda W| \qquad (43) \\
&\quad -\frac{1}{2\sigma_u^2}\left[y_L - X_L\beta\right]'\left[y_L - X_L\beta\right]
\end{aligned}$$

where $X_L = (I_N - \lambda W)X$ and $y_L = (I_N - \lambda W)y$.

Maximum likelihood estimation of the SEM

ML estimators given λ

- Suppose now that λ is known, the first order conditions for β and σ_u^2 are:

$$\frac{\partial \ln L}{\partial \beta'} = -\frac{1}{2\sigma_u^2}(X_L' y_L + 2\beta X_L' X_L) = 0 \tag{44}$$

$$\frac{\partial \ln L}{\partial \sigma_u^2} = -\frac{N}{\sigma_u^2} + \frac{1}{\sigma_u^4}(y_L - X_L\beta)'(y_L - X_L\beta) = 0 \tag{45}$$

- It follows that the maximum likelihood estimators for β and σ_u^2, given λ, are obtained as:

$$\hat{\beta}_{ML}(\lambda) = (X_L' X_L)^{-1} X_L' y_L \tag{46}$$

$$\hat{\sigma}^2_{ML}(\lambda) = \frac{1}{N}(y_L - X_L\beta)'(y_L - X_L\beta) \tag{47}$$

Maximum likelihood estimation of the SEM

ML estimators given λ

- Suppose now that λ is known, the first order conditions for β and σ_u^2 are:

$$\frac{\partial \ln L}{\partial \beta'} = -\frac{1}{2\sigma_u^2}(X_L' y_L + 2\beta X_L' X_L) = 0 \tag{44}$$

$$\frac{\partial \ln L}{\partial \sigma_u^2} = -\frac{N}{\sigma_u^2} + \frac{1}{\sigma_u^4}(y_L - X_L\beta)'(y_L - X_L\beta) = 0 \tag{45}$$

- It follows that the maximum likelihood estimators for β and σ_u^2, given λ, are obtained as:

$$\hat{\beta}_{ML}(\lambda) = (X_L' X_L)^{-1} X_L' y_L \tag{46}$$

$$\hat{\sigma}^2_{ML}(\lambda) = \frac{1}{N}(y_L - X_L\beta)'(y_L - X_L\beta) \tag{47}$$

Maximum likelihood estimation of the SEM

Concentrated log-likelihood function

- Substitution of (46) and (47) in the log-likelihood function (43) yields a concentrated log-likelihood function as a non-linear function of a single parameter λ:

$$\ln L = \frac{N}{2}[1 + \ln(2\pi)] - \frac{N}{2}\ln\left(\frac{\hat{u}'\hat{u}}{N}\right) + \sum_{i=1}^{N}\ln(1 - \lambda\omega_i) \tag{48}$$

  where $\hat{u} = y_L - X_L\hat{\beta}_{ML}(\lambda)$, then $\hat{u}'\hat{u} = y_L'Y_L - y_L'X_L(X_L'X_L)^{-1}X_L'y_L$, where $y_L = y - \lambda Wy$ and $X_L = X - \lambda WX$ are the spatially filtered variables.

- A maximum likelihood estimate for λ is obtained from a numerical optimization of the concentrated log-likelihood function.

Maximum likelihood estimation of the SEM

Concentrated log-likelihood function

- Substitution of (46) and (47) in the log-likelihood function (43) yields a concentrated log-likelihood function as a non-linear function of a single parameter λ:

$$\ln L = \frac{N}{2}[1 + \ln(2\pi)] - \frac{N}{2}\ln\left(\frac{\hat{u}'\hat{u}}{N}\right) + \sum_{i=1}^{N}\ln(1 - \lambda\omega_i) \tag{48}$$

  where $\hat{u} = y_L - X_L\hat{\beta}_{ML}(\lambda)$, then $\hat{u}'\hat{u} = y_L'Y_L - y_L'X_L(X_L'X_L)^{-1}X_L'y_L$, where $y_L = y - \lambda Wy$ and $X_L = X - \lambda WX$ are the spatially filtered variables.
- A maximum likelihood estimate for λ is obtained from a numerical optimization of the concentrated log-likelihood function.

Estimation of the SEM model

Maximum likelihood estimators

- The estimation procedure is hence more complicated than for the SAR model since $\hat{u}$ in the concentrated log-likelihood function is indirectly also a function of λ as $\hat{\beta}$ is obtained for a value for λ. Therefore a one-time optimization of the concentrated log-likelihood function with respect to λ does not suffice to obtain ML estimates of all the unknown parameters.
- An iterative approach is thus required. This would essentially alternate back and forth between the estimation of λ conditional upon a vector of residuals u generated for a value of β, and an estimation of β and σ^2 conditional upon a value for λ until numerical convergence is obtained.
- This procedure can be described as follows.
  - We first regress y on X by OLS and compute the initial set of residuals $\hat{u} = y - X\beta_{OLS}$.
  - Given $\hat{u}$, we then find λ that maximizes the concentrated log-lokelihood function.
  - Given λ, we carry out FGLS that yields $\hat{\beta}_{FGLS}$.
  - We then compute a new set of residuals $\hat{u} = y - X\hat{\beta}_{FGLS}$.
  - If the numerical convergence criterion is met, that is, if values for both the residuals and $\hat{\beta}_{FGLS}$ fail to change from one iteration to the next, given $\hat{u}$ and $\hat{\lambda}_{ML}$ we compute $\hat{\sigma}^2_{ML}$,
  - else we go back to the maximization of the concentrated log-likelihood function to get a new λ.

Estimation of the SEM model

Maximum likelihood estimators

- The estimation procedure is hence more complicated than for the SAR model since $\hat{u}$ in the concentrated log-likelihood function is indirectly also a function of λ as $\hat{\beta}$ is obtained for a value for λ. Therefore a one-time optimization of the concentrated log-likelihood function with respect to λ does not suffice to obtain ML estimates of all the unknown parameters.
- An iterative approach is thus required. This would essentially alternate back and forth between the estimation of λ conditional upon a vector of residuals u generated for a value of β, and an estimation of β and σ^2 conditional upon a value for λ until numerical convergence is obtained.
- This procedure can be described as follows.
  - We first regress y on X by OLS and compute the initial set of residuals $\hat{u} = y - X\beta_{OLS}$.
  - Given $\hat{u}$, we then find λ that maximizes the concentrated log-lokelihood function.
  - Given λ, we carry out FGLS that yields $\hat{\beta}_{FGLS}$.
  - We then compute a new set of residuals $\hat{u} = y - X\hat{\beta}_{FGLS}$.
  - If the numerical convergence criterion is met, that is, if values for both the residuals and $\hat{\beta}_{FGLS}$ fail to change from one iteration to the next, given $\hat{u}$ and $\hat{\lambda}_{ML}$ we compute $\hat{\sigma}^2_{ML}$,
  - else we go back to the maximization of the concentrated log-likelihood function to get a new λ.

Estimation of the SEM model

Maximum likelihood estimators

- The estimation procedure is hence more complicated than for the SAR model since $\hat{u}$ in the concentrated log-likelihood function is indirectly also a function of λ as $\hat{\beta}$ is obtained for a value for λ. Therefore a one-time optimization of the concentrated log-likelihood function with respect to λ does not suffice to obtain ML estimates of all the unknown parameters.
- An iterative approach is thus required. This would essentially alternate back and forth between the estimation of λ conditional upon a vector of residuals u generated for a value of β, and an estimation of β and σ^2 conditional upon a value for λ until numerical convergence is obtained.
- This procedure can be described as follows.
  - We first regress y on X by OLS and compute the initial set of residuals $\hat{u} = y - X\beta_{OLS}$.
  - Given $\hat{u}$, we then find λ that maximizes the concentrated log-lokelihood function.
  - Given λ, we carry out FGLS that yields $\hat{\beta}_{FGLS}$.
  - We then compute a new set of residuals $\hat{u} = y - X\hat{\beta}_{FGLS}$.
  - If the numerical convergence criterion is met, that is, if values for both the residuals and $\hat{\beta}_{FGLS}$ fail to change from one iteration to the next, given $\hat{u}$ and $\hat{\lambda}_{ML}$ we compute $\hat{\sigma}^2_{ML}$,
  - else we go back to the maximization of the concentrated log-likelihood function to get a new λ.

Estimation of the SEM model

Maximum likelihood estimators

- The estimation procedure is hence more complicated than for the SAR model since $\hat{u}$ in the concentrated log-likelihood function is indirectly also a function of λ as $\hat{\beta}$ is obtained for a value for λ. Therefore a one-time optimization of the concentrated log-likelihood function with respect to λ does not suffice to obtain ML estimates of all the unknown parameters.
- An iterative approach is thus required. This would essentially alternate back and forth between the estimation of λ conditional upon a vector of residuals u generated for a value of β, and an estimation of β and σ^2 conditional upon a value for λ until numerical convergence is obtained.
- This procedure can be described as follows.
  - We first regress y on X by OLS and compute the initial set of residuals $\hat{u} = y - X\beta_{OLS}$.
  - Given $\hat{u}$, we then find λ that maximizes the concentrated log-lokelihood function.
  - Given λ, we carry out FGLS that yields $\hat{\beta}_{FGLS}$.
  - We then compute a new set of residuals $\hat{u} = y - X\hat{\beta}_{FGLS}$.
  - If the numerical convergence criterion is met, that is, if values for both the residuals and $\hat{\beta}_{FGLS}$ fail to change from one iteration to the next, given $\hat{u}$ and $\hat{\lambda}_{ML}$ we compute $\hat{\sigma}^2_{ML}$,
  - else we go back to the maximization of the concentrated log-likelihood function to get a new λ.

Estimation of the SEM model

Maximum likelihood estimators

- The estimation procedure is hence more complicated than for the SAR model since $\hat{u}$ in the concentrated log-likelihood function is indirectly also a function of λ as $\hat{\beta}$ is obtained for a value for λ. Therefore a one-time optimization of the concentrated log-likelihood function with respect to λ does not suffice to obtain ML estimates of all the unknown parameters.
- An iterative approach is thus required. This would essentially alternate back and forth between the estimation of λ conditional upon a vector of residuals u generated for a value of β, and an estimation of β and σ^2 conditional upon a value for λ until numerical convergence is obtained.
- This procedure can be described as follows.
  - We first regress y on X by OLS and compute the initial set of residuals $\hat{u} = y - X\beta_{OLS}$.
  - Given $\hat{u}$, we then find λ that maximizes the concentrated log-lokelihood function.
  - Given λ, we carry out FGLS that yields $\hat{\beta}_{FGLS}$.
  - We then compute a new set of residuals $\hat{u} = y - X\hat{\beta}_{FGLS}$.
  - If the numerical convergence criterion is met, that is, if values for both the residuals and $\hat{\beta}_{FGLS}$ fail to change from one iteration to the next, given $\hat{u}$ and $\hat{\lambda}_{ML}$ we compute $\hat{\sigma}^2_{ML}$,
  - else we go back to the maximization of the concentrated log-likelihood function to get a new λ.

Estimation of the SEM model

Maximum likelihood estimators

- The estimation procedure is hence more complicated than for the SAR model since $\hat{u}$ in the concentrated log-likelihood function is indirectly also a function of λ as $\hat{\beta}$ is obtained for a value for λ. Therefore a one-time optimization of the concentrated log-likelihood function with respect to λ does not suffice to obtain ML estimates of all the unknown parameters.
- An iterative approach is thus required. This would essentially alternate back and forth between the estimation of λ conditional upon a vector of residuals u generated for a value of β, and an estimation of β and σ^2 conditional upon a value for λ until numerical convergence is obtained.
- This procedure can be described as follows.
  - We first regress y on X by OLS and compute the initial set of residuals $\hat{u} = y - X\beta_{OLS}$.
  - Given $\hat{u}$, we then find λ that maximizes the concentrated log-lokelihood function.
  - Given λ, we carry out FGLS that yields $\hat{\beta}_{FGLS}$.
  - We then compute a new set of residuals $\hat{u} = y - X\hat{\beta}_{FGLS}$.
  - If the numerical convergence criterion is met, that is, if values for both the residuals and $\hat{\beta}_{FGLS}$ fail to change from one iteration to the next, given $\hat{u}$ and $\hat{\lambda}_{ML}$ we compute $\hat{\sigma}^2_{ML}$,
  - else we go back to the maximization of the concentrated log-likelihood function to get a new λ.

Estimation of the SEM model

Maximum likelihood estimators

- The estimation procedure is hence more complicated than for the SAR model since $\hat{u}$ in the concentrated log-likelihood function is indirectly also a function of λ as $\hat{\beta}$ is obtained for a value for λ. Therefore a one-time optimization of the concentrated log-likelihood function with respect to λ does not suffice to obtain ML estimates of all the unknown parameters.
- An iterative approach is thus required. This would essentially alternate back and forth between the estimation of λ conditional upon a vector of residuals u generated for a value of β, and an estimation of β and σ^2 conditional upon a value for λ until numerical convergence is obtained.
- This procedure can be described as follows.
  - We first regress y on X by OLS and compute the initial set of residuals $\hat{u} = y - X\beta_{OLS}$.
  - Given $\hat{u}$, we then find λ that maximizes the concentrated log-lokelihood function.
  - Given λ, we carry out FGLS that yields $\hat{\beta}_{FGLS}$.
  - We then compute a new set of residuals $\hat{u} = y - X\hat{\beta}_{FGLS}$.
  - If the numerical convergence criterion is met, that is, if values for both the residuals and $\hat{\beta}_{FGLS}$ fail to change from one iteration to the next, given $\hat{u}$ and $\hat{\lambda}_{ML}$ we compute $\hat{\sigma}^2_{ML}$,
  - else we go back to the maximization of the concentrated log-likelihood function to get a new λ.

Estimation of the SEM model

Maximum likelihood estimators

- The estimation procedure is hence more complicated than for the SAR model since $\hat{u}$ in the concentrated log-likelihood function is indirectly also a function of λ as $\hat{\beta}$ is obtained for a value for λ. Therefore a one-time optimization of the concentrated log-likelihood function with respect to λ does not suffice to obtain ML estimates of all the unknown parameters.
- An iterative approach is thus required. This would essentially alternate back and forth between the estimation of λ conditional upon a vector of residuals u generated for a value of β, and an estimation of β and σ^2 conditional upon a value for λ until numerical convergence is obtained.
- This procedure can be described as follows.
  - We first regress y on X by OLS and compute the initial set of residuals $\hat{u} = y - X\beta_{OLS}$.
  - Given $\hat{u}$, we then find λ that maximizes the concentrated log-lokelihood function.
  - Given λ, we carry out FGLS that yields $\hat{\beta}_{FGLS}$.
  - We then compute a new set of residuals $\hat{u} = y - X\hat{\beta}_{FGLS}$.
  - If the numerical convergence criterion is met, that is, if values for both the residuals and $\hat{\beta}_{FGLS}$ fail to change from one iteration to the next, given $\hat{u}$ and $\hat{\lambda}_{ML}$ we compute $\hat{\sigma}^2_{ML}$,
  - else we go back to the maximization of the concentrated log-likelihood function to get a new λ.

Estimation of the SEM model

Maximum likelihood estimators

- The estimation procedure is hence more complicated than for the SAR model since $\hat{u}$ in the concentrated log-likelihood function is indirectly also a function of λ as $\hat{\beta}$ is obtained for a value for λ. Therefore a one-time optimization of the concentrated log-likelihood function with respect to λ does not suffice to obtain ML estimates of all the unknown parameters.
- An iterative approach is thus required. This would essentially alternate back and forth between the estimation of λ conditional upon a vector of residuals u generated for a value of β, and an estimation of β and σ^2 conditional upon a value for λ until numerical convergence is obtained.
- This procedure can be described as follows.
  - We first regress y on X by OLS and compute the initial set of residuals $\hat{u} = y - X\beta_{OLS}$.
  - Given $\hat{u}$, we then find λ that maximizes the concentrated log-lokelihood function.
  - Given λ, we carry out FGLS that yields $\hat{\beta}_{FGLS}$.
  - We then compute a new set of residuals $\hat{u} = y - X\hat{\beta}_{FGLS}$.
  - If the numerical convergence criterion is met, that is, if values for both the residuals and $\hat{\beta}_{FGLS}$ fail to change from one iteration to the next, given $\hat{u}$ and $\hat{\lambda}_{ML}$ we compute $\hat{\sigma}^2_{ML}$,
  - else we go back to the maximization of the concentrated log-likelihood function to get a new λ.

Estimation of the SEM model

Asymptotic variance-covariance matrix

- As for the SAR model, under the regularity conditions described for instance in Lee (2004, p.1902-1904), it can be shown that the maximum likelihood estimators have the usual asymptotic properties, including consistency, normality, and asymptotic efficiency.
- The asymptotic variance-covariance matrix follows as the inverse of the information matrix. Defining $W_B = W(I - \lambda W)^{-1}$ to simplify notation, we have then:
AsyVar$[\beta', \lambda, \sigma^2] =$

(49)

$$\begin{bmatrix} \frac{1}{\sigma^2} X_L' X_L & 0 & 0 \\ 0 & \text{tr } W_B^2 + \text{tr } W_B' W_B & \frac{1}{\sigma^2} trW_B \\ 0 & \frac{1}{\sigma^2} trW_B & \frac{n}{2\sigma^4} \end{bmatrix}^{-1}$$

Due to the block-diagonal form of the asymptotic variance matrix, knowledge of the precision of λ does not affect the precision of the β estimates. Consequently, if the latter is the primary interest, the complex inverse and trace expressions need not be computed.

Estimation of the SEM model

Asymptotic variance-covariance matrix

- As for the SAR model, under the regularity conditions described for instance in Lee (2004, p.1902-1904), it can be shown that the maximum likelihood estimators have the usual asymptotic properties, including consistency, normality, and asymptotic efficiency.
- The asymptotic variance-covariance matrix follows as the inverse of the information matrix. Defining $W_B = W(I - \lambda W)^{-1}$ to simplify notation, we have then:
AsyVar$[\beta', \lambda, \sigma^2] =$

(49)

$$\begin{bmatrix} \frac{1}{\sigma^2} X_L' X_L & 0 & 0 \\ 0 & \text{tr } W_B^2 + \text{tr } W_B' W_B & \frac{1}{\sigma^2} trW_B \\ 0 & \frac{1}{\sigma^2} trW_B & \frac{n}{2\sigma^4} \end{bmatrix}^{-1}$$

Due to the block-diagonal form of the asymptotic variance matrix, knowledge of the precision of λ does not affect the precision of the β estimates. Consequently, if the latter is the primary interest, the complex inverse and trace expressions need not be computed.

Spatial or interaction multipliers: Motivation 1

Impacts in the classical regression model

- A simple cross-section *no interaction* model

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i \quad \varepsilon_i \sim i.i.d.(0, \sigma^2) \qquad i = 1, \ldots, N$$

- Impact of a variation of x_{ik} on y_i for $k = 1, ..., K$

$$\frac{\partial y_i}{\partial x_{ik}} = \beta_k \quad \text{for all } i$$

- Impact of a variation of x_{jk} on y_i for $k = 1, ..., K$

$$\frac{\partial y_i}{\partial x_j} = 0 \quad \text{for all } j \neq i$$

$\Rightarrow$ *No spatial spillover effect*

Spatial or interaction multipliers: Motivation 1

Impacts in the classical regression model

- A simple cross-section *no interaction* model

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i \quad \varepsilon_i \sim i.i.d.(0, \sigma^2) \qquad i = 1, \ldots, N$$

- Impact of a variation of x_{ik} on y_i for $k = 1, ..., K$

$$\frac{\partial y_i}{\partial x_{ik}} = \beta_k \quad \text{for all } i$$

- Impact of a variation of x_{jk} on y_i for $k = 1, ..., K$

$$\frac{\partial y_i}{\partial x_j} = 0 \quad \text{for all } j \neq i$$

$\Rightarrow$ *No spatial spillover effect*

Spatial or interaction multipliers: Motivation 1

Impacts in the classical regression model

- A simple cross-section *no interaction* model

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i \quad \varepsilon_i \sim i.i.d.(0, \sigma^2) \qquad i = 1, \ldots, N$$

- Impact of a variation of x_{ik} on y_i for $k = 1, ..., K$

$$\frac{\partial y_i}{\partial x_{ik}} = \beta_k \quad \text{for all } i$$

- Impact of a variation of x_{jk} on y_i for $k = 1, ..., K$

$$\frac{\partial y_i}{\partial x_j} = 0 \quad \text{for all } j \neq i$$

$\Rightarrow$ *No spatial spillover effect*

Spatial or interaction multipliers: Motivation 1

Impacts in the classical regression model

- A simple cross-section *no interaction* model

$$y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ik} + \varepsilon_i \quad \varepsilon_i \sim i.i.d.(0, \sigma^2) \qquad i = 1, \ldots, N$$

- Impact of a variation of x_{ik} on y_i for $k = 1, ..., K$

$$\frac{\partial y_i}{\partial x_{ik}} = \beta_k \quad \text{for all } i$$

- Impact of a variation of x_{jk} on y_i for $k = 1, ..., K$

$$\frac{\partial y_i}{\partial x_j} = 0 \quad \text{for all } j \neq i$$

$\Rightarrow$ *No spatial spillover effect*

Spatial or interaction multipliers: Motivation 2

Impacts in the classical regression model

- Stacking over all individuals $i = 1, ..., N$

$$\mathbf{y} = \beta_0 \iota + \sum_{k=1}^{K} \beta_k \mathbf{x}_k + \epsilon$$

- Impact matrices for $k = 1, ...K$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}'_k} = \begin{pmatrix} \frac{\partial y_1}{\partial x_{1k}} & \cdots & \frac{\partial y_1}{\partial x_{Nk}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_N}{\partial x_{1k}} & \cdots & \frac{\partial y_N}{\partial x_{Nk}} \end{pmatrix} = \begin{pmatrix} \beta_k & 0 & \cdots & 0 \\ 0 & \beta_k & 0 \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \beta_k \end{pmatrix} = \mathbf{I}_N \beta_k$$

Spatial or interaction multipliers: Motivation 2

Impacts in the classical regression model

- Stacking over all individuals $i = 1, ..., N$

$$\mathbf{y} = \beta_0 \iota + \sum_{k=1}^{K} \beta_k \mathbf{x}_k + \epsilon$$

- Impact matrices for $k = 1, ...K$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}_k'} = \begin{pmatrix} \frac{\partial y_1}{\partial x_{1k}} & \cdots & \frac{\partial y_1}{\partial x_{Nk}} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_N}{\partial x_{1k}} & \cdots & \frac{\partial y_N}{\partial x_{Nk}} \end{pmatrix} = \begin{pmatrix} \beta_k & 0 & \dots & 0 \\ 0 & \beta_k & 0 \dots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \dots & 0 & \beta_k \end{pmatrix} = \mathbf{I}_N \beta_k$$

Spatial or interaction multipliers

Impacts in the Spatial Durbin Model

- Let us consider the general case of the Spatial Durbin Model:

$$y_i = \beta_0 + \rho \sum_{j=1}^{N} w_{ij} y_j + \sum_{k=1}^{K} x_{ik} \beta_k + \sum_{k=1}^{K} w_{ij} z_{ik} \gamma_k + \varepsilon_i \text{ for } i = 1, ..., N \tag{50}$$

- or in matrix form:

$$y = \beta_0 \iota_N + \rho W y + X\beta + WZ\gamma + \varepsilon \tag{51}$$

- where y_N is the $N \times 1$ vector of the dependent variable, $W_N y_N$ is the spatially lagged dependent variable, ι_N is the $N \times 1$ unit vector and X is the $N \times K$ matrix of the explanatory variables; β is the $K \times 1$ vector of the associated coefficients.
- Note that Z may include explanatory variables not included in X, supposed affecting y only through their spatial lag: Z may then be a $N \times (K + M)$ matrix $Z = [X \ \widetilde{Z}]$ where X is the previous $N \times K$ matrix of the explanatory variables and $\tilde{Z}$ is the $N \times M$ matrix of extra explanatory variables not included in X which are supposed to affect y only through their spatial lag. WZ is the $(N \times (K + M))$ matrix of the spatially lagged explanatory variables and γ is the $(K + M) \times 1$ vector of the associated coefficients.

Spatial or interaction multipliers

Impacts in the Spatial Durbin Model

- Let us consider the general case of the Spatial Durbin Model:

$$y_i = \beta_0 + \rho \sum_{j=1}^{N} w_{ij} y_j + \sum_{k=1}^{K} x_{ik} \beta_k + \sum_{k=1}^{K} w_{ij} z_{ik} \gamma_k + \varepsilon_i \text{ for } i = 1, ..., N \tag{50}$$

- or in matrix form:

$$y = \beta_0 \iota_N + \rho W y + X\beta + WZ\gamma + \varepsilon \tag{51}$$

- where y_N is the $N \times 1$ vector of the dependent variable, $W_N y_N$ is the spatially lagged dependent variable, ι_N is the $N \times 1$ unit vector and X is the $N \times K$ matrix of the explanatory variables; β is the $K \times 1$ vector of the associated coefficients.
- Note that Z may include explanatory variables not included in X, supposed affecting y only through their spatial lag: Z may then be a $N \times (K + M)$ matrix $Z = [X \ \tilde{Z}]$ where X is the previous $N \times K$ matrix of the explanatory variables and $\tilde{Z}$ is the $N \times M$ matrix of extra explanatory variables not included in X which are supposed to affect y only through their spatial lag. WZ is the $(N \times (K + M))$ matrix of the spatially lagged explanatory variables and γ is the $(K + M) \times 1$ vector of the associated coefficients.

Spatial or interaction multipliers

Impacts in the Spatial Durbin Model

- Let us consider the general case of the Spatial Durbin Model:

$$y_i = \beta_0 + \rho \sum_{j=1}^{N} w_{ij} y_j + \sum_{k=1}^{K} x_{ik} \beta_k + \sum_{k=1}^{K} w_{ij} z_{ik} \gamma_k + \varepsilon_i \text{ for } i = 1, ..., N \tag{50}$$

- or in matrix form:

$$y = \beta_0 \iota_N + \rho W y + X\beta + WZ\gamma + \varepsilon \tag{51}$$

- where y_N is the $N \times 1$ vector of the dependent variable, $W_N y_N$ is the spatially lagged dependent variable, ι_N is the $N \times 1$ unit vector and X is the $N \times K$ matrix of the explanatory variables; β is the $K \times 1$ vector of the associated coefficients.
- Note that Z may include explanatory variables not included in X, supposed affecting y only through their spatial lag: Z may then be a $N \times (K + M)$ matrix $Z = [X \ \tilde{Z}]$ where X is the previous $N \times K$ matrix of the explanatory variables and $\tilde{Z}$ is the $N \times M$ matrix of extra explanatory variables not included in X which are supposed to affect y only through their spatial lag. WZ is the $(N \times (K + M))$ matrix of the spatially lagged explanatory variables and γ is the $(K + M) \times 1$ vector of the associated coefficients.

Spatial or interaction multipliers

Impacts in the Spatial Durbin Model

- Let us consider the general case of the Spatial Durbin Model:

$$y_i = \beta_0 + \rho \sum_{j=1}^{N} w_{ij} y_j + \sum_{k=1}^{K} x_{ik} \beta_k + \sum_{k=1}^{K} w_{ij} z_{ik} \gamma_k + \varepsilon_i \text{ for } i = 1, ..., N \tag{50}$$

- or in matrix form:

$$y = \beta_0 \iota_N + \rho W y + X\beta + WZ\gamma + \varepsilon \tag{51}$$

- where y_N is the $N \times 1$ vector of the dependent variable, $W_N y_N$ is the spatially lagged dependent variable, ι_N is the $N \times 1$ unit vector and X is the $N \times K$ matrix of the explanatory variables; β is the $K \times 1$ vector of the associated coefficients.
- Note that Z may include explanatory variables not included in X, supposed affecting y only through their spatial lag: Z may then be a $N \times (K + M)$ matrix $Z = [X\ \widetilde{Z}]$ where X is the previous $N \times K$ matrix of the explanatory variables and $\tilde{Z}$ is the $N \times M$ matrix of extra explanatory variables not included in X which are supposed to affect y only through their spatial lag. WZ is the $(N \times (K + M))$ matrix of the spatially lagged explanatory variables and γ is the $(K + M) \times 1$ vector of the associated coefficients.

Spatial or interaction multipliers

Impacts in the Spatial Durbin Model

- Suppose for simplicity that Z only includes X, the model may then be written as follows:

$$y = \beta_0 \iota_N + \rho W y + \sum_{k=1}^{K} (I_N \beta_k + W \gamma_k) X_k + \varepsilon \tag{52}$$

and the reduced form is then:

$$y = (I - \rho W)^{-1} \iota_N \beta_0 + \sum_{k=1}^{K} (I - \rho W)^{-1} (I_N \beta_k + W \gamma_k) X_k + (I - \rho W)^{-1} \varepsilon \tag{53}$$

- Note that if W is row normalized then $W^q \iota_N = \iota_N$ for $q \geq 0$ and we have:

$$\begin{aligned} (I_N - \rho W)^{-1} \iota_N &= (I_N + \rho W + \rho^2 W^2 + \ldots) \iota_N \\ &= (1 + \rho + \rho^2 + \ldots) \iota_N = \frac{1}{1 - \rho} \iota_N \end{aligned} \tag{54}$$

$$y = \frac{\beta_0}{1 - \rho} \iota_N + \sum_{k=1}^{K} (I_N - \rho W)^{-1} (I_N \beta_k + W \gamma_k) X_k + (I_N - \rho W)^{-1} \epsilon \tag{55}$$

Spatial or interaction multipliers

Impacts in the Spatial Durbin Model

- Suppose for simplicity that Z only includes X, the model may then be written as follows:

$$y = \beta_0 \iota_N + \rho W y + \sum_{k=1}^{K} (I_N \beta_k + W \gamma_k) X_k + \varepsilon \tag{52}$$

and the reduced form is then:

$$y = (I - \rho W)^{-1} \iota_N \beta_0 + \sum_{k=1}^{K} (I - \rho W)^{-1} (I_N \beta_k + W \gamma_k) X_k + (I - \rho W)^{-1} \varepsilon \tag{53}$$

- Note that if W is row normalized then $W^q \iota_N = \iota_N$ for $q \geq 0$ and we have:

$$\begin{aligned} (I_N - \rho W)^{-1} \iota_N &= (I_N + \rho W + \rho^2 W^2 + \ldots) \iota_N \\ &= (1 + \rho + \rho^2 + \ldots) \iota_N = \frac{1}{1 - \rho} \iota_N \end{aligned} \tag{54}$$

$$y = \frac{\beta_0}{1 - \rho} \iota_N + \sum_{k=1}^{K} (I_N - \rho W)^{-1} (I_N \beta_k + W \gamma_k) X_k + (I_N - \rho W)^{-1} \epsilon \tag{55}$$

Spatial or interaction multipliers

Partial derivatives: Impact matrix

- Let us take the partial derivatives of y relative to X_k for $k = 1, \ldots, K$:

$$\frac{\partial y}{\partial X_k'} = S_k(W) = (I_N - \rho W)^{-1}(I_N \beta_k + W\gamma_k)$$
$$= (I_N + \rho W + \rho^2 W^2 + \ldots)(I_N \beta_k + W\gamma_k)$$

  where $(I_N - \rho W)^{-1}$ is the so-called *global spatial multiplier* or *global interaction multiplier*.

- Let us define, as the *impact matrix* associated to the k^{th} explanatory variable, the $N \times N$ matrix $S_k(W) = (I_N - \rho W)^{-1}(I_N \beta_k + W\gamma_k)$. $S_k(W)$ is a $N \times N$ full matrix whose elements are:

$$S_k(W) = \begin{pmatrix} S_k(W)_{11} & S_k(W)_{12} & \cdots & S_k(W)_{1N} \\ S_k(W)_{21} & S_k(W)_{22} & & S_k(W)_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ S_k(W)_{N1} & S_k(W)_{N2} & \cdots & S_k(W)_{NN} \end{pmatrix} \quad (56)$$

Spatial or interaction multipliers

Partial derivatives: Impact matrix

- Let us take the partial derivatives of y relative to X_k for $k = 1, \ldots, K$:

$$\frac{\partial y}{\partial X_k'} = S_k(W) = (I_N - \rho W)^{-1}(I_N \beta_k + W\gamma_k)$$
$$= (I_N + \rho W + \rho^2 W^2 + \ldots)(I_N \beta_k + W\gamma_k)$$

  where $(I_N - \rho W)^{-1}$ is the so-called *global spatial multiplier* or *global interaction multiplier*.
- Let us define, as the *impact matrix* associated to the k^{th} explanatory variable, the $N \times N$ matrix $S_k(W) = (I_N - \rho W)^{-1}(I_N \beta_k + W\gamma_k)$. $S_k(W)$ is a $N \times N$ full matrix whose elements are:

$$S_k(W) = \begin{pmatrix} S_k(W)_{11} & S_k(W)_{12} & \ldots & S_k(W)_{1N} \\ S_k(W)_{21} & S_k(W)_{22} & & S_k(W)_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ S_k(W)_{N1} & S_k(W)_{N2} & \ldots & S_k(W)_{NN} \end{pmatrix} \tag{56}$$

Spatial or interaction multipliers

Impact matrix: Interpretations

- The partial derivatives of y_i relative to x_{ik} or x_{jk} for $i, j = 1, \ldots, N$, $j \neq i$ and for $k = 1, \ldots, K$ are then:

$$\frac{\partial y_i}{\partial x_{ik}} = S_k(W)_{ii}, \qquad \frac{\partial y_i}{\partial x_{jk}} = S_k(W)_{ij} \tag{57}$$

- In general $S_k(W)_{ii} \neq 0$ and $S_k(W)_{ij} \neq 0$ for $i, j = 1, \ldots, N$, $j \neq i$ and for $k = 1, \ldots, K$.
- The diagonal elements of this matrix, $\text{diag}(S_k(W))$, represent the *direct impacts* including "own spillover" effects, which are inherently heterogenous in presence of spatial autocorrelation due to differentiated interaction terms in the W matrix.
- This type of heterogeneity is called *interactive heterogeneity*, in opposition to standard individual heterogeneity in panel data models (Debarsy and Ertur, 2010).
- The off-diagonal elements of the impact matrix represent *indirect impacts*: they are collected in the matrix $Q_k(W) = S_k(W) - \text{diag}(S_k(W))$.

Spatial or interaction multipliers

Impact matrix: Interpretations

- The partial derivatives of y_i relative to x_{ik} or x_{jk} for $i, j = 1, \ldots, N$, $j \neq i$ and for $k = 1, \ldots, K$ are then:

$$\frac{\partial y_i}{\partial x_{ik}} = S_k(W)_{ii}, \qquad \frac{\partial y_i}{\partial x_{jk}} = S_k(W)_{ij} \tag{57}$$

- In general $S_k(W)_{ii} \neq 0$ and $S_k(W)_{ij} \neq 0$ for $i, j = 1, \ldots, N$, $j \neq i$ and for $k = 1, \ldots, K$.
- The diagonal elements of this matrix, $\text{diag}(S_k(W))$, represent the *direct impacts* including "own spillover" effects, which are inherently heterogenous in presence of spatial autocorrelation due to differentiated interaction terms in the W matrix.
- This type of heterogeneity is called *interactive heterogeneity*, in opposition to standard individual heterogeneity in panel data models (Debarsy and Ertur, 2010).
- The off-diagonal elements of the impact matrix represent *indirect impacts*: they are collected in the matrix $Q_k(W) = S_k(W) - \text{diag}(S_k(W))$.

Spatial or interaction multipliers

Impact matrix: Interpretations

- The partial derivatives of y_i relative to x_{ik} or x_{jk} for $i, j = 1, \ldots, N$, $j \neq i$ and for $k = 1, \ldots, K$ are then:

$$\frac{\partial y_i}{\partial x_{ik}} = S_k(W)_{ii}, \qquad \frac{\partial y_i}{\partial x_{jk}} = S_k(W)_{ij} \tag{57}$$

- In general $S_k(W)_{ii} \neq 0$ and $S_k(W)_{ij} \neq 0$ for $i, j = 1, \ldots, N$, $j \neq i$ and for $k = 1, \ldots, K$.
- The diagonal elements of this matrix, $\mathrm{diag}(S_k(W))$, represent the *direct impacts* including "own spillover" effects, which are inherently heterogenous in presence of spatial autocorrelation due to differentiated interaction terms in the W matrix.
- This type of heterogeneity is called *interactive heterogeneity*, in opposition to standard individual heterogeneity in panel data models (Debarsy and Ertur, 2010).
- The off-diagonal elements of the impact matrix represent *indirect impacts*: they are collected in the matrix $Q_k(W) = S_k(W) - \mathrm{diag}(S_k(W))$.

Spatial or interaction multipliers

Impact matrix: Interpretations

- The partial derivatives of y_i relative to x_{ik} or x_{jk} for $i, j = 1, \ldots, N$, $j \neq i$ and for $k = 1, \ldots, K$ are then:

$$\frac{\partial y_i}{\partial x_{ik}} = S_k(W)_{ii}, \qquad \frac{\partial y_i}{\partial x_{jk}} = S_k(W)_{ij} \tag{57}$$

- In general $S_k(W)_{ii} \neq 0$ and $S_k(W)_{ij} \neq 0$ for $i, j = 1, \ldots, N$, $j \neq i$ and for $k = 1, \ldots, K$.
- The diagonal elements of this matrix, $\mathrm{diag}(S_k(W))$, represent the *direct impacts* including "own spillover" effects, which are inherently heterogenous in presence of spatial autocorrelation due to differentiated interaction terms in the W matrix.
- This type of heterogeneity is called *interactive heterogeneity*, in opposition to standard individual heterogeneity in panel data models (Debarsy and Ertur, 2010).
- The off-diagonal elements of the impact matrix represent *indirect impacts*: they are collected in the matrix $Q_k(W) = S_k(W) - \mathrm{diag}(S_k(W))$.

Spatial or interaction multipliers

Impact matrix: Interpretations

- The partial derivatives of y_i relative to x_{ik} or x_{jk} for $i, j = 1, \ldots, N$, $j \neq i$ and for $k = 1, \ldots, K$ are then:

$$\frac{\partial y_i}{\partial x_{ik}} = S_k(W)_{ii}, \qquad \frac{\partial y_i}{\partial x_{jk}} = S_k(W)_{ij} \tag{57}$$

- In general $S_k(W)_{ii} \neq 0$ and $S_k(W)_{ij} \neq 0$ for $i, j = 1, \ldots, N$, $j \neq i$ and for $k = 1, \ldots, K$.
- The diagonal elements of this matrix, $\mathrm{diag}(S_k(W))$, represent the *direct impacts* including "own spillover" effects, which are inherently heterogenous in presence of spatial autocorrelation due to differentiated interaction terms in the W matrix.
- This type of heterogeneity is called *interactive heterogeneity*, in opposition to standard individual heterogeneity in panel data models (Debarsy and Ertur, 2010).
- The off-diagonal elements of the impact matrix represent *indirect impacts*: they are collected in the matrix $Q_k(W) = S_k(W) - \mathrm{diag}(S_k(W))$.

Spatial or interaction multipliers

Impact matrix: Interpretations

- Note that the own derivative for country *i* includes the feedback effects where country *i* affects country *j* and country *j* also affects country *i* as well as longer paths which might go from country *i* to *j* to *k* and back to *i*.
- The magnitude of those direct effects will depend on: (1) the degree of interaction between countries, which is governed by the *W* matrix, (2) the parameter ρ, measuring the strength of spatial correlation between countries and (3) the parameter β_k.
- Note also that the magnitude of pure feedback effects are then given by $S_k(W)_{ii} - \beta_k$, where β_k could be interpreted as representing the direct impact of the explanatory variable if there was no spatial autocorrelation, i.e. if ρ was equal to zero.

Spatial or interaction multipliers

Impact matrix: Interpretations

- Note that the own derivative for country *i* includes the feedback effects where country *i* affects country *j* and country *j* also affects country *i* as well as longer paths which might go from country *i* to *j* to *k* and back to *i*.
- The magnitude of those direct effects will depend on: (1) the degree of interaction between countries, which is governed by the *W* matrix, (2) the parameter ρ, measuring the strength of spatial correlation between countries and (3) the parameter β_k.
- Note also that the magnitude of pure feedback effects are then given by $S_k(W)_{ii} - \beta_k$, where β_k could be interpreted as representing the direct impact of the explanatory variable if there was no spatial autocorrelation, i.e. if ρ was equal to zero.

Spatial or interaction multipliers

Impact matrix: Interpretations

- Note that the own derivative for country *i* includes the feedback effects where country *i* affects country *j* and country *j* also affects country *i* as well as longer paths which might go from country *i* to *j* to *k* and back to *i*.
- The magnitude of those direct effects will depend on: (1) the degree of interaction between countries, which is governed by the *W* matrix, (2) the parameter ρ, measuring the strength of spatial correlation between countries and (3) the parameter β_k.
- Note also that the magnitude of pure feedback effects are then given by $S_k(W)_{ii} - \beta_k$, where β_k could be interpreted as representing the direct impact of the explanatory variable if there was no spatial autocorrelation, i.e. if ρ was equal to zero.

Spatial or interaction multipliers

Column interpretation

- Moreover, considering column j, we note that a variation Δx_{jk} of the k^{th} explanatory variable in spatial unit j differently affects each of the spatial units of the sample:

$$\begin{pmatrix} S_k(W)_{1j} \\ S_k(W)_{2j} \\ \vdots \\ S_k(W)_{jj} \\ \vdots \\ S_k(W)_{Nj} \end{pmatrix}$$

- The sum down the j^{th} column yields the total impact on y_i $(i = 1, ..., N)$ for all the N spatial units of the sample of a change of x_{jk} in spatial unit j.
- The total impacts, direct and indirect, from each of the units $j = 1, ..., N$ are then collected in the row vector $\iota_N' S_k(W)$.
- However it may be of interest to distinguish direct and indirect effects in applied papers where the direct impacts are actually higher than each of the indirect impacts and potentially higher than their sum (if the impact matrix is strictly diagonally dominant).
- The total indirect impacts from each of the units $j = 1, ..., N$ may then be usefully collected in the row vector $\iota_N' Q_k(W)$.

Spatial or interaction multipliers

Column interpretation

- Moreover, considering column j, we note that a variation Δx_{jk} of the k^{th} explanatory variable in spatial unit j differently affects each of the spatial units of the sample:

$$\begin{pmatrix} S_k(W)_{1j} \\ S_k(W)_{2j} \\ \vdots \\ S_k(W)_{jj} \\ \vdots \\ S_k(W)_{Nj} \end{pmatrix}$$

- The sum down the j^{th} column yields the total impact on y_i ($i = 1, ..., N$) for all the N spatial units of the sample of a change of x_{jk} in spatial unit j.
- The total impacts, direct and indirect, from each of the units $j = 1, ..., N$ are then collected in the row vector $\iota_N' S_k(W)$.
- However it may be of interest to distinguish direct and indirect effects in applied papers where the direct impacts are actually higher than each of the indirect impacts and potentially higher than their sum (if the impact matrix is strictly diagonally dominant).
- The total indirect impacts from each of the units $j = 1, ..., N$ may then be usefully collected in the row vector $\iota_N' Q_k(W)$.

Spatial or interaction multipliers

Column interpretation

- Moreover, considering column j, we note that a variation Δx_{jk} of the k^{th} explanatory variable in spatial unit j differently affects each of the spatial units of the sample:

$$\begin{pmatrix} S_k(W)_{1j} \\ S_k(W)_{2j} \\ \vdots \\ S_k(W)_{jj} \\ \vdots \\ S_k(W)_{Nj} \end{pmatrix}$$

- The sum down the j^{th} column yields the total impact on y_i ($i = 1, ..., N$) for all the N spatial units of the sample of a change of x_{jk} in spatial unit j.
- The total impacts, direct and indirect, from each of the units $j = 1, ..., N$ are then collected in the row vector $\iota'_N S_k(W)$.
- However it may be of interest to distinguish direct and indirect effects in applied papers where the direct impacts are actually higher than each of the indirect impacts and potentially higher than their sum (if the impact matrix is strictly diagonally dominant).
- The total indirect impacts from each of the units $j = 1, ..., N$ may then be usefully collected in the row vector $\iota'_N Q_k(W)$.

Spatial or interaction multipliers

Column interpretation

- Moreover, considering column j, we note that a variation Δx_{jk} of the k^{th} explanatory variable in spatial unit j differently affects each of the spatial units of the sample:

$$\begin{pmatrix} S_k(W)_{1j} \\ S_k(W)_{2j} \\ \vdots \\ S_k(W)_{jj} \\ \vdots \\ S_k(W)_{Nj} \end{pmatrix}$$

- The sum down the j^{th} column yields the total impact on y_i ($i = 1, ..., N$) for all the N spatial units of the sample of a change of x_{jk} in spatial unit j.
- The total impacts, direct and indirect, from each of the units $j = 1, ..., N$ are then collected in the row vector $\iota'_N S_k(W)$.
- However it may be of interest to distinguish direct and indirect effects in applied papers where the direct impacts are actually higher than each of the indirect impacts and potentially higher than their sum (if the impact matrix is strictly diagonally dominant).
- The total indirect impacts from each of the units $j = 1, ..., N$ may then be usefully collected in the row vector $\iota'_N Q_k(W)$.

Spatial or interaction multipliers

Column interpretation

- Moreover, considering column j, we note that a variation Δx_{jk} of the k^{th} explanatory variable in spatial unit j differently affects each of the spatial units of the sample:

$$\begin{pmatrix} S_k(W)_{1j} \\ S_k(W)_{2j} \\ \vdots \\ S_k(W)_{jj} \\ \vdots \\ S_k(W)_{Nj} \end{pmatrix}$$

- The sum down the j^{th} column yields the total impact on y_i ($i = 1, ..., N$) for all the N spatial units of the sample of a change of x_{jk} in spatial unit j.
- The total impacts, direct and indirect, from each of the units $j = 1, ..., N$ are then collected in the row vector $\iota'_N S_k(W)$.
- However it may be of interest to distinguish direct and indirect effects in applied papers where the direct impacts are actually higher than each of the indirect impacts and potentially higher than their sum (if the impact matrix is strictly diagonally dominant).
- The total indirect impacts from each of the units $j = 1, ..., N$ may then be usefully collected in the row vector $\iota'_N Q_k(W)$.

Spatial or interaction multipliers

Row interpretation

- Considering row i, we note that an identical variation ΔX_k of the k^{th} explanatory variable across all the units of the sample differently affects spatial unit i:

$$\begin{pmatrix} S_k(W)_{i1} & S_k(W)_{i2} & \dots & S_k(W)_{ii} & \dots & S_k(W)_{iN} \end{pmatrix}$$

- The sum across the i^{th} row represents the total impact on y_i of an identical change of x_{jk} $(j = 1, ..., N)$ across all the N spatial units in the sample.
- The total impacts, direct and indirect, on each of the units $i = 1, ..., N$ are then collected in the column vector $S_k(W)\iota_N$.
- Again the total indirect impacts on each of the units $i = 1, ..., N$ may be collected in the row vector $Q_k(W)\iota_N$.
- Of course, in both cases the corresponding elements of main diagonal of $S_k(W)$ may be taken into account or not in those sums, depending on the inclusion or the exclusion of the direct effects in those computations.

Spatial or interaction multipliers

Row interpretation

- Considering row i, we note that an identical variation ΔX_k of the k^{th} explanatory variable across all the units of the sample differently affects spatial unit i:

$$\begin{pmatrix} S_k(W)_{i1} & S_k(W)_{i2} & \dots & S_k(W)_{ii} & \dots & S_k(W)_{iN} \end{pmatrix}$$

- The sum across the i^{th} row represents the total impact on y_i of an identical change of x_{jk} ($j = 1, ..., N$) across all the N spatial units in the sample.
- The total impacts, direct and indirect, on each of the units $i = 1, ..., N$ are then collected in the column vector $S_k(W)\iota_N$.
- Again the total indirect impacts on each of the units $i = 1, ..., N$ may be collected in the row vector $Q_k(W)\iota_N$.
- Of course, in both cases the corresponding elements of main diagonal of $S_k(W)$ may be taken into account or not in those sums, depending on the inclusion or the exclusion of the direct effects in those computations.

Spatial or interaction multipliers

Row interpretation

- Considering row i, we note that an identical variation ΔX_k of the k^{th} explanatory variable across all the units of the sample differently affects spatial unit i:

$$\begin{pmatrix} S_k(W)_{i1} & S_k(W)_{i2} & \dots & S_k(W)_{ii} & \dots & S_k(W)_{iN} \end{pmatrix}$$

- The sum across the i^{th} row represents the total impact on y_i of an identical change of x_{jk} ($j = 1, ..., N$) across all the N spatial units in the sample.
- The total impacts, direct and indirect, on each of the units $i = 1, ..., N$ are then collected in the column vector $S_k(W)\iota_N$.
- Again the total indirect impacts on each of the units $i = 1, ..., N$ may be collected in the row vector $Q_k(W)\iota_N$.
- Of course, in both cases the corresponding elements of main diagonal of $S_k(W)$ may be taken into account or not in those sums, depending on the inclusion or the exclusion of the direct effects in those computations.

Spatial or interaction multipliers

Row interpretation

- Considering row i, we note that an identical variation ΔX_k of the k^{th} explanatory variable across all the units of the sample differently affects spatial unit i:

$$\begin{pmatrix} S_k(W)_{i1} & S_k(W)_{i2} & \dots & S_k(W)_{ii} & \dots & S_k(W)_{iN} \end{pmatrix}$$

- The sum across the i^{th} row represents the total impact on y_i of an identical change of x_{jk} ($j = 1, ..., N$) across all the N spatial units in the sample.
- The total impacts, direct and indirect, on each of the units $i = 1, ..., N$ are then collected in the column vector $S_k(W)\iota_N$.
- Again the total indirect impacts on each of the units $i = 1, ..., N$ may be collected in the row vector $Q_k(W)\iota_N$.
- Of course, in both cases the corresponding elements of main diagonal of $S_k(W)$ may be taken into account or not in those sums, depending on the inclusion or the exclusion of the direct effects in those computations.

Spatial or interaction multipliers

Row interpretation

- Considering row i, we note that an identical variation ΔX_k of the k^{th} explanatory variable across all the units of the sample differently affects spatial unit i:

$$\begin{pmatrix} S_k(W)_{i1} & S_k(W)_{i2} & \dots & S_k(W)_{ii} & \dots & S_k(W)_{iN} \end{pmatrix}$$

- The sum across the i^{th} row represents the total impact on y_i of an identical change of x_{jk} ($j = 1, ..., N$) across all the N spatial units in the sample.
- The total impacts, direct and indirect, on each of the units $i = 1, ..., N$ are then collected in the column vector $S_k(W)\iota_N$.
- Again the total indirect impacts on each of the units $i = 1, ..., N$ may be collected in the row vector $Q_k(W)\iota_N$.
- Of course, in both cases the corresponding elements of main diagonal of $S_k(W)$ may be taken into account or not in those sums, depending on the inclusion or the exclusion of the direct effects in those computations.

Spatial or interaction multipliers

Scalar summaries

- Given the complexity and the amount of the information available in such impact matrices, LeSage and Pace (2009) suggest some useful summary scalar measures.
- The average direct impact, including feedback effects, is defined as:

$$N^{-1}\,\mathrm{tr}(S_k(W))$$

- whereas the average global impact is defined as:

$$N^{-1}\iota_N' S_k(W)\iota_N$$

  where ι_N is the $N \times 1$ sum vector.
- Finally the average indirect impact is, by definition, the difference between the average global impact and the average direct impact:

$$N^{-1}\iota_N' S_k(W)\iota_N - N^{-1}\,\mathrm{tr}(S_k(W)) = N^{-1}\iota_N' Q_k(W)\iota_N$$

Spatial or interaction multipliers

Scalar summaries

- Given the complexity and the amount of the information available in such impact matrices, LeSage and Pace (2009) suggest some useful summary scalar measures.
- The average direct impact, including feedback effects, is defined as:

$$N^{-1}\,\mathrm{tr}(S_k(W))$$

- whereas the average global impact is defined as:

$$N^{-1}\iota_N' S_k(W)\iota_N$$

where ι_N is the $N \times 1$ sum vector.

- Finally the average indirect impact is, by definition, the difference between the average global impact and the average direct impact:

$$N^{-1}\iota_N' S_k(W)\iota_N - N^{-1}\,\mathrm{tr}(S_k(W)) = N^{-1}\iota_N' Q_k(W)\iota_N$$

Spatial or interaction multipliers

Scalar summaries

- Given the complexity and the amount of the information available in such impact matrices, LeSage and Pace (2009) suggest some useful summary scalar measures.
- The average direct impact, including feedback effects, is defined as:

$$N^{-1}\,\mathrm{tr}(S_k(W))$$

- whereas the average global impact is defined as:

$$N^{-1}\iota_N' S_k(W)\iota_N$$

where ι_N is the $N \times 1$ sum vector.

- Finally the average indirect impact is, by definition, the difference between the average global impact and the average direct impact:

$$N^{-1}\iota_N' S_k(W)\iota_N - N^{-1}\,\mathrm{tr}(S_k(W)) = N^{-1}\iota_N' Q_k(W)\iota_N$$

Spatial or interaction multipliers

Scalar summaries

- Given the complexity and the amount of the information available in such impact matrices, LeSage and Pace (2009) suggest some useful summary scalar measures.
- The average direct impact, including feedback effects, is defined as:

$$N^{-1}\,\mathrm{tr}(S_k(W))$$

- whereas the average global impact is defined as:

$$N^{-1}\iota_N' S_k(W)\iota_N$$

  where ι_N is the $N \times 1$ sum vector.
- Finally the average indirect impact is, by definition, the difference between the average global impact and the average direct impact:

$$N^{-1}\iota_N' S_k(W)\iota_N - N^{-1}\,\mathrm{tr}(S_k(W)) = N^{-1}\iota_N' Q_k(W)\iota_N$$

Spatial or interaction multipliers: Special cases 1

The spatial autoregressive model (SAR)

- Let us now consider the SAR specification, excluding the spatial lags of the explanatory variables from the specification ($\rho \neq 0$, $\beta \neq 0$ and $\gamma = 0$), the partial derivatives of y relative to X_k for $k = 1, \ldots, K$ are then simply:

$$\frac{\partial y}{\partial X_k'} = S_k(W) = (I_N - \rho W)^{-1} I_N \beta_k = (I_N + \rho W + \rho^2 W^2 + \ldots) I_N \beta_k \tag{58}$$

  The diagonal elements of this *impact matrix* represent the direct effects including "own spillover" effects whereas the off-diagonal terms represent indirect effects.

- Note that in this special case, using a row-normalized interaction matrix W, since then $(I_N - \rho W)^{-1} \iota_N = \frac{1}{1-\rho} \iota_N$, the total impacts on each of the units $i = 1, ..., N$ collected in the column vector $S_k(W)\iota_N$ may be written as follows:

$$S_k(W)\iota_N = (I_N - \rho W)^{-1} \beta_k \iota_N = \frac{\beta_k}{1-\rho} \iota_N$$

  The average global impact of a variation of the k^{th} explanatory variable simplifies then to:

$$N^{-1} \iota_N' S_k(W) \iota_N = N^{-1} \frac{\beta_k}{1-\rho} \iota_N' \iota_N = \frac{\beta_k}{1-\rho} \tag{59}$$

Spatial or interaction multipliers: Special cases 1

The spatial autoregressive model (SAR)

- Let us now consider the SAR specification, excluding the spatial lags of the explanatory variables from the specification ($\rho \neq 0$, $\beta \neq 0$ and $\gamma = 0$), the partial derivatives of y relative to X_k for $k = 1, \ldots, K$ are then simply:

$$\frac{\partial y}{\partial X_k'} = S_k(W) = (I_N - \rho W)^{-1} I_N \beta_k = (I_N + \rho W + \rho^2 W^2 + \ldots) I_N \beta_k \tag{58}$$

  The diagonal elements of this *impact matrix* represent the direct effects including "own spillover" effects whereas the off-diagonal terms represent indirect effects.
- Note that in this special case, using a row-normalized interaction matrix W, since then $(I_N - \rho W)^{-1} \iota_N = \frac{1}{1-\rho} \iota_N$, the total impacts on each of the units $i = 1, ..., N$ collected in the column vector $S_k(W)\iota_N$ may be written as follows:

$$S_k(W)\iota_N = (I_N - \rho W)^{-1} \beta_k \iota_N = \frac{\beta_k}{1-\rho} \iota_N$$

  The average global impact of a variation of the k^{th} explanatory variable simplifies then to:

$$N^{-1} \iota_N' S_k(W) \iota_N = N^{-1} \frac{\beta_k}{1-\rho} \iota_N' \iota_N = \frac{\beta_k}{1-\rho} \tag{59}$$

Spatial or interaction multipliers: Special cases 2

The cross regressive model

- Consider now the following simpler model, which does not include the spatially lagged endogenous variable ($\rho = 0$) but includes exogenous variables together with spatially lagged exogenous variables ($\beta \neq 0$ and $\gamma \neq 0$) which may be estimated by OLS under the usual set of assumptions:

$$y = \beta_0 \iota_N + \sum_{k=1}^{K} (I_N \beta_k + W\gamma_k) X_k + \varepsilon \tag{60}$$

Let us take the partial derivatives of y relative to X_k for $k = 1, \ldots, K$:

$$\frac{\partial y}{\partial X_k'} = S_k(W) = I_N \beta_k + W\gamma_k \tag{61}$$

- Again, the diagonal elements of this *impact matrix* represent the direct effects whereas the off-diagonal terms represent indirect effects. Contrary to the Spatial Durbin Model, here the impacts are only *local*, coming from neighboring units, W playing the role of a *local spatial multiplier* or *local interaction multiplier*.

Spatial or interaction multipliers: Special cases 2

The cross regressive model

- Consider now the following simpler model, which does not include the spatially lagged endogenous variable ($\rho = 0$) but includes exogenous variables together with spatially lagged exogenous variables ($\beta \neq 0$ and $\gamma \neq 0$) which may be estimated by OLS under the usual set of assumptions:

$$y = \beta_0 \iota_N + \sum_{k=1}^{K} (I_N \beta_k + W\gamma_k) X_k + \varepsilon \tag{60}$$

Let us take the partial derivatives of y relative to X_k for $k = 1, \ldots, K$:

$$\frac{\partial y}{\partial X_k'} = S_k(W) = I_N \beta_k + W\gamma_k \tag{61}$$

- Again, the diagonal elements of this *impact matrix* represent the direct effects whereas the off-diagonal terms represent indirect effects. Contrary to the Spatial Durbin Model, here the impacts are only *local*, coming from neighboring units, W playing the role of a *local spatial multiplier* or *local interaction multiplier*.

Spatial or interaction multipliers: Special cases 3

The spatially autocorrelated error model (SEM)

- Let us finally consider the following regression model with spatially autocorrelated errors:

$$y = \beta_0 \iota_N + X\beta + \varepsilon \qquad \varepsilon = \lambda W\varepsilon + u \tag{62}$$

which may be written as follows:

$$y = \beta_0 \iota_N + X\beta + (I_N - \lambda W)^{-1} u \tag{63}$$

Let us take the partial derivatives of y relative to X_k for $k = 1, \ldots, K$:

$$\frac{\partial y}{\partial X_k'} = S_k(W) = I_N \beta_k \tag{64}$$

which is exactly the same result as in the standard a-spatial regression model, where there are no spatial spillovers.

Spatial or interaction multipliers: Special cases 4

The constrained spatial Durbin model

- Note that this model may also be interpreted as a constrained Spatial Durbin Model:

$$(I_N - \rho W)y = (I_N - \lambda W)(\beta_0 \iota_N + X\beta) + u \tag{65}$$

$$y = (I_N - \rho W)\beta_0 \iota_N + \lambda W y + X\beta - \lambda WX\beta + u \tag{66}$$

As before, using a row-normalized interaction matrix W, one gets:

$$y = \frac{\beta_0}{1-\lambda}\iota_N + \lambda W y + X\beta - \lambda WX\beta + u \tag{67}$$

which is the Spatial Durbin Model:

$$y = \frac{\beta_0}{1-\lambda}\iota_N + \lambda W y + X\beta + WX\gamma + \varepsilon \tag{68}$$

with non linear constraints : $\gamma = -\lambda\beta$.

Impact of a random shock

The spatial Durbin model

- Reconsider the SDM model, in matrix form:

$$y = \beta_0 \iota_N + \rho W y + X\beta + WX\gamma + \varepsilon \tag{69}$$

- the reduced form is then:

$$y = (I - \rho W)^{-1} \iota_N \beta_0 + (I - \rho W)^{-1}(X\beta + WX\gamma) + (I - \rho W)^{-1}\varepsilon \tag{70}$$

- This specification also implies that a shock affecting one unit propagates to all the other units of the sample through the spatial transformation $(I - \rho W)^{-1}$.
- Note that the derivative of y with respect to ε is:

$$\frac{\partial y}{\partial \varepsilon'} = (I - \rho W)^{-1} = (I_N + \rho W + \rho^2 W^2 + \ldots) \tag{71}$$

which is the so-called *global spatial multiplier* or the *interaction multiplier*.

Impact of a random shock

The spatial Durbin model

- Reconsider the SDM model, in matrix form:

$$y = \beta_0 \iota_N + \rho W y + X\beta + WX\gamma + \varepsilon \tag{69}$$

- the reduced form is then:

$$y = (I - \rho W)^{-1} \iota_N \beta_0 + (I - \rho W)^{-1}(X\beta + WX\gamma) + (I - \rho W)^{-1}\varepsilon \tag{70}$$

- This specification also implies that a shock affecting one unit propagates to all the other units of the sample through the spatial transformation $(I - \rho W)^{-1}$.
- Note that the derivative of y with respect to ε is:

$$\frac{\partial y}{\partial \varepsilon'} = (I - \rho W)^{-1} = (I_N + \rho W + \rho^2 W^2 + \ldots) \tag{71}$$

which is the so-called *global spatial multiplier* or the *interaction multiplier*.

Impact of a random shock

The spatial Durbin model

- Reconsider the SDM model, in matrix form:

$$y = \beta_0 \iota_N + \rho W y + X\beta + WX\gamma + \varepsilon \tag{69}$$

- the reduced form is then:

$$y = (I - \rho W)^{-1} \iota_N \beta_0 + (I - \rho W)^{-1}(X\beta + WX\gamma) + (I - \rho W)^{-1}\varepsilon \tag{70}$$

- This specification also implies that a shock affecting one unit propagates to all the other units of the sample through the spatial transformation $(I - \rho W)^{-1}$.
- Note that the derivative of y with respect to ε is:

$$\frac{\partial y}{\partial \varepsilon'} = (I - \rho W)^{-1} = (I_N + \rho W + \rho^2 W^2 + \ldots) \tag{71}$$

which is the so-called *global spatial multiplier* or the *interaction multiplier*.

Impact of a random shock

The spatial Durbin model

- Reconsider the SDM model, in matrix form:

$$y = \beta_0 \iota_N + \rho W y + X\beta + WX\gamma + \varepsilon \tag{69}$$

- the reduced form is then:

$$y = (I - \rho W)^{-1} \iota_N \beta_0 + (I - \rho W)^{-1}(X\beta + WX\gamma) + (I - \rho W)^{-1}\varepsilon \tag{70}$$

- This specification also implies that a shock affecting one unit propagates to all the other units of the sample through the spatial transformation $(I - \rho W)^{-1}$.
- Note that the derivative of y with respect to ε is:

$$\frac{\partial y}{\partial \varepsilon'} = (I - \rho W)^{-1} = (I_N + \rho W + \rho^2 W^2 + \ldots) \tag{71}$$

which is the so-called *global spatial multiplier* or the *interaction multiplier*.

Impact of a random shock

The spatial Durbin model

- Let us define, as the *impact matrix* associated to the random shock, the $(N \times N)$ matrix $U(W) = (I_N - \rho W)^{-1}$. $U(W)$ is a $(N \times N)$ full matrix whose elements are:

$$U(W) = \begin{pmatrix} U(W)_{11} & U(W)_{12} & \dots & U(W)_{1N} \\ U(W)_{21} & U(W)_{22} & & U(W)_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ U(W)_{N1} & U(W)_{N2} & \dots & U(W)_{NN} \end{pmatrix} \tag{72}$$

- The partial derivatives of y_i relative to ε_i or ε_j for $i, j = 1, \dots, N, j \neq i$ are then:

$$\frac{\partial y_i}{\partial \varepsilon_i} = U(W)_{ii}, \qquad \frac{\partial y_i}{\varepsilon_j} = U(W)_{ij} \tag{73}$$

In general $U(W)_{ii} \neq 0$ and $U(W)_{ij} \neq 0$ for $i, j = 1, \dots, N, j \neq i$.

Impact of a random shock

The spatial Durbin model

- Let us define, as the *impact matrix* associated to the random shock, the $(N \times N)$ matrix $U(W) = (I_N - \rho W)^{-1}$. $U(W)$ is a $(N \times N)$ full matrix whose elements are:

$$U(W) = \begin{pmatrix} U(W)_{11} & U(W)_{12} & \dots & U(W)_{1N} \\ U(W)_{21} & U(W)_{22} & & U(W)_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ U(W)_{N1} & U(W)_{N2} & \dots & U(W)_{NN} \end{pmatrix} \tag{72}$$

- The partial derivatives of y_i relative to ε_i or ε_j for $i, j = 1, \dots, N$, $j \neq i$ are then:

$$\frac{\partial y_i}{\partial \varepsilon_i} = U(W)_{ii}, \qquad \frac{\partial y_i}{\varepsilon_j} = U(W)_{ij} \tag{73}$$

In general $U(W)_{ii} \neq 0$ and $U(W)_{ij} \neq 0$ for $i, j = 1, \dots, N$, $j \neq i$.

Impact of a random shock

The spatial Durbin model

- The diagonal elements of this matrix, $\mathrm{diag}(U(W))$, represent the *direct impacts* of a unitary random shock including "own spillover" effects.
- Again those impacts are heterogenous due to differentiated interaction terms in the W matrix.
- The off-diagonal elements of the impact matrix represent *indirect impacts* of the unitary random shock.
- Note also that the own derivative for unit i includes as previously the feedback effects where the unitary random shock on unit i affects unit j and unit j also affects unit i as well as longer paths which might go from unit i to j to k and back to i.
- The magnitude of those direct effects will now depend on: (1) the degree of interaction between countries, which is governed by the W matrix and (2) the parameter ρ, measuring the strength of spatial correlation between units.
- The magnitude of pure feedback effects are given by $U(W)_{ii} - 1$.
- Considering column j, we note that an unitary random shock in spatial unit j differently affects each of the spatial units of the sample.It represents the emission side of the spatial diffusion process.
- Considering row i, we note that unitary random shocks across all the units of the sample differently affects spatial unit i. It represents the reception side of the spatial diffusion process.

Impact of a random shock

The spatial Durbin model

- The diagonal elements of this matrix, $\mathrm{diag}(U(W))$, represent the *direct impacts* of a unitary random shock including "own spillover" effects.
- Again those impacts are heterogenous due to differentiated interaction terms in the W matrix.
- The off-diagonal elements of the impact matrix represent *indirect impacts* of the unitary random shock.
- Note also that the own derivative for unit i includes as previously the feedback effects where the unitary random shock on unit i affects unit j and unit j also affects unit i as well as longer paths which might go from unit i to j to k and back to i.
- The magnitude of those direct effects will now depend on: (1) the degree of interaction between countries, which is governed by the W matrix and (2) the parameter ρ, measuring the strength of spatial correlation between units.
- The magnitude of pure feedback effects are given by $U(W)_{ii} - 1$.
- Considering column j, we note that an unitary random shock in spatial unit j differently affects each of the spatial units of the sample.It represents the emission side of the spatial diffusion process.
- Considering row i, we note that unitary random shocks across all the units of the sample differently affects spatial unit i. It represents the reception side of the spatial diffusion process.

Impact of a random shock

The spatial Durbin model

- The diagonal elements of this matrix, $\mathrm{diag}(U(W))$, represent the *direct impacts* of a unitary random shock including "own spillover" effects.
- Again those impacts are heterogenous due to differentiated interaction terms in the W matrix.
- The off-diagonal elements of the impact matrix represent *indirect impacts* of the unitary random shock.
- Note also that the own derivative for unit i includes as previously the feedback effects where the unitary random shock on unit i affects unit j and unit j also affects unit i as well as longer paths which might go from unit i to j to k and back to i.
- The magnitude of those direct effects will now depend on: (1) the degree of interaction between countries, which is governed by the W matrix and (2) the parameter ρ, measuring the strength of spatial correlation between units.
- The magnitude of pure feedback effects are given by $U(W)_{ii} - 1$.
- Considering column j, we note that an unitary random shock in spatial unit j differently affects each of the spatial units of the sample.It represents the emission side of the spatial diffusion process.
- Considering row i, we note that unitary random shocks across all the units of the sample differently affects spatial unit i. It represents the reception side of the spatial diffusion process.

Impact of a random shock

The spatial Durbin model

- The diagonal elements of this matrix, $\mathrm{diag}(U(W))$, represent the *direct impacts* of a unitary random shock including "own spillover" effects.
- Again those impacts are heterogenous due to differentiated interaction terms in the W matrix.
- The off-diagonal elements of the impact matrix represent *indirect impacts* of the unitary random shock.
- Note also that the own derivative for unit i includes as previously the feedback effects where the unitary random shock on unit i affects unit j and unit j also affects unit i as well as longer paths which might go from unit i to j to k and back to i.
- The magnitude of those direct effects will now depend on: (1) the degree of interaction between countries, which is governed by the W matrix and (2) the parameter ρ, measuring the strength of spatial correlation between units.
- The magnitude of pure feedback effects are given by $U(W)_{ii} - 1$.
- Considering column j, we note that an unitary random shock in spatial unit j differently affects each of the spatial units of the sample.It represents the emission side of the spatial diffusion process.
- Considering row i, we note that unitary random shocks across all the units of the sample differently affects spatial unit i. It represents the reception side of the spatial diffusion process.

Impact of a random shock

The spatial Durbin model

- The diagonal elements of this matrix, $\mathrm{diag}(U(W))$, represent the *direct impacts* of a unitary random shock including "own spillover" effects.
- Again those impacts are heterogenous due to differentiated interaction terms in the W matrix.
- The off-diagonal elements of the impact matrix represent *indirect impacts* of the unitary random shock.
- Note also that the own derivative for unit i includes as previously the feedback effects where the unitary random shock on unit i affects unit j and unit j also affects unit i as well as longer paths which might go from unit i to j to k and back to i.
- The magnitude of those direct effects will now depend on: (1) the degree of interaction between countries, which is governed by the W matrix and (2) the parameter ρ, measuring the strength of spatial correlation between units.
- The magnitude of pure feedback effects are given by $U(W)_{ii} - 1$.
- Considering column j, we note that an unitary random shock in spatial unit j differently affects each of the spatial units of the sample.It represents the emission side of the spatial diffusion process.
- Considering row i, we note that unitary random shocks across all the units of the sample differently affects spatial unit i. It represents the reception side of the spatial diffusion process.

Impact of a random shock

The spatial Durbin model

- The diagonal elements of this matrix, $\mathrm{diag}(U(W))$, represent the *direct impacts* of a unitary random shock including "own spillover" effects.
- Again those impacts are heterogenous due to differentiated interaction terms in the W matrix.
- The off-diagonal elements of the impact matrix represent *indirect impacts* of the unitary random shock.
- Note also that the own derivative for unit i includes as previously the feedback effects where the unitary random shock on unit i affects unit j and unit j also affects unit i as well as longer paths which might go from unit i to j to k and back to i.
- The magnitude of those direct effects will now depend on: (1) the degree of interaction between countries, which is governed by the W matrix and (2) the parameter ρ, measuring the strength of spatial correlation between units.
- The magnitude of pure feedback effects are given by $U(W)_{ii} - 1$.
- Considering column j, we note that an unitary random shock in spatial unit j differently affects each of the spatial units of the sample.It represents the emission side of the spatial diffusion process.
- Considering row i, we note that unitary random shocks across all the units of the sample differently affects spatial unit i. It represents the reception side of the spatial diffusion process.

Impact of a random shock

The spatial Durbin model

- The diagonal elements of this matrix, $\mathrm{diag}(U(W))$, represent the *direct impacts* of a unitary random shock including "own spillover" effects.
- Again those impacts are heterogenous due to differentiated interaction terms in the W matrix.
- The off-diagonal elements of the impact matrix represent *indirect impacts* of the unitary random shock.
- Note also that the own derivative for unit i includes as previously the feedback effects where the unitary random shock on unit i affects unit j and unit j also affects unit i as well as longer paths which might go from unit i to j to k and back to i.
- The magnitude of those direct effects will now depend on: (1) the degree of interaction between countries, which is governed by the W matrix and (2) the parameter ρ, measuring the strength of spatial correlation between units.
- The magnitude of pure feedback effects are given by $U(W)_{ii} - 1$.
- Considering column j, we note that an unitary random shock in spatial unit j differently affects each of the spatial units of the sample.It represents the emission side of the spatial diffusion process.
- Considering row i, we note that unitary random shocks across all the units of the sample differently affects spatial unit i. It represents the reception side of the spatial diffusion process.

Impact of a random shock

The spatial Durbin model

- The diagonal elements of this matrix, $\mathrm{diag}(U(W))$, represent the *direct impacts* of a unitary random shock including "own spillover" effects.
- Again those impacts are heterogenous due to differentiated interaction terms in the W matrix.
- The off-diagonal elements of the impact matrix represent *indirect impacts* of the unitary random shock.
- Note also that the own derivative for unit i includes as previously the feedback effects where the unitary random shock on unit i affects unit j and unit j also affects unit i as well as longer paths which might go from unit i to j to k and back to i.
- The magnitude of those direct effects will now depend on: (1) the degree of interaction between countries, which is governed by the W matrix and (2) the parameter ρ, measuring the strength of spatial correlation between units.
- The magnitude of pure feedback effects are given by $U(W)_{ii} - 1$.
- Considering column j, we note that an unitary random shock in spatial unit j differently affects each of the spatial units of the sample.It represents the emission side of the spatial diffusion process.
- Considering row i, we note that unitary random shocks across all the units of the sample differently affects spatial unit i. It represents the reception side of the spatial diffusion process.

Simulating the impacts of heterogenous random shocks

- Let a_i be the magnitude of the shock affecting unit i and $\hat{\varepsilon}^i$ be the $(N \times 1)$ vector containing the estimated error of the model with a shock on the error term in unit i:

$$\hat{\varepsilon}^i = (\hat{\varepsilon}_1, ..., \hat{\varepsilon}_i + a_i, ..., \hat{\varepsilon}_N)' \tag{74}$$

- Therefore, the $(N \times 1)$ vector y^{i*} of the simulated dependent variable with a shock in unit i is:

$$y^{*i} = (I - \hat{\rho}W)^{-1}(\iota_N \hat{\beta}_0 + X\hat{\beta} + WX\hat{\gamma}) + (I - \hat{\rho}W)^{-1}\hat{\varepsilon}^i \tag{75}$$

$$y^{*i} = (I - \hat{\rho}W)^{-1}\widetilde{X}\hat{\delta} + (I - \hat{\rho}W)^{-1}\hat{\varepsilon}^i \tag{76}$$

where $\widetilde{X} = [\iota_N \;\; X \;\; WX]$, $\hat{\delta} = [\hat{\beta}_0 \;\; \hat{\beta} \;\; \hat{\gamma}]'$ and $\hat{\beta}_0$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\rho}$ are the Maximum Likelihood estimates of the unknown parameters in equation (69).

Simulating the impacts of heterogenous random shocks

- Let a_i be the magnitude of the shock affecting unit i and $\hat{\varepsilon}^i$ be the $(N \times 1)$ vector containing the estimated error of the model with a shock on the error term in unit i:

$$\hat{\varepsilon}^i = (\hat{\varepsilon}_1, ..., \hat{\varepsilon}_i + a_i, ..., \hat{\varepsilon}_N)' \tag{74}$$

- Therefore, the $(N \times 1)$ vector y^{i*} of the simulated dependent variable with a shock in unit i is:

$$y^{*i} = (I - \hat{\rho}W)^{-1}(\iota_N \hat{\beta}_0 + X\hat{\beta} + WX\hat{\gamma}) + (I - \hat{\rho}W)^{-1}\hat{\varepsilon}^i \tag{75}$$

$$y^{*i} = (I - \hat{\rho}W)^{-1}\widetilde{X}\hat{\delta} + (I - \hat{\rho}W)^{-1}\hat{\varepsilon}^i \tag{76}$$

where $\widetilde{X} = [\iota_N \;\; X \;\; WX]$, $\hat{\delta} = [\hat{\beta}_0 \;\; \hat{\beta} \;\; \hat{\gamma}]'$ and $\hat{\beta}_0$, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\rho}$ are the Maximum Likelihood estimates of the unknown parameters in equation (69).

Simulating the impacts of heterogenous random shocks 2

- Furthermore, let Y^* be the $(N \times N)$ matrix where each column i represents the simulated dependent variable for all units in the sample with a shock in unit i:

$$Y^* = [y^{*1} \ ... \ y^{*n}] = (I - \hat{\rho}W)^{-1}[\widetilde{X}\hat{\delta} \ ... \ \widetilde{X}\hat{\delta}] + (I - \hat{\rho}W)^{-1}\hat{\varepsilon}^* \tag{77}$$

where $\hat{\varepsilon}^* = [\hat{\varepsilon}^1 \ ... \ \hat{\varepsilon}^N]$ is a $(N \times N)$ matrix.

- Given the definition of $\hat{\varepsilon}^i$, the matrix $\hat{\varepsilon}^*$ can be rewritten as follows:

$$\hat{\varepsilon}^* = \begin{bmatrix} \hat{\varepsilon}_1 + a_1 & \hat{\varepsilon}_1 & \cdots & \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 & \hat{\varepsilon}_2 + a_2 & \cdots & \hat{\varepsilon}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\varepsilon}_N & \hat{\varepsilon}_N & \cdots & \hat{\varepsilon}_N + a_N \end{bmatrix}$$

$$\hat{\varepsilon}^* = \iota_N' \otimes \hat{\varepsilon} + A \tag{78}$$

where ι_N is the unit vector of dimension N and A is a diagonal matrix of order N, whose i^{th} diagonal element corresponds to a_i.

Simulating the impacts of heterogenous random shocks 2

- Furthermore, let Y^* be the $(N \times N)$ matrix where each column i represents the simulated dependent variable for all units in the sample with a shock in unit i:

$$Y^* = [y^{*1} \ ... \ y^{*n}] = (I - \hat{\rho}W)^{-1}[\widetilde{X}\hat{\delta} \ ... \ \widetilde{X}\hat{\delta}] + (I - \hat{\rho}W)^{-1}\hat{\varepsilon}^* \tag{77}$$

where $\hat{\varepsilon}^* = [\hat{\varepsilon}^1 \ ... \ \hat{\varepsilon}^N]$ is a $(N \times N)$ matrix.

- Given the definition of $\hat{\varepsilon}^i$, the matrix $\hat{\varepsilon}^*$ can be rewritten as follows:

$$\hat{\varepsilon}^* = \begin{bmatrix} \hat{\varepsilon}_1 + a_1 & \hat{\varepsilon}_1 & \dots & \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 & \hat{\varepsilon}_2 + a_2 & \dots & \hat{\varepsilon}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\varepsilon}_N & \hat{\varepsilon}_N & \dots & \hat{\varepsilon}_N + a_N \end{bmatrix}$$

$$\hat{\varepsilon}^* = \iota_N' \otimes \hat{\varepsilon} + A \tag{78}$$

where ι_N is the unit vector of dimension N and A is a diagonal matrix of order N, whose i^{th} diagonal element corresponds to a_i.

Simulating the impacts of heterogenous random shocks 3

- Therefore, we may write:

$$Y^* = (I - \hat{\rho}W)^{-1}(\iota_N' \otimes \widetilde{X}\hat{\delta}) + (I - \hat{\rho}W)^{-1}(\iota_N' \otimes \hat{\varepsilon} + A) \tag{79}$$

which may be also written:

$$Y^* = \iota_N' \otimes (I - \hat{\rho}W)^{-1}\widetilde{X}\hat{\delta} + \iota_N' \otimes (I - \hat{\rho}W)^{-1}(\hat{\varepsilon} + A) \tag{80}$$

- Finally, we compute the impact of a shock on unit i on all values of the dependent variable by calculating the difference between the simulated dependent variable Y^* and the matrix of observed dependent variable $Y = \iota_N' \otimes y$ with $y = (I - \hat{\rho}W)^{-1}(\widetilde{X}\hat{\delta} + \hat{\varepsilon})$:

$$Y^* - Y = (I - \hat{\rho}W)^{-1}A \tag{81}$$

- Note that when $A = I$ we retrieve the previous result on the impact of a unitary random shock. The implementation of shocks different from unity and different from each others just requires the post-multiplication of the impact matrix associated to the random shock by the matrix A.

Simulating the impacts of heterogenous random shocks 3

- Therefore, we may write:

$$Y^* = (I - \hat{\rho}W)^{-1}(\iota_N' \otimes \widetilde{X}\hat{\delta}) + (I - \hat{\rho}W)^{-1}(\iota_N' \otimes \hat{\varepsilon} + A) \tag{79}$$

  which may be also written:

$$Y^* = \iota_N' \otimes (I - \hat{\rho}W)^{-1}\widetilde{X}\hat{\delta} + \iota_N' \otimes (I - \hat{\rho}W)^{-1}(\hat{\varepsilon} + A) \tag{80}$$

- Finally, we compute the impact of a shock on unit i on all values of the dependent variable by calculating the difference between the simulated dependent variable Y^* and the matrix of observed dependent variable $Y = \iota_N' \otimes y$ with $y = (I - \hat{\rho}W)^{-1}(\widetilde{X}\hat{\delta} + \hat{\varepsilon})$:

$$Y^* - Y = (I - \hat{\rho}W)^{-1}A \tag{81}$$

- Note that when $A = I$ we retrieve the previous result on the impact of a unitary random shock. The implementation of shocks different from unity and different from each others just requires the post-multiplication of the impact matrix associated to the random shock by the matrix A.

Simulating the impacts of heterogenous random shocks 3

- Therefore, we may write:

$$Y^* = (I - \hat{\rho}W)^{-1}(\iota_N' \otimes \widetilde{X}\hat{\delta}) + (I - \hat{\rho}W)^{-1}(\iota_N' \otimes \hat{\varepsilon} + A) \tag{79}$$

  which may be also written:

$$Y^* = \iota_N' \otimes (I - \hat{\rho}W)^{-1}\widetilde{X}\hat{\delta} + \iota_N' \otimes (I - \hat{\rho}W)^{-1}(\hat{\varepsilon} + A) \tag{80}$$

- Finally, we compute the impact of a shock on unit i on all values of the dependent variable by calculating the difference between the simulated dependent variable Y^* and the matrix of observed dependent variable $Y = \iota_N' \otimes y$ with $y = (I - \hat{\rho}W)^{-1}(\widetilde{X}\hat{\delta} + \hat{\varepsilon})$:

$$Y^* - Y = (I - \hat{\rho}W)^{-1}A \tag{81}$$

- Note that when $A = I$ we retrieve the previous result on the impact of a unitary random shock. The implementation of shocks different from unity and different from each others just requires the post-multiplication of the impact matrix associated to the random shock by the matrix A.

Special cases

- The SAR model, with $\gamma = 0$ is a straightforward special case of the previous development. The pure SAR model with $\beta_0 = \beta = \gamma = 0$ is also a trivial special case.
- It must be stressed that the spatial diffusion of a random shock is the only property of the spatially autocorrelated error model (SEM), in which there are otherwise no spatial spillovers.
- For some examples see Ertur, Baumont and Legallo (2003) and LeGallo, Baumont, Dallerba and Ertur (2005).

Special cases

- The SAR model, with $\gamma = 0$ is a straightforward special case of the previous development. The pure SAR model with $\beta_0 = \beta = \gamma = 0$ is also a trivial special case.
- It must be stressed that the spatial diffusion of a random shock is the only property of the spatially autocorrelated error model (SEM), in which there are otherwise no spatial spillovers.
- For some examples see Ertur, Baumont and Legallo (2003) and LeGallo, Baumont, Dallerba and Ertur (2005).

Special cases

- The SAR model, with $\gamma = 0$ is a straightforward special case of the previous development. The pure SAR model with $\beta_0 = \beta = \gamma = 0$ is also a trivial special case.
- It must be stressed that the spatial diffusion of a random shock is the only property of the spatially autocorrelated error model (SEM), in which there are otherwise no spatial spillovers.
- For some examples see Ertur, Baumont and Legallo (2003) and LeGallo, Baumont, Dallerba and Ertur (2005).

References

- Anselin, L. (1988), *Spatial Econometrics: Method and Models*, Kluwer Academic Publishers, London, England.
- Anselin, L. (2006), Spatial Econometrics, in T. Mills and K. Patterson, eds., Palgrave *Handbook of Econometrics: Econometric Theory 1*, Palgrave Macmillan, Basingstoke, p.901-969.
- Ertur, C., Baumont, C. and Le Gallo, J. (2003), A spatial econometric analysis of convergence across european regions, 1980-1995, in B. Fingleton, ed., *European Regional Growth*, Springer-Verlag, p.99-130.
- Horn, R. and Johnson, C. (1985), *Matrix Analysis*, Cambridge University Press.
- Kelejian, H. H. and Prucha, I. R. (1998), A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *Journal of Real Estate Finance and Economics* 17, p.99-121.
- Kelejian, H. H. and Prucha, I. R. (1999), A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review* 40, p.509-533.
- Kelejian, H. H. and Prucha, I. R. (2010), Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, *Journal of Econometrics* 157, p.53-67.

References

- Le Gallo, J., Baumont, C., Dallerba, S. and Ertur, C. (2005), On the property of diffusion in the spatial error model, *Applied Economics Letters* 12, p.533-536.
- Lee, L.-F. (2002), Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models, *Econometric Theory* 18, p.252-277.
- Lee, L.-F. (2003), Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances, *Econometric Reviews* 22, p.307-335.
- Lee, L.-F. (2004a), Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models, *Econometrica* 72, p.1899-1925.
- Lee, L.-F. (2004b), Web appendix for the Econometrica 2004a paper:
- LeSage, J. and Pace, K. R. (2009), Introduction to Spatial Econometrics, CRC Press Taylor and Francis Group, New York.
- Ord, J. (1975), Estimation methods for models of spatial interaction, *Journal of the American Statistical Association* 70, p.120-126.