Maximum likelihood and generalized least squares estimation of spatial lag models with endogenous spatial coefficients: a Monte Carlo simulation

Myasnikov, Alexander

Plekhanov Russian University of Economics

March 2018

Оценка пространственных моделей с переменными коэффициентами пространственной чувствительности методом максимального правдоподобия и обобщенным методом наименьших квадратов

1. Введение

В отличие от национальных экономик, которые в экономическом анализе зачастую явно или неявно считают закрытыми системами, экономики субнациональных территориальных единиц – регионов, районов, графств и т.д.[1] – необходимо рассматривать как системы открытые, то есть испытывающие существенное влияние со стороны внешних по отношению к ним сил. В частности, региональные экономики, не имея существенных барьеров к движению капитала, труда или иных движимых факторов производства, неизбежно находятся в состоянии динамического неравновесия, ощущая влияние не только своих собственных, присущих каждому региону самому по себе, характеристик и особенностей, но также и воздействие со стороны соседних регионов. Поскольку те, в свою очередь, находятся в зависимости от экономических процессов, происходящих в их соседних регионах, а эти последние зависят от своих соседей и т.д., то вся система региональных экономик оказывается сложной средой распространения шоков. При этом сигнал, возникший в одном из регионов, по мере его распространения по всей сети регионов через цепочку соседей, постепенно затухает. Несмотря на это, каждый регион, входящий в систему, оказывается зависимым – по крайней мере в теории – от событий, происходящих даже в самых удаленных от него регионах из той же системы. Тем самым взаимодействие регионов друг с другом оказывается сложным процессом с многочисленными обратными связями.

Для анализа подобных взаимодействий между региональными экономиками обычно используется достаточно хорошо разработанный к настоящему времени инструментарий пространственной эконометрики. Главными моделями пространственной эконометрики являются модель пространственных лагов и модель пространственных ошибок. Первая используется в случаях, когда предполагается, что пространственные взаимодействия проявляются в том, что значения зависимой переменной для каждого региона зависят от значений той же самой переменной в соседних регионах. Примерами подобных взаимодействий могут быть технологические «переливы» между соседними регионами. Вторая применяется в случаях, когда подобные пространственные взаимодействия между значениями зависимой переменной в соседних регионах маловероятны или несущественны, но при этом предполагается, что соседние регионы все же влияют друг на друга некоторым образом, который не отражается во включенных в модель регрессорах. Обе модели предполагают, что степень чувствительности регионов к влиянию соседей постоянна. Однако в отношении стран, включающих большое количество весьма разнородных регионов, такое предположение может оказаться слишком сильным. Например, как показано в работе Демидовой, Иванова (2016), для валового регионального продукта (ВРП) российских регионов характерны пространственные взаимодействия с различными степенями чувствительности, зависящими от тех или иных характеристик каждого региона, воспринимающего сигналы от соседей.

В настоящей статье мы проводим имитационное моделирование по методу Монте-Карло модели пространственных лагов, включающей переменные коэффициенты пространственной чувствительности, которая может использоваться для анализа описанного выше типа межрегиональных взаимодействий через значения зависимой переменной, когда предполагается, что степень чувствительности каждого региона к внешним пространственным

[1] Для лаконичности мы везде далее будем говорить только о региональных экономиках и о взаимодействии соседних регионов, подразумевая при этом, что те же закономерности действуют и для более мелких территориальных образований.

эффектам может быть переменной, а именно может зависеть от тех или иных присущих этим регионам характеристик. При этом мы также описываем алгоритмы расширения описанных Anselin (1988) и Kelejian, Prucha (1998) методов оценки моделей пространственных лагов на рассматриваемый нами случай с переменными коэффициентами пространственной чувствительности.

Методология пространственного эконометрического анализа может быть применена не только для оценки собственно пространственных моделей как таковых, но также и в некоторых других ситуациях, подразумевающих сетевое взаимодействие между обособленными субъектами (регионами, городами, фирмами, индивидами и т.д.) – в том числе в случаях, когда чувствительность к внешним эффектам определяется иерархически или на основе социального (а не географического) расстояния (Corrado, Fingleton 2012). Рассматриваемая нами модель также может применяться не только для проведения собственно пространственных исследований, но также для анализа иных видов сетевых взаимодействий.

2. Модель и проблема выбора метода ее оценки

Традиционная модель пространственных лагов может быть записана в виде:

$$y = \rho W y + X^* \beta^* + \varepsilon, \qquad (1)$$

где y – вектор значений зависимой переменной размера $n \times 1$; ρ – постоянный коэффициент пространственной чувствительности; W – пространственная матрица («матрица соседства») размера $n \times n$; X^* – матрица значений независимых переменных размера $n \times k^*$; β^* – вектор коэффициентов при независимых переменных, имеющий размер $k^* \times 1$; ε – вектор ошибок размера $n \times 1$.

В случае, когда предполагается влияние на зависимую переменную не только со стороны значений той же переменной в соседних регионах, но также и со стороны независимых переменных, характеризующих соседние регионы, вместо этой модели может использоваться пространственная модель Дарбина. По сути, это та же модель (1), в которой матрица независимых переменных является блочной (Anselin 1988; Bekti et al. 2013):

$$y = \rho W y + X\beta + \varepsilon, \quad X = (X^* \mid W X^*), \qquad (2)$$

где X – блочная матрица независимых переменных и их пространственных лагов размера $n \times k$, β – вектор коэффициентов при независимых переменных и их лагах, имеющий размер $k \times 1$.

Поскольку базовая модель (1) может рассматриваться как частный случай модели (2), везде далее в тексте мы будем называть модель (2) базовой моделью пространственных лагов, имея при этом в виду, что те же самые замечания могли бы быть сделаны и в отношении модели (1).

Как базовая модель пространственных лагов, так и модель Дарбина основаны на допущении, что все регионы одинаково восприимчивы к воздействиям извне, что проявляется в постоянстве значения коэффициента ρ. Однако на практике это может быть не так: например, можно предположить, что интенсивность торговых отношений с соседними регионами (а значит, и интенсивность взаимовлияния ВРП) находится в зависимости от транспортной инфраструктуры региона и от численности и/или плотности его населения; степень восприимчивости региона к новым научным идеям может зависеть от количества расположенных в нем образовательных и научных учреждений и т.д. Для стран, чьи регионы характеризуются весьма высоким разнообразием климатических, социальных, экономических и прочих условий – таких, как Россия, – возможность подобного непостоянства чувствительности к пространственным внешним эффектам представляется особенно вероятной. Для эконометрического оценивания пространственных взаимодействий в условиях, когда пространственная чувствительность различных регионов непостоянна, но может быть смоделирована как зависящая от некоторых присущих регионам характеристик, в работе Демидовой, Иванова (2016) была предложена модель, которая может быть записана в общем виде следующим образом:

$$y = (\rho_0 I + \rho_1 Z)Wy + X\beta + \varepsilon, \qquad (3)$$

где I – единичная матрица размера $n \times 1$; $Z = diag(z)$ – диагональная матрица размера $n \times n$, значения элементов главной диагонали которой совпадают с элементами вектора z (имеющего размер $n \times 1$) значений переменной, которая, как предполагается, определяет степень чувствительности регионов к пространственным взаимодействиям; ρ_0 – скалярная величина, отражающая степень пространственной чувствительности регионов с нулевыми значениями z; ρ_1 – скалярный параметр, характеризующий степень влияния переменной z на пространственную чувствительность регионов.

Допуская наличие более чем одной переменной, определяющей степень пространственной чувствительности, мы можем переписать модель (3) в более общем виде следующим образом:

$$y = \left(\sum_{i=0}^{m} \rho_i Z_i \, W\right) y + X\beta + \varepsilon, \qquad (4)$$

где m – количество параметров регионов, влияющих на их чувствительность к пространственным взаимодействиям; ρ_i – коэффициент чувствительности к значениям i-го параметра z_i; $Z_i = diag(z_i)$ – диагональная матрица размера $n \times n$, составленная из значений i-го по счету параметра z_i ($i \in [0..m]$), имеющего размер $n \times 1$ и определяющего пространственную чувствительность (причем мы полагаем $Z_0 = I$). При $m = 0$ модель вырождается в базовую модель пространственного лага (2); при $m = 1$ имеем модель, описанную в работе Демидовой, Иванова (2016).

Нам не удалось найти в существующей литературе описания метода оценки модели пространственных лагов с переменным коэффициентом, аналогичной (4). При этом подходы к оценке базовой модели пространственных лагов (2) к настоящему моменту достаточно хорошо разработаны. Традиционным инструментом оценки базовой модели пространственных лагов (2) является метод максимального правдоподобия, подробно описанный в Anselin (1988). Данный подход реализован в основных программных пакетах для анализа пространственных регрессий, что делает его весьма удобным для практического применения. Однако он не лишен некоторых недостатков. А именно, этот метод основан на допущении о нормальности ошибок и, кроме того, подразумевает многократный расчет логарифма детерминанта матрицы $I - \rho W$ в процессе нахождения оптимального значения пространственного параметра ρ. Соответственно, проведение такого многократного расчета логарифма детерминанта создает определенные требования к вычислительной мощности, а также может требовать значительных затрат времени при больших размерах выборки. Впрочем, с учетом произошедшего за последние десятилетия существенного роста вычислительной мощности компьютеров можно предположить, что ограничения на использование метода максимального правдоподобия для оценки пространственных моделей, связанные с необходимостью длительных вычислений детерминанта матрицы $I - \rho W$, отошли на второй план – по крайней мере, для не слишком больших выборок.[2]

[2] Рост вычислительной мощности, вероятно, также означает, что применение приближенных вычислений для расчета якобиана в функции правдоподобия более не является необходимостью – по крайней мере, вплоть до некоторого размера выборки. Это означает рост точности результатов, получаемых с помощью метода максимального правдоподобия, вследствие снятия данных технических ограничений.

Главной альтернативой методу максимального правдоподобия при оценке пространственных моделей в настоящее время является подход, основанный на применении обобщенного метода наименьших квадратов. В работе Kelejian, Prucha (1999) описан порядок применения обобщенного метода моментов к оценке модели пространственных ошибок, в то время как в Kelejian, Prucha (1998) предлагается основанный на этом же методе трехступенчатый метод оценки пространственных моделей, содержащих одновременно как пространственный лаг зависимой переменной, так и пространственно-скоррелированные ошибки. Первый шаг данного метода оценки основан на обобщенном методе наименьших квадратов – этого шага достаточно, если пространственная модель содержит только пространственные лаги и не содержит пространственных ошибок. Таким образом, для оценки модели (4) достаточно применения только первого шага данной методики. Обобщенный метод наименьших квадратов (и – в более общем случае – обобщенный метод моментов) при оценивании пространственных данных предъявляет менее высокие требования к вычислительной мощности и времени расчета при больших объемах выборки (Bell, Bockstael (2000); Walde et al. (2008)) и, кроме того, допускает использование пространственных матриц более произвольной структуры, чем при применении метода максимального правдоподобия (Bell, Bockstael (2000)).

На больших выборках пространственных данных эффективность обобщенного метода моментов приближается к эффективности метода максимального правдоподобия (Bell, Bockstael (2000)). Мы не нашли исследований, посвященных анализу сравнительной эффективности двух методов оценки обобщенной модели пространственных лагов (4). Между тем сама модель (4) представляется достаточно перспективной для исследований пространственных взаимодействий в странах с высокой неоднородностью территориальных образований. Проблема выбора конкретного метода оценки моделей пространственных лагов с переменным пространственным коэффициентом при небольших объемах выборки, на наш взгляд, является достаточно актуальной для проведения пространственного анализа, например, российских региональных данных. В самом деле, достоверные статистические сведения о российской экономике на уровнях иерархии ниже регионального весьма скудны, что делает основным объектом пространственных исследований российской экономики именно регионы, а не субрегиональные единицы. Но размерность пространственной матрицы в случае проведения анализа на региональном уровне оказывается ограничена количеством российских регионов, которое находится вблизи 80 (точное количество регионов зависит от анализируемого периода). Тем самым, для значительной части пространственных исследований российской экономики размер выборки неизбежно является не очень большим. Можно предположить, что в ряде других стран с формирующимися рынками субрегиональная статистика также может быть сравнительно скудной, как в России, что означает небольшие размеры выборок при проведении пространственных исследований экономик и этих стран.

С учетом указанных обстоятельств, мы поставили перед собой задачу описать конкретные алгоритмы оценки модели (4) с помощью метода максимального правдоподобия и обобщенного метода моментов и сравнить эффективность оценок, полученных с применением этих методик на небольших выборках.

3. Подходы к оценке модели пространственных лагов с переменным коэффициентом чувствительности к пространственным взаимодействиям

Прежде всего, опишем алгоритмы оценки модели (4) с помощью двух рассматриваемых нами методов – начнем с метода максимального правдоподобия.

Уравнение (4) может быть переписано следующим образом:

$$\left(I - \sum_{i=0}^{m} \rho_i Z_i W\right) y = X\beta + \varepsilon \qquad (5)$$

По аналогии с Anselin (1988), введем обозначение $A = I - \sum_{i=0}^{m} \rho_i Z_i W$, тогда уравнение (5) можно будет записать в более простом виде:

$$Ay = X\beta + \varepsilon \qquad (6)$$

Предполагая, что ошибки ε независимы, будем иметь $E(\varepsilon\varepsilon') = \Omega$, где Ω – диагональная матрица, так что существует вектор гомоскедастичных случайных возмущений $u = \Omega^{-\frac{1}{2}}\varepsilon$, откуда $\varepsilon = \Omega^{\frac{1}{2}}u$, так что уравнение (6) можно переписать следующим образом:

$$Ay = X\beta + \Omega^{\frac{1}{2}}u \qquad (7)$$

Решая его относительно гомоскедастичного возмущения u, получаем:

$$u = \Omega^{-\frac{1}{2}}(Ay - X\beta) \qquad (8)$$

Ввиду ненаблюдаемости гомоскедастичных ошибок u для нахождения оценок параметров нам потребуется трансформация переменной u в y – соответственно, запишем якобиан трансформации:

$$J \equiv \frac{du}{dy} = \left|\Omega^{-\frac{1}{2}}A\right| \qquad (9)$$

Для использования в логарифмической функции правдоподобия нам потребуется логарифм якобиана, который может быть записан следующим образом:

$$\ln J = -\frac{1}{2}\ln|\Omega| + \ln|A| \qquad (9')$$

Подставляя выражение $9'$ для логарифма якобиана в стандартное уравнение логарифмической функции правдоподобия для нормально распределенных случайных переменных получаем два эквивалентных выражения для логарифма функции правдоподобия:

$$L = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Omega| + \ln|A| - \frac{1}{2}u'u \qquad (10)$$

$$L = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Omega| + \ln|A| - \frac{1}{2}(Ay - X\beta)'\Omega^{-1}(Ay - X\beta) \qquad (10')$$

Перепишем уравнение $(10')$, заменив коэффициенты β на их оценки $\hat{\beta}$:

$$L = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Omega| + \ln|A| - \frac{1}{2}\left(Ay - X\hat{\beta}\right)'\Omega^{-1}\left(Ay - X\hat{\beta}\right) \qquad (11)$$

Из (6) следует, что в общем виде оценки коэффициентов $\hat{\beta}$ по доступному обобщенному методу наименьших квадратов могут быть записаны следующим образом:

$$\hat{\beta} = (X'X)^{-1}X'Ay \qquad (12)$$

Раскрывая выражение для A, перепишем уравнение (12):

$$\hat{\beta} = (X'X)^{-1}X'Ay = (X'X)^{-1}X'\left(I - \sum_{i=0}^{m} \rho_i Z_i\, W\right)y$$
$$= (X'X)^{-1}X'Iy - (X'X)^{-1}X'\left(\sum_{i=0}^{m} \rho_i Z_i\, W\right)y \quad (13)$$

Воспользуемся тем, что $Z_0 = I$, а ρ_i являются скалярными величинами, и перепишем выражение (13) для $\hat{\beta}$, дополнительно убрав из него незначащие множители в виде единичных матриц:

$$\hat{\beta} = (X'X)^{-1}X'y - \rho_0(X'X)^{-1}X'Wy - \sum_{i=1}^{m}[\rho_i(X'X)^{-1}X'Z_iWy] \qquad (14)$$

Вводя обозначения $b_0 = (X'X)^{-1}X'y$, $b_L = (X'X)^{-1}X'Wy$, $b_{Z_iL} = (X'X)^{-1}X'Z_iWy$, перепишем выражение (14) в более простом виде:

$$\hat{\beta} = b_0 - \rho_0 b_L - \sum_{i=1}^{m} \rho_i b_{Z_iL} \qquad (15)$$

Векторы b_0, b_L, b_{Z_iL}, очевидно, имеют следующий эконометрический смысл:

- b_0 – это вектор оценок коэффициентов, полученных методом наименьших квадратов из обычной регрессии без пространственных лагов;
- b_L – вектор оценок коэффициентов, полученных методом наименьших квадратов из регрессии, в которой вектор значений эндогенной переменной был предумножен на матрицу пространственных весов – иными словами, b_L показывает, каким образом значения регрессоров в данном территориальном образовании влияют на средневзвешенные значения эндогенной переменной в соседних регионах (для того, чтобы далее, через пространственный лаг, повлиять обратно на значение зависимой переменной в данном регионе);
- b_{Z_iL} – вектор оценок коэффициентов, полученных методом наименьших квадратов из n регрессий, в каждой из которых вектор значений эндогенной переменной был предумножен на значения n-ой собственной характеристики соответствующих территориальных образований и на средневзвешенные значения эндогенной переменной в соседних регионах. Тем самым b_{Z_iL} показывает, в какой части значения регрессоров в самом регионе влияют на средневзвешенные значения эндогенной переменной в соседних регионах (с учетом их чувствительностей к пространственным внешним эффектам).

Соответственно, каждая из этих регрессий, определяющих b_0, b_L, b_{Z_iL}, порождает остатки:

$$e_0 = y - Xb_0$$
$$e_L = Wy - Xb_L$$
$$e_{Z_iL} = Z_iWy - Xb_{Z_iL}$$

Введем обозначение e для остатков пространственной регрессии и, пользуясь равенством (14), выведем соответствующее выражение для суммарного остатка:

$$e \equiv Ay - X\hat{\beta} = \left(I - \rho_0 W - \sum_{i=1}^{m} \rho_i Z_i W\right) y - X\left(b_0 - b_L - \sum_{i=1}^{m} b_{Z_i L}\right)$$
$$= (y - Xb_0) - \rho_0 (Wy - Xb_L) - \sum_{i=1}^{m} \rho_i \left(Z_i Wy - Xb_{Z_i L}\right)$$
$$= e_0 - \rho_0 e_L - \sum_{i=1}^{m} \rho_i e_{Z_i L} \quad (16)$$

По аналогии с наиболее распространенным подходом к анализу пространственных моделей типа (2), в дальнейшем будем предполагать ошибки ε гомоскедастичными: $\Omega \equiv \varepsilon' \varepsilon = \sigma^2 I$ – тогда будет иметь место равенство $\ln|\Omega| = n \ln \sigma^2$. По определению выборочной дисперсии остатков имеем: $\hat{\sigma}^2 = \frac{1}{n} e'e \Rightarrow e'e = n\hat{\sigma}^2$. С учетом этого, логарифмическая функция правдоподобия (11), записанная с использованием оценок параметров $\hat{\beta}$ и $\hat{\sigma}^2$, может быть сконцентрирована относительно них:

$$L = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \hat{\sigma}^2 + \ln|A| - \frac{1}{2\sigma^2}\left(Ay - X\hat{\beta}\right)'\left(Ay - X\hat{\beta}\right) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \hat{\sigma}^2 + \ln|A| - \frac{1}{2\hat{\sigma}^2} e'e$$
$$= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \hat{\sigma}^2 + \ln|A| - \frac{1}{2\hat{\sigma}^2} n\hat{\sigma}^2$$
$$= -\frac{n}{2} - \frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\left(\frac{1}{n} e'e\right) + \ln|A| \quad (17)$$

Для оценки параметров ρ_i необходимо осуществить оптимизацию логарифмической функции правдоподобия (17).[3] Подставляя полученные оптимальные значения параметров ρ_i в (15), получим искомые оценки коэффициентов $\hat{\beta}$.

Для расчета асимптотических стандартных ошибок оценок $\hat{\beta}$ воспользуемся асимптотической ковариационной матрицей $\hat{V} = \widehat{Var}_H(\Theta)$, полученной путем обращения информационной матрицы $I(\Theta)$, совпадающей с гессианом $H(\Theta)$ логарифмической функции правдоподобия (10), взятым с противоположным знаком:

$$\hat{V} = \widehat{Var}_H(\Theta) = I^{-1}(\Theta) = -H^{-1}(\Theta)$$

Гессиан логарифмической функции правдоподобия в общем случае может быть записан следующим образом (мы предполагаем наличие m переменных, определяющих чувствительность регионов к внешним эффектам, так что коэффициенты чувствительности ρ_i формируют вектор ρ размера $(m+1) \times 1$):

$$H = \begin{pmatrix} \frac{\partial^2 L}{\partial \rho^2} & \frac{\partial^2 L}{\partial \beta \partial \rho} & \frac{\partial^2 L}{\partial \sigma^2 \partial \rho} \\ \frac{\partial^2 L}{\partial \rho \partial \beta} & \frac{\partial^2 L}{\partial \beta^2} & \frac{\partial^2 L}{\partial \sigma^2 \partial \beta} \\ \frac{\partial^2 L}{\partial \rho \partial \sigma^2} & \frac{\partial^2 L}{\partial \beta \partial \sigma^2} & \frac{\partial^2 L}{\partial (\sigma^2)^2} \end{pmatrix} \quad (18)$$

Элементы гессиана (18), расположенные на главной диагонали и над ней, имеют следующий общий вид:

[3] Именно на этом шаге возникает потребность в высокой вычислительной мощности. Однако с учетом мощностей современной компьютерной техники, как мы уже отмечали выше, в случае использования не слишком больших матриц пространственных весов проблем при проведении такой оптимизации возникать не должно.

$$\frac{\partial^2 L}{\partial\rho_i\partial\rho_j} = -tr\left(A^{-1}Z_jWA^{-1}Z_iW\right) - \frac{y'W'Z_j'Z_iWy}{\sigma^2} \quad (19)$$

$$\frac{\partial^2 L}{\partial\beta\rho_i} = -\frac{y'W'Z_iX}{\sigma^2} \quad (20)$$

$$\frac{\partial^2 L}{\partial\sigma^2\partial\rho_i} = -\frac{(Ay - X\beta)'Z_iWy}{\sigma^4} \quad (21)$$

$$\frac{\partial^2 L}{\partial\beta^2} = -\frac{X'X}{\sigma^2} \quad (22)$$

$$\frac{\partial^2 L}{\partial\sigma^2\partial\beta} = -\frac{X'(Ay - X\beta)}{\sigma^4} \quad (23)$$

$$\frac{\partial^2 L}{\partial(\sigma^2)^2} = \frac{-n}{2\sigma^4} \quad (24)$$

Подробный ход выведения выражений (19)–(24) представлен в приложении.

В простом случае лишь одной переменной, влияющей на чувствительность регионов к пространственным внешним эффектам (т.е. в случае, когда $m = 1$ и мы имеем дело с двумя пространственными коэффициентами ρ_0, ρ_1 – как в работе Демидовой, Иванова (2016)) матрица A задается уравнением $A = I - \rho_0 W - \rho_1 ZW$, где матрица Z содержит на главной диагонали значения единственной переменной z, определяющей пространственную чувствительность регионов, и имеет размер $n \times n$. Гессиан логарифмической функции правдоподобия в этом случае приобретает следующий вид:

$$H = \begin{pmatrix} \frac{\partial^2 L}{\partial\rho_0^2} & \frac{\partial^2 L}{\partial\rho_1\partial\rho_0} & \frac{\partial^2 L}{\partial\beta\partial\rho_0} & \frac{\partial^2 L}{\partial\sigma^2\partial\rho_0} \\ \frac{\partial^2 L}{\partial\rho_0\partial\rho_1} & \frac{\partial^2 L}{\partial\rho_1^2} & \frac{\partial^2 L}{\partial\beta\partial\rho_1} & \frac{\partial^2 L}{\partial\sigma^2\partial\rho_1} \\ \frac{\partial^2 L}{\partial\rho_0\partial\beta} & \frac{\partial^2 L}{\partial\rho_1\partial\beta} & \frac{\partial^2 L}{\partial\beta^2} & \frac{\partial^2 L}{\partial\sigma^2\partial\beta} \\ \frac{\partial^2 L}{\partial\rho_0\partial\sigma^2} & \frac{\partial^2 L}{\partial\rho_1\partial\sigma^2} & \frac{\partial^2 L}{\partial\beta\partial\sigma^2} & \frac{\partial^2 L}{\partial(\sigma^2)^2} \end{pmatrix} \quad (25)$$

Как несложно убедиться, используя равенства (19)–(24), выражения для элементов гессиана (25), расположенных на главной диагонали или над ней, имеют следующий вид:

$$\frac{\partial^2 L}{\partial\rho_0^2} = -tr(A^{-1}WA^{-1}W) - \frac{y'W'Wy}{\sigma^2} \quad (26)$$

$$\frac{\partial^2 L}{\partial\rho_1\partial\rho_0} = -tr(A^{-1}WA^{-1}ZW) - \frac{y'W'ZWy}{\sigma^2} \quad (27)$$

$$\frac{\partial^2 L}{\partial\beta\partial\rho_0} = -\frac{y'W'X}{\sigma^2} \quad (28)$$

$$\frac{\partial^2 L}{\partial\sigma^2\partial\rho_0} = -\frac{(Ay - X\beta)'Wy}{\sigma^4} \quad (29)$$

$$\frac{\partial^2 L}{\partial\rho_1^2} = -tr(A^{-1}ZWA^{-1}ZW) - \frac{y'W'Z'ZWy}{\sigma^2} \quad (30)$$

$$\frac{\partial^2 L}{\partial \beta \partial \rho_1} = -\frac{y'W'Z'X}{\sigma^2} \quad (31)$$

$$\frac{\partial^2 L}{\partial \sigma^2 \partial \rho_1} = -\frac{(Ay - X\beta)'ZWy}{\sigma^4} \quad (32)$$

$$\frac{\partial^2 L}{\partial \beta^2} = -\frac{X'X}{\sigma^2} \quad (33)$$

$$\frac{\partial^2 L}{\partial \sigma^2 \partial \beta} = -\frac{X'(Ay - X\beta)}{\sigma^4} \quad (34)$$

$$\frac{\partial^2 L}{\partial (\sigma^2)^2} = -\frac{n}{2\sigma^4} \quad (35)$$

Алгоритм оценки модели (4) с помощью обобщенного метода наименьших квадратов еще проще, чем с помощью метода максимального правдоподобия. Kelejian, Prucha (1998) предлагают для оценки пространственного коэффициента в модели, аналогичной (2), применять матрицу инструментов K,[4] составленную из линейно независимых столбцов матриц $X, WX, W^2X, \ldots$ Используя эту матрицу, они записывают оценку вектора $\hat{\delta}$ коэффициентов пространственной модели, аналогичной (2), в следующем виде:

$$\tilde{\delta} = \left(\hat{Z}'\hat{Z}\right)^{-1}\hat{Z}'y, \quad (36)$$

где $\hat{\delta} = \left(\hat{\beta} \mid \rho\right)'$ – вектор размера $(k+1) \times 1$ оценок коэффициентов при экзогенных параметрах и при пространственной матрице, соответственно; $\hat{Z} = P_K(X \mid Wy)$ – матрица размера $n \times (k+1)$, в которой $P_K = K(K'K)^{-1}K'$ – матрица размера $n \times n$.

Для оценки модели (4) может быть использовано то же самое уравнение (36) – необходимо лишь поменять определение матрицы $\hat{Z}$ следующим образом:

$$\hat{Z} = P_K(X \mid \widehat{\widehat{Wy}}), \quad (37)$$

где $\widehat{\widehat{Wy}} = (Wy \mid Z_1Wy \mid \ldots \mid Z_mWy)$ – матрица размера $n \times (m+1)$, так что блочная матрица $(X \mid \widehat{\widehat{Wy}})$ имеет размер $n \times (k+m+1)$.

Доказательство того, что выражения (36) и (37) позволяют получить оценки коэффициентов модели (4), полностью аналогично приведенному в Kelejian, Prucha (1998); необходимые при этом допущения также аналогичны допущениям, описанным в той же работе.

4. Результаты имитационного моделирования

Для сравнения эффективности двух описанных выше методов оценки модели (4) мы провели имитационное моделирование по методу Монте-Карло. Для этого мы использовали модель (4), включающую три пространственных коэффициента: $\rho_0 = 0.5$, $\rho_1 = 0.7$ и $\rho_2 = -0.3$.[5] В качестве матрицы пространственных весов мы использовали фактическую матрицу пространственных весов по российским регионам, различные ее усеченные (в соответствии с размером экспериментальной выборки) версии, а также сгенерированные случайным образом

---

[4] Во избежание путаницы обозначений между двумя рассматриваемыми нами методами оценки модели (4) мы изменили обозначения некоторых переменных, векторов и матриц по сравнению с обозначениями, использованными в Kelejian, Prucha (1998).

[5] Мы также проводили имитационное моделирование с использованием нескольких наборов случайных значений ρ_0, ρ_1, ρ_2, получая при этом те же качественные результаты.

матрицы пространственных весов необходимых размерностей. В качестве переменных z_1, z_2, определяющих степень чувствительности к внешним эффектам, мы также использовали как фактические данные по российским регионам (в частности, такие переменные, как «Отношение суммы кредитов нефинансовым организациям к ВРП» и «Плотность населения региона»), так и сгенерированные случайным образом векторы значений. Для каждого варианта модели мы использовали по 500 случайных выборок значений экзогенных переменных X, предполагая при этом наличие в модели двух экзогенных переменных (не считая вектора единиц).

Моделирование мы проводили в среде языка R, используя для этого функции его базовых библиотек. Аппаратная среда состояла из персонального компьютера под управлением 64-битной Windows 10 Pro с процессором Intel® Core™ i5 (2.4 ГГц) и 8 ГБ оперативной памяти.[6]

Поскольку результаты имитаций с использованием реальных пространственных весов и значений переменных z_1, z_2 по российским регионам качественно совпали с результатами имитаций с использованием полностью случайных значений тех же параметров, для выборок с размерами, отличающимися от количества российских регионов, ниже мы приводим результаты имитации только для моделей со случайными параметрами.

Поскольку качество оценок коэффициентов модели (4), полученных с помощью метода максимального правдоподобия, может зависеть от того, следуют ли ошибки нормальному распределению, для имитации мы используем три разных распределения (аналогично Kelejian, Prucha (1998)):

1) Стандартное нормальное распределение ($\varepsilon_j \sim N(0{,}1)$). При описании результатов моделирования мы используем для данного распределения обозначение D_1;
2) логнормальное распределение ($\varepsilon_j = \frac{e^{\xi_j} - e^{0.5}}{\sqrt{e^2 - e}}$, $\xi_j \sim N(0{,}1)$; $\varepsilon_j \sim iid(0{,}1)$). При описании результатов моделирования мы используем для данного распределения обозначение D_2;
3) распределение случайной величины, которая определяется следующим образом: в 95% случаев она совпадает со стандартной нормальной случайной величиной, а в 5% случаев – с нормально распределенной случайной величиной, имеющей нулевое среднее значение и дисперсию 100 ($\varepsilon_j = \frac{\lambda_j \xi_j + (1-\lambda_j)\zeta_j}{\sqrt{5.95}}$, $\xi_j \sim N(0{,}1)$, $\zeta_j \sim N(0{,}100)$, $p(\lambda_j = 1) = 0.95$; $\varepsilon_j \sim iid(0{,}1)$). При описании результатов моделирования мы используем для данного распределения обозначение D_3.

В таблицу 1 (см. приложение) сведены результаты имитационного моделирования пространственной модели (4), основанные на неверном предположении о том, что данные порождены моделью (2). Эти оценки получены с помощью метода наименьших квадратов, а также с помощью традиционного подхода к анализу пространственных эффектов, основанного на применении метода максимального правдоподобия. В обоих случаях используется базовая спецификация модели пространственных лагов (2). Поскольку такая спецификация является ошибочной по сравнению с истинной спецификацией (4), использованной для генерации случайных данных, не удивительно, что оба подхода дают сильно смещенные результаты в части коэффициентов β при экзогенных регрессорах.

Таблица 2 приводит результаты имитационного моделирования на тех же данных, но уже на основе модели с правильной спецификацией (4). При этом используются метод максимального правдоподобия, описанный в уравнениях (10)–(35), и обобщенный метод наименьших квадратов, описанный в уравнениях (36)–(37). Из таблицы хорошо видно, что метод максимального

[6] Windows, Intel и Core являются торговыми марками соответствующих правообладателей.

правдоподобия практически во всех случаях характеризуется меньшей смещенностью оценок и их более высокой эффективностью по сравнению с обобщенным методом наименьших квадратов – однако на сравнительно большой выборке из 78 наблюдений эта разница в случаях нормального (D_1) и логнормального (D_2) распределений истинных ошибок оказывается на уровне погрешности. Бросаются в глаза более высокие значения смещения и дисперсии оценок в случае небольшой выборки, состоящей лишь из 10 наблюдений. При этом на такой выборке оценки, полученные по методу максимального правдоподобия, также оказываются менее смещенными и более эффективными, чем оценки по обобщенному методу наименьших квадратов на всех трех типах распределения истинных ошибок.

Наконец, из таблицы 2 также видно, что в случае распределения истинных ошибок D_3 (Kelejian, Prucha (1998) называют такой тип распределения «загрязненным нормальным») оба метода оценки показывают худшие результаты по сравнению с данными, содержащими истинные ошибки по двум другим распределениям, – при этом и в данном случае метод максимального правдоподобия оказывается лучше обобщенного метода наименьших квадратов с точки зрения смещенности и эффективности. Однако разница в смещенности и эффективности оценок между двумя методами сравнительно невелика для всех коэффициентов, кроме коэффициента при векторе единиц (β_0).

С учетом полученных нами результатов имитационного моделирования нам представляется, что на небольших выборках пространственных данных, в которых предполагаются переменные значения коэффициентов пространственной чувствительности, зависящие от тех или иных характеристик объектов исследования (то есть если модель (4) предполагается верной спецификацией), предпочтительным является использование метода максимального правдоподобия. Данный метод оказывается лучше обобщенного метода наименьших квадратов с точки зрения смещенности и эффективности, а при современном уровне вычислительной мощности компьютерных систем он не требует слишком больших затрат времени для оценки пространственных моделей на небольших выборках. В то же время применение обобщенного метода наименьших квадратов выглядит целесообразным как способ проверки робастности оценок, полученных методом максимального правдоподобия. С учетом существенно более высокой простоты реализации обобщенного метода наименьших квадратов данный метод также может применяться для проведения «пристрелочного» оценивания пространственных моделей.

5. Выводы

Модели пространственного лага с переменными коэффициентами пространственной чувствительности могут быть полезны для анализа пространственных взаимодействий между территориальными единицами, достаточно сильно отличающимися друг от друга по тем или иным параметрам и в этой связи способными по-разному реагировать на внешние пространственные эффекты – например, можно предположить, что именно такова ситуация с российскими регионами. Если при моделировании такого рода пространственных взаимодействий предполагается нормальное распределение ошибок и если выборка не слишком велика (так что не возникает чрезмерных требований к вычислительной мощности), то для эконометрической оценки подобных моделей может использоваться алгоритм, основанный на методе максимального правдоподобия. Результаты проведенного нами имитационного моделирования показали, что метод максимального правдоподобия также показывает более хорошие результаты и в случае некоторых других распределений ошибок – а именно, логнормального и «загрязненного» нормального (в терминологии Kelejian, Prucha (1998)). Обобщенный метод наименьших квадратов, алгоритм применения которого к оценке моделей пространственных лагов описан в Kelejian, Prucha (1998) и расширен в настоящей статье на случай моделей с переменными коэффициентами пространственной чувствительности, имеет смысл использовать на больших выборках, а в случаях оценивания моделей на небольших выборках может быть полезен в качестве средства для проверки робастности оценок, полученных с помощью метода максимального правдоподобия. Степени смещенности и эффективности оценок, полученных с помощью обобщенного метода наименьших квадратов при оценивании описанной в статье модели пространственных лагов с переменными коэффициентами пространственной чувствительности, достаточно близки к результатам, получаемым при использовании метода максимального правдоподобия, особенно на сравнительно больших выборках.

ЛИТЕРАТУРА

1. Anselin, L. (1988) Spatial econometrics: methods and models. Kluwer Academic Publishers.
2. Anselin, L. (2001). Spatial econometrics. A Companion to Theoretical Econometrics, 310–330. http://doi.org/10.1002/9780470996249.
3. Bekti, R. D., Rahayu, A., & Sutikno. (2013). Maximum Likelihood Estimation for Spatial Durbin Model. Journal of Mathematics and Statistics, 9(3), 169–174. http://doi.org/10.3844/jmssp.2013.169.174.
4. Bell, K. P., & Bockstael, N. E. (2000). Applying the Generalized-Moments Estimation Approach To Spatial Problems Involving Microlevel Data. Journal of Economic Growth, 82(February), 72–82. http://doi.org/10.1162/003465300558641.
5. Corrado, L., & Fingleton, B. (2012). Where is the economics in spatial econometrics? Journal of Regional Science, 52(2), 210–239. http://doi.org/10.1111/j.1467-9787.2011.00726.x.
6. Kelejian, H. H., & 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, 171, 99–121. http://doi.org/http://www.springerlink.com/link.asp?id=102945.
7. Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review, 40(2), 509–533.
8. Walde, J., Larch, M., & Tappeiner, G. (2008). Performance contest between MLE and GMM for huge spatial autoregressive models. Journal of Statistical Computation and Simulation, 78(2), 151–166. http://doi.org/10.1080/10629360600954109.
9. Демидова О.А., Иванов Д.С. (2016) Модели экономического роста с неоднородными пространственными эффектами (на примере российских регионов) // Экономический журнал ВШЭ. – Т. 20. – №1. – С.52 – 75.

ПРИЛОЖЕНИЯ

Приложение 1. Результаты имитационного моделирования

Таблица 1. Результаты имитационного моделирования модели (4) с ее оценкой методом наименьших квадратов и методом максимального правдоподобия, исходя из уравнения (2).

Распределение ошибок	**Источник значений W, z_1, z_2**	**Показатель**	**Метод наименьших квадратов, модель (2)**				**Метод максимального правдоподобия, модель (2)**			
			$\hat{\rho}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
D_1	**Реальные данные по российским регионам**	$\hat{\delta}-\delta$	-0.991	-1337.470	37.104	32.011	-0.764	-913.782	39.556	24.109
		σ	0.001	0.897	0.008	0.105	0.000	1.099	0.010	0.126
		RMSE	0.991	1337.470	37.104	32.011	0.764	913.782	39.556	24.109
		n	78							
	Случайные данные	$\hat{\delta}-\delta$	1.616	17067.989	22.972	-46.739	-0.248	-3111.225	-37.776	-179.285
		σ	0.001	10.451	0.038	0.133	0.000	2.791	0.017	0.068
		RMSE	1.616	17067.993	22.972	46.739	0.248	3111.226	37.776	179.285
		n	10							

Условные обозначения в таблице: $\hat{\delta}$ – средняя оценка параметра; δ – истинное значение параметра; $\hat{\delta}-\delta$ – среднее смещение оценки параметра; σ – стандартное отклонение оценки параметра; RMSE – корень среднего квадрата ошибки; n – размер выборки. Описание использованных в исследовании типов распределений ошибок дано в тексте работы.

Таблица 2. Результаты имитационного моделирования методом максимального правдоподобия и обобщенным методом наименьших квадратов, исходя из уравнения (4).

Распределение ошибок	Источник значений W, z_1, z_2	Показатель	Метод максимального правдоподобия, модель (4)						Обобщенный метод наименьших квадратов, модель (4)					
			$\hat{\rho}_0$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}_0$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
D_1	Реальные данные по российским регионам	$\hat{\delta}-\delta$	10^{-6}	-10^{-6}	10^{-7}	-0.015	10^{-5}	-10^{-4}	-10^{-5}	10^{-4}	-10^{-6}	0.024	10^{-4}	10^{-5}
		σ	10^{-6}	10^{-4}	10^{-5}	0.357	0.004	0.022	10^{-4}	0.005	10^{-5}	0.425	0.007	0.029
		RMSE	10^{-6}	10^{-4}	10^{-5}	0.358	0.004	0.022	10^{-4}	0.005	10^{-5}	0.425	0.007	0.029
		t	215						<1					
		n	78											
D_1	Случайные данные	$\hat{\delta}-\delta$	-10^{-5}	10^{-6}	-10^{-6}	-0.13	-0.002	-0.003	-10^{-5}	10^{-5}	10^{-5}	-0.224	-10^{-4}	0.002
		σ	10^{-4}	10^{-4}	10^{-4}	7.88	0.029	0.093	0.001	10^{-4}	10^{-4}	9.264	0.035	0.125
		RMSE	10^{-4}	10^{-4}	10^{-4}	7.89	0.029	0.093	0.001	10^{-4}	10^{-4}	9.266	0.035	0.125
		t	19						1					
		n	10											
D_2	Реальные данные по российским регионам	$\hat{\delta}-\delta$	10^{-5}	-10^{-5}	10^{-7}	0.001	10^{-4}	-10^{-4}	-10^{-5}	10^{-4}	10^{-6}	-0.004	0.001	0.001
		σ	10^{-4}	0.007	10^{-5}	0.32	0.004	0.019	0.001	0.006	10^{-5}	0.416	0.009	0.036
		RMSE	10^{-4}	0.007	10^{-5}	0.32	0.004	0.019	0.001	0.006	10^{-5}	0.416	0.009	0.036
		t	227						1					
		n												
D_2	Случайные данные	$\hat{\delta}-\delta$	-10^{-5}	10^{-6}	10^{-6}	-0.274	0.001	0.001	10^{-5}	10^{-5}	-10^{-5}	0.376	0.001	0.001
		σ	10^{-4}	10^{-4}	10^{-4}	8.665	0.017	0.079	0.001	10^{-4}	10^{-4}	17.149	0.021	0.08
		RMSE	10^{-4}	10^{-4}	10^{-4}	8.669	0.017	0.079	0.001	10^{-4}	10^{-4}	17.153	0.021	0.08
		t	24						<1					
		n												

Таблица 2 (продолжение). Результаты имитационного моделирования методом максимального правдоподобия и обобщенным методом наименьших квадратов, исходя из уравнения (4).

Распределение ошибок	**Источник значений W, z_1, z_2**	**Показатель**	**Метод максимального правдоподобия, модель (4)**						**Обобщенный метод наименьших квадратов, модель (4)**					
			$\hat{\rho}_0$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}_0$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
D_3	**Реальные данные по российским регионам**	$\hat{\delta}-\delta$	-10^{-5}	10^{-4}	10^{-6}	0.1	0.001	−0.008	-10^{-4}	0.002	-10^{-6}	0.317	−0.002	−0.009
		σ	0.001	0.005	10^{-4}	2.509	0.038	0.181	0.006	0.032	10^{-4}	4.551	0.063	0.206
		RMSE	0.001	0.005	10^{-4}	2.511	0.038	0.181	0.006	0.033	10^{-4}	4.562	0.063	0.206
		t	189						1					
		n	78											
D_3	**Случайные данные**	$\hat{\delta}-\delta$	-10^{-4}	10^{-4}	10^{-5}	−1.311	-10^{-4}	−0.005	-10^{-4}	10^{-4}	-10^{-4}	−2.387	0.006	0.042
		σ	0.004	0.002	0.002	40.686	0.23	0.774	0.004	0.003	0.003	47.024	0.253	1.04
		RMSE	0.004	0.002	0.002	40.707	0.23	0.774	0.004	0.003	0.003	47.085	0.253	1.04
		t	19						<1					
		n	10											

Условные обозначения в таблице: $\hat{\delta}$ – средняя оценка параметра по итогам расчета на базе 500 случайных выборок; δ – истинное значение параметра; $\hat{\delta}-\delta$ – среднее смещение оценки параметра; σ – стандартное отклонение оценки параметра; RMSE – корень среднего квадрата ошибки; t – суммарное время расчета модели для 500 выборок (в секундах); n – размер выборки. Описание использованных в исследовании типов распределений ошибок дано в тексте работы. Значения выше 0.0005 округляются с точностью до тысячных; для значений ниже 0.0005 приведены только их десятичные порядки.

Приложение 2. Выведение выражений для элементов гессиана логарифма функции правдоподобия

Логарифм функции правдоподобия в общем случае (когда количество пространственных коэффициентов равно m) при $\Omega = \sigma^2 I$ имеет вид $L = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 + \ln|A| - \frac{1}{2}u'u$, где $A = I - \sum_{i=0}^{m}\rho_i Z_i W$, $Z_0 = I, Z_i = Iz_i, u = \sigma^{-1}(Ay - X\beta)$.

Найдем первые частные производные логарифма функции правдоподобия:

$$\frac{\partial L}{\partial \rho_i} = \frac{\partial\left(-\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 + \ln|A| - \frac{1}{2}u'u\right)}{\partial \rho_i} = \frac{\partial(\ln|A|)}{\partial\rho_i} - \frac{1}{2}\frac{\partial(u'u)}{\partial\rho_i} = tr\left(A^{-1}\frac{\partial A}{\partial\rho_i}\right) - \frac{1}{2}\left(2\frac{\partial u}{\partial\rho_i}u\right)$$
$$= tr[A^{-1}(-Z_iW)] - \left\{\frac{\partial[\sigma^{-1}(Ay - X\beta)]}{\partial\rho_0}u\right\} = -tr(A^{-1}Z_iW) - \frac{1}{\sigma}(-Z_iWy)'u$$
$$= -tr(A^{-1}Z_iW) + \frac{y'W'Z_i'u}{\sigma}$$

$$\frac{\partial L}{\partial \beta} = \frac{\partial\left(-\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 + \ln|A| - \frac{1}{2}u'u\right)}{\partial \beta} = -\frac{1}{2}\frac{\partial(u'u)}{\partial\beta} = -\frac{1}{2}\left(2\frac{\partial u}{\partial\beta}u\right)$$
$$= -\left\{\frac{\partial[\sigma^{-1}(Ay - X\beta)]}{\partial\beta}u\right\} = -\frac{1}{\sigma}(-X'u) = \frac{X'u}{\sigma}$$

$$\frac{\partial L}{\partial \sigma^2} = \frac{\partial\left(-\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 + \ln|A| - \frac{1}{2}u'u\right)}{\partial \sigma^2} = -\frac{n}{2}\frac{\partial(\ln\sigma^2)}{\partial\sigma^2} - \frac{1}{2}\frac{\partial(u'u)}{\partial\sigma^2} = -\frac{n}{2\sigma^2} - \frac{1}{2}\left(2\frac{\partial u}{\partial\sigma^2}u\right)$$
$$= -\frac{n}{2\sigma^2} - \left\{\frac{\partial[(\sigma^2)^{-0.5}(Ay - X\beta)]}{\partial\sigma^2}u\right\} = -\frac{n}{2\sigma^2} + \frac{(Ay - X\beta)'u}{2(\sigma^2)^{1.5}}$$
$$= -\frac{n}{2\sigma^2} + \frac{(y'A' - \beta'X')u}{2\sigma^3}$$

Используя полученные выражения для первых частных производных, найдем вторые частные производные логарифма функции правдоподобия:

$$\frac{\partial^2 L}{\partial\rho_i\partial\rho_j} = \frac{\partial\left[-tr(A^{-1}Z_iW) + \frac{y'W'Z_i'u}{\sigma}\right]}{\partial\rho_j} = -\frac{\partial[tr(A^{-1}Z_iW)]}{\partial\rho_j} + \frac{1}{\sigma}\frac{\partial(y'W'Z_i'u)}{\partial\rho_j}$$
$$= -tr\left[\frac{\partial(A^{-1}Z_iW)}{\partial\rho_j}\right] + \frac{1}{\sigma}\left(\frac{\partial u}{\partial\rho_j}Z_iWy\right)$$
$$= tr\left(A^{-1}\frac{\partial A}{\partial\rho_j}A^{-1}Z_iW\right) + \frac{1}{\sigma}\left\{\frac{\partial[\sigma^{-1}(Ay - X\beta)]}{\partial\rho_j}Z_iWy\right\}$$
$$= tr\left[A^{-1}(-Z_jW)A^{-1}Z_iW\right] + \frac{1}{\sigma^2}\left(-Z_jWy\right)'Z_iWy$$
$$= -tr\left(A^{-1}Z_jWA^{-1}Z_iW\right) - \frac{y'W'Z_j'Z_iWy}{\sigma^2}$$

$$\frac{\partial^2 L}{\partial\rho_i\beta} = \frac{\partial\left[-tr(A^{-1}Z_iW) + \frac{y'W'Z_i'u}{\sigma}\right]}{\partial\beta} = \frac{1}{\sigma}\frac{\partial(y'W'Z_i'u)}{\partial\beta} = \frac{1}{\sigma}\frac{\partial u}{\partial\beta}Z_iWy = \frac{1}{\sigma}\frac{\partial[\sigma^{-1}(Ay - X\beta)]}{\partial\beta}Z_iWy$$
$$= \frac{1}{\sigma^2}(-X')Z_iWy = -\frac{X'Z_iWy}{\sigma^2}$$

$$\frac{\partial^2 L}{\partial\rho_i\partial\sigma^2} = \frac{\partial\left[-tr(A^{-1}Z_iW) + \frac{y'W'Z_i'u}{\sigma}\right]}{\partial\sigma^2} = \frac{\partial(\sigma^{-1}y'W'Z_i'u)}{\partial\sigma^2} = \frac{\partial[\sigma^{-1}y'W'Z_i'\sigma^{-1}(Ay - X\beta)]}{\partial\sigma^2}$$
$$= \frac{\partial[(\sigma^2)^{-1}y'W'Z_i'(Ay - X\beta)]}{\partial\sigma^2} = -\frac{(Ay - X\beta)'Z_iWy}{\sigma^4}$$

$$\frac{\partial^2 L}{\partial\beta\rho_i} = \frac{\partial\left(\frac{X'u}{\sigma}\right)}{\partial\rho_i} = \frac{1}{\sigma}\frac{\partial(X'u)}{\partial\rho_i} = \frac{1}{\sigma}\frac{\partial u}{\partial\rho_i}X = \frac{1}{\sigma}\frac{\partial[\sigma^{-1}(Ay - X\beta)]}{\partial\rho_i}X = \frac{1}{\sigma^2}(-Z_iWy)'X = -\frac{y'W'Z_iX}{\sigma^2}$$

$$\frac{\partial^2 L}{\partial\beta^2} = \frac{\partial\left(\frac{X'u}{\sigma}\right)}{\partial\beta} = \frac{1}{\sigma}\frac{\partial(X'u)}{\partial\beta} = \frac{1}{\sigma}\frac{\partial u}{\partial\beta}X = \frac{1}{\sigma}\frac{\partial[\sigma^{-1}(Ay - X\beta)]}{\partial\beta}X = \frac{1}{\sigma^2}(-X')X = -\frac{X'X}{\sigma^2}$$

$$\frac{\partial^2 L}{\partial\beta\partial\sigma^2} = \frac{\partial\left(\frac{X'u}{\sigma}\right)}{\partial\sigma^2} = \frac{\partial[\sigma^{-1}X'\sigma^{-1}(Ay - X\beta)]}{\partial\sigma^2} = \frac{\partial[(\sigma^2)^{-1}X'(Ay - X\beta)]}{\partial\sigma^2} = -\frac{(Ay - X\beta)'X}{\sigma^4}$$

$$\frac{\partial^2 L}{\partial\sigma^2\partial\rho_i} = \frac{\partial\left[-\frac{n}{2\sigma^2} + \frac{(y'A' - \beta'X')u}{2\sigma^3}\right]}{\partial\rho_i} = \frac{\partial\left[\frac{(y'A' - \beta'X')u}{2\sigma^3}\right]}{\partial\rho_i} = \frac{1}{2\sigma^3}\frac{\partial[(y'A' - \beta'X')u]}{\partial\rho_i}$$
$$= \frac{1}{2\sigma^3}\left[\frac{\partial(Ay - X\beta)}{\partial\rho_i}u + (y'A' - \beta'X')\left(\frac{\partial u}{\partial\rho_i}\right)'\right]$$
$$= \frac{1}{2\sigma^3}\left\{-y'W'Z_i'\sigma^{-1}(Ay - X\beta) + (y'A' - \beta'X')\left[\frac{\partial[\sigma^{-1}(Ay - X\beta)]}{\partial\rho_i}\right]'\right\}$$
$$= -\frac{1}{2\sigma^4}[y'W'Z_i'(Ay - X\beta) + (y'A' - \beta'X')Z_iWy]$$
$$= -\frac{y'W'Z_i'(Ay - X\beta) + (y'A' - \beta'X')Z_iWy}{2\sigma^4} = -\frac{(Ay - X\beta)'Z_iWy}{\sigma^4},$$

где последнее равенство получено благодаря тому, что оба слагаемых в числителе являются скалярными величинами, которые, очевидно, нечувствительны к транспонированию.

$$\frac{\partial^2 L}{\partial\sigma^2\partial\beta} = \frac{\partial\left[-\frac{n}{2\sigma^2} + \frac{(y'A' - \beta'X')u}{2\sigma^3}\right]}{\partial\beta} = \frac{\partial\left[\frac{(y'A' - \beta'X')u}{2\sigma^3}\right]}{\partial\beta} = \frac{1}{2\sigma^3}\frac{\partial[(y'A' - \beta'X')u]}{\partial\beta}$$
$$= \frac{1}{2\sigma^3}\left[\frac{\partial(Ay - X\beta)}{\partial\beta}u + \frac{\partial u}{\partial\beta}(Ay - X\beta)\right]$$
$$= \frac{1}{2\sigma^3}[-X'\sigma^{-1}(Ay - X\beta) - \sigma^{-1}X'(Ay - X\beta)] = -\frac{X'(Ay - X\beta)}{\sigma^4}$$

$$\frac{\partial^2 L}{\partial(\sigma^2)^2} = \frac{\partial\left[-\frac{n}{2\sigma^2} + \frac{(y'A' - \beta'X')u}{2\sigma^3}\right]}{\partial\sigma^2} = \frac{n}{2\sigma^4} + \frac{1}{2}\frac{\partial[\sigma^{-3}(y'A' - \beta'X')\sigma^{-1}(Ay - X\beta)]}{\partial\sigma^2}$$
$$= \frac{n}{2\sigma^4} + \frac{1}{2}\frac{\partial[(\sigma^2)^{-2}(y'A' - \beta'X')(Ay - X\beta)]}{\partial\sigma^2}$$
$$= \frac{n}{2\sigma^4} - (\sigma^2)^{-3}(y'A' - \beta'X')(Ay - X\beta) = \frac{n}{2\sigma^4} - \frac{(Ay - X\beta)'(Ay - X\beta)}{\sigma^6}$$
$$= \frac{n}{2\sigma^4} - \frac{e'e}{\sigma^6} = \frac{n}{2\sigma^4} - \frac{n\sigma^2}{\sigma^6} = \frac{n}{2\sigma^4} - \frac{n}{\sigma^4} = -\frac{n}{2\sigma^4}$$

Проверим, что полученные нами выражения для $\frac{\partial^2 L}{\partial\rho_i\partial\rho_j}$ и $\frac{\partial^2 L}{\partial\rho_j\partial\rho_i}$ соответствуют теореме Юнга (теореме о смешанных производных), то есть что $\frac{\partial^2 L}{\partial\rho_i\partial\rho_j} = \frac{\partial^2 L}{\partial\rho_j\partial\rho_i}$. Воспользуемся тем, что след произведения матриц нечувствителен к циклическим перестановкам $(tr(AB) = tr(BA))$ и коммутативностью скалярного произведения векторов ($a'b = b'a$). В самом деле, мы можем записать:

$$\frac{\partial^2 L}{\partial\rho_j\partial\rho_i} = -tr\left(A^{-1}Z_iWA^{-1}Z_jW\right) - \frac{y'W'Z_i'Z_jWy}{\sigma^2} = -tr\left[(A^{-1}Z_iW)\left(A^{-1}Z_jW\right)\right] - \frac{(Z_iWy)'(Z_jWy)}{\sigma^2}$$
$$= -tr\left[\left(A^{-1}Z_jW\right)(A^{-1}Z_iW)\right] - \frac{\left(Z_jWy\right)'(Z_iWy)}{\sigma^2}$$
$$= -tr\left(A^{-1}Z_jWA^{-1}Z_iW\right) - \frac{y'W'Z_j'Z_iWy}{\sigma^2} = \frac{\partial^2 L}{\partial\rho_i\partial\rho_j}$$