МОДЕЛИРОВАНИЕ СТРУКТУРНЫХ СДВИГОВ КОЭФФИЦИЕНТОВ В МОДЕЛЯХ ГЕОГРАФИЧЕСКИ ВЗВЕШЕННОЙ РЕГРЕССИИ

Балаш Ольга Сергеевна
Балаш Владимир Алексеевич
Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского

Географически взвешенная регрессия (GWR)

- Обобщает классическую модель линейной регрессии, коэффициенты оцениваются локально

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \xi_i$$

- (u_i, v_i) – географические координаты i-го объекта.

$$\beta = \begin{bmatrix} \beta_{0(u_i,v_i)} & \beta_{0(u_i,v_i)} & \beta_{0(u_i,v_i)} & .. & \beta_{0(u_i,v_i)} \\ \beta_{0(u_i,v_i)} & \beta_{0(u_i,v_i)} & \beta_{0(u_i,v_i)} & .. & \beta_{0(u_i,v_i)} \\ .. & .. & .. & .. & .. \\ \beta_{0(u_i,v_i)} & \beta_{0(u_i,v_i)} & \beta_{0(u_i,v_i)} & .. & \beta_{0(u_i,v_i)} \end{bmatrix}$$

- Коэффициенты изменяются в пространстве (в каждом местоположении свои коэффициенты)

$$\hat{\beta}(i) = (X^T W(i) X)^{-1} X^T W(i) Y$$

- Для каждой точки оценивается собственная модель, но соседи учитываются с большим весом

$$W(i) = \begin{bmatrix} w_{i1} & 0 & .. & 0 \\ 0 & w_{i2} & .. & 0 \\ .. & .. & .. & .. \\ 0 & 0 & 0 & w_{in} \end{bmatrix}$$

Оценка коэффициентов GWR

- Локально взвешенный метод наименьших квадратов
- Так как расчеты коэффициентов проводятся для всех измерений, то в результате получают матрицу оценок параметров:

$$\widehat{B}=\begin{bmatrix} \widehat{\beta}_0(u_1,v_1) & \widehat{\beta}_1(u_1,v_1) & \dots & \widehat{\beta}_p(u_1,v_1) \\ \widehat{\beta}_0(u_2,v_2) & \widehat{\beta}_1(u_2,v_2) & \dots & \widehat{\beta}_p(u_2,v_2) \\ \dots & \dots & \dots & \dots \\ \widehat{\beta}_0(u_n,v_n) & \widehat{\beta}_1(u_n,v_n) & \dots & \widehat{\beta}_p(u_n,v_n) \end{bmatrix}$$

Оценка коэффициентов GWR

- где *i*–ая строка представляет собой вектор оценок коэффициентов в точке (u_i, v_i)

$$\beta(i) = (X^T W(i) X)^{-1} X^T W(i) Y$$

Взвешивание

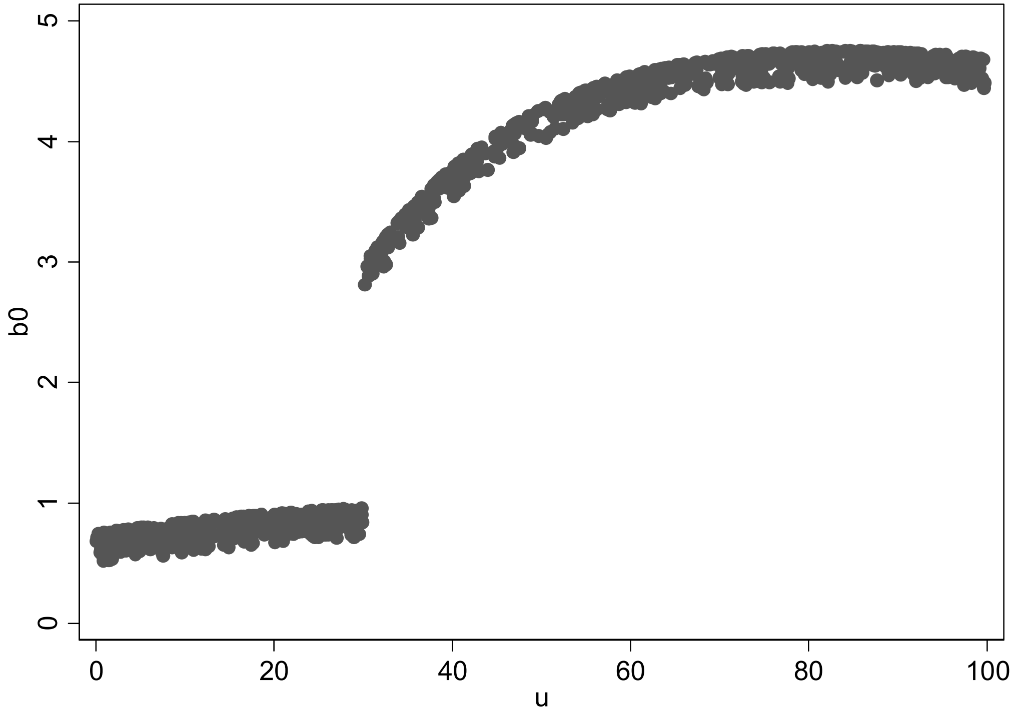
- Определение весовой функции
  - Наиболее часто применяют ядра Гаусса
  - и им подобные, например, би-квадрат или три-куб:
- Взвешивание может проводиться с использование фиксированной или адаптивной шириной полосы пропускания *b*

$$w_{ij} = \exp\left(-\frac{\alpha}{2}\left(\frac{d_{ij}}{b}\right)^2 \right)$$

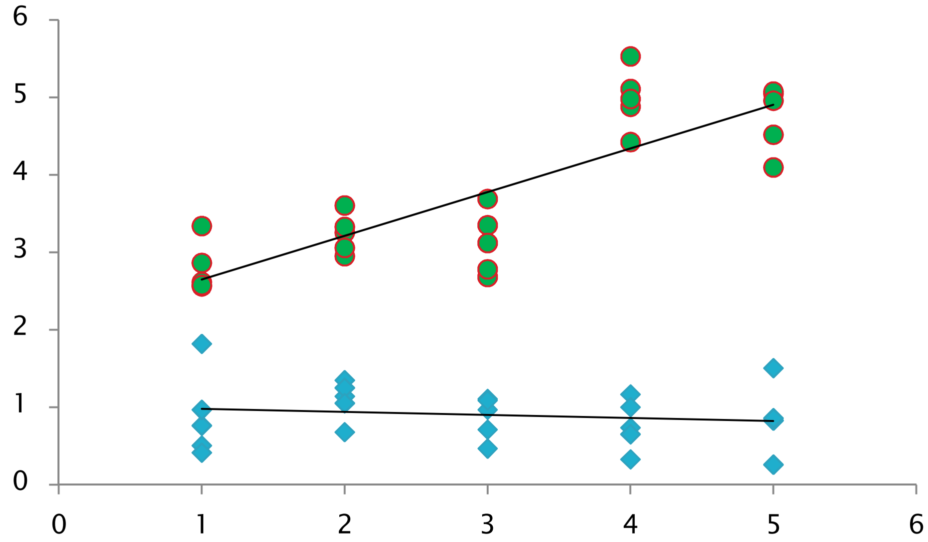
$$w_{ij} = \begin{cases} (1-(d_{ij}/b)^2)^2, \ d_{ij} < b \\ 0, \ d_{ij} \geq b \end{cases}$$

$$w_{ij} = \begin{cases} (1-(d_{ij}/b)^3)^4, \ d_{ij} < b \\ 0, \ d_{ij} \geq b \end{cases}$$

Скачкообразное изменение коэффициента β_0



Скачкообразное изменение зависимой переменной в подвыборке



✓остатки GWR модели не зависят от координат (*u,v)*;

✓есть скачок, неучтенный моделью GWR.

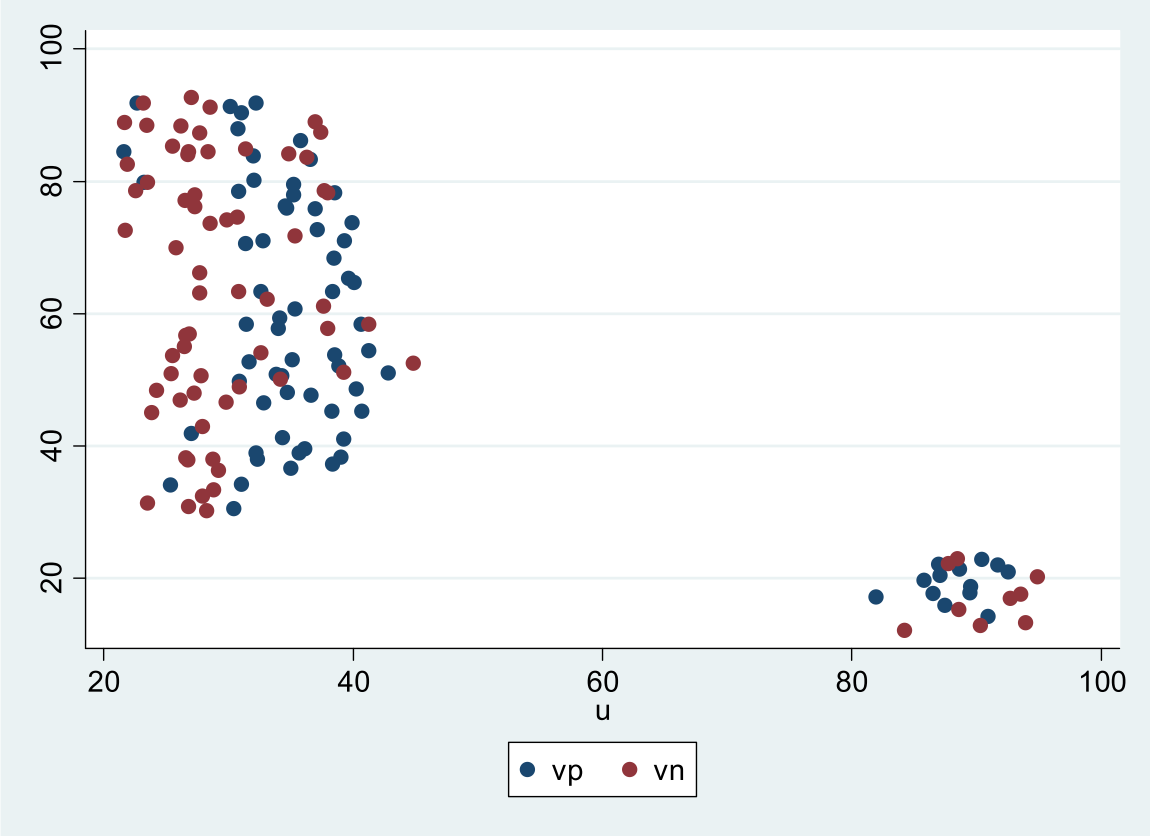
- Проверяется гипотеза о независимости величины (или знаков) остатков от координат. Строим регрессию для остатков:

$$e_i = \gamma_0 + \gamma_1 u_i + \gamma_2 v_i + \delta_i$$

- или логит-регрессию

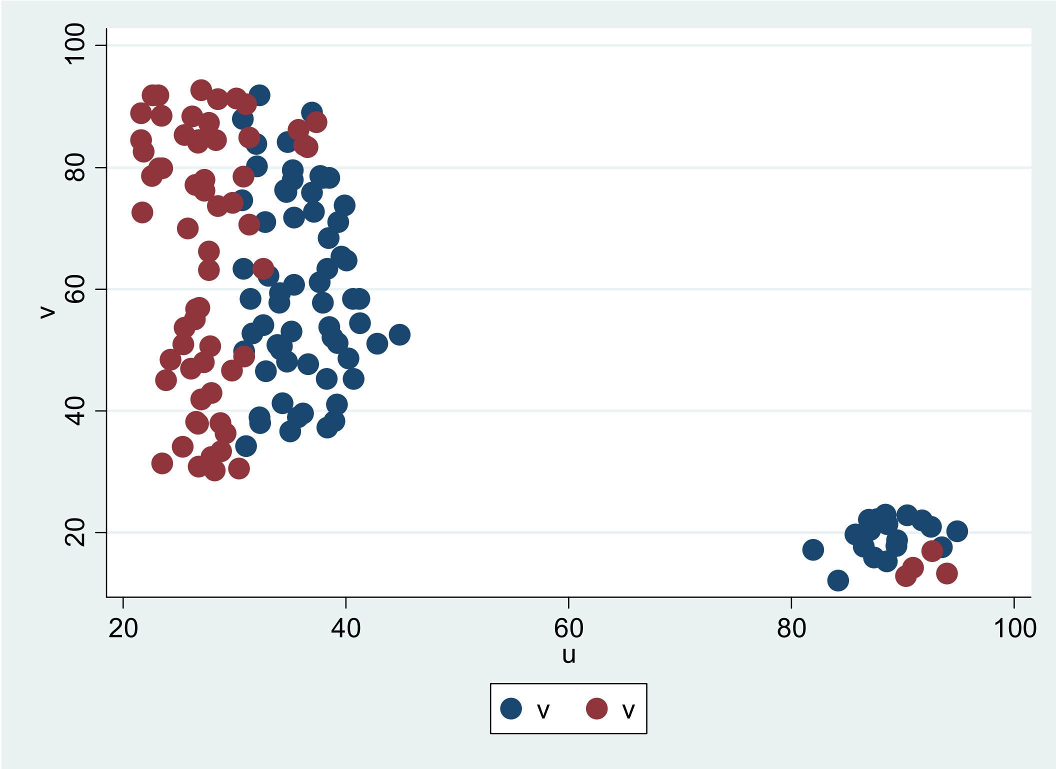
$$P(e_i) = \Lambda(\gamma_0 + \gamma_1 u_i + \gamma_2 v_i)$$

Значения меток



- M_1 – локальный прогноз по ГВР больше наблюдаемого значения;
- M_{-1} – локальный прогноз меньше наблюдаемого значения.

Уточнение границ областей



➢ Построение решающего правила методом непараметрического дискриминантного анализа (*k*–ближайших соседей);

➢ Отнесение к классам M_1 и M_{-1} с использованием полученного решающего правила.

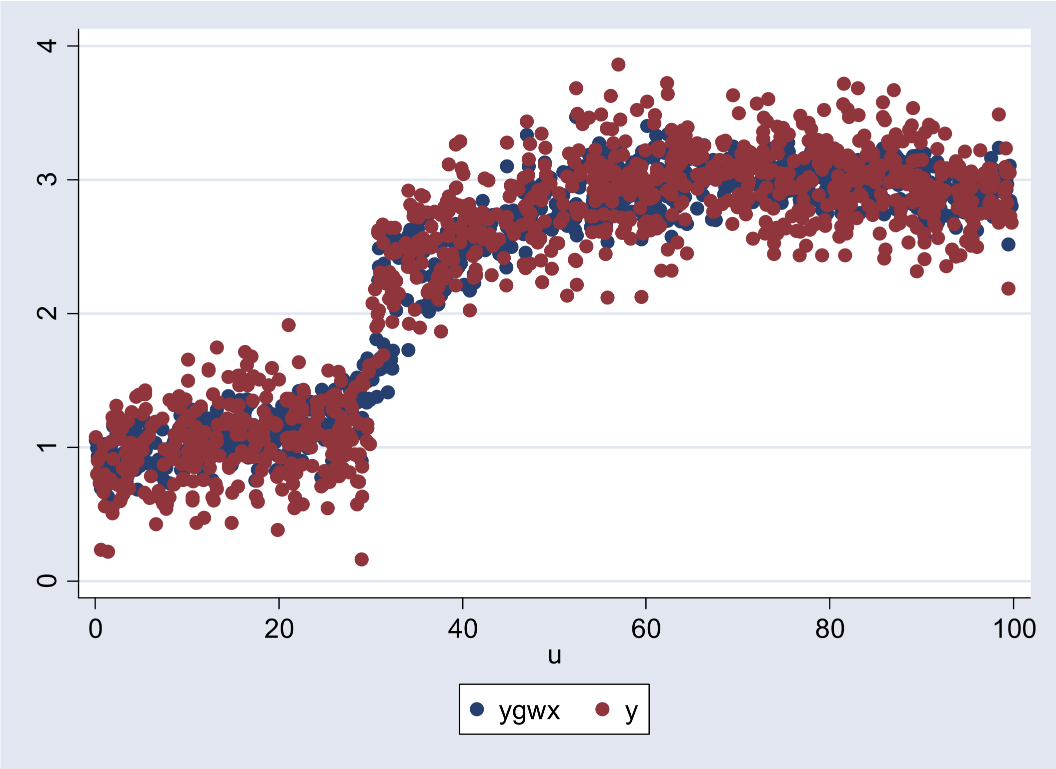
- Вид модели:

$$y_i = \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \beta_2(u_i, v_i)x_{i2} + \beta_3(u_i, v_i)d_i + \varepsilon_i$$

- где d_i фиктивная переменная

$$d_i = \begin{cases} 1, \text{если } (u_i . v_i) \in \mathrm{M}_1 \\ 0, \text{если } (u_i . v_i) \notin \mathrm{M}_1 \end{cases}$$

Прогноз модели ГВР с фиктивной переменной



Благодарю за внимание!