Theoretical Foundations of GWR with GNN-Based Adaptive Kernels

1. Model (Intuition and Formal Definition)

1.1 Intuisi Model (dengan GNN sejak awal)

Kita mempelajari hubungan antara variabel respon dan kovariat yang **bervariasi secara spasial**. Namun, kita **tidak ingin**:

- memaksakan bentuk kernel spasial tertentu (Gaussian, bisquare, dsb),
- karena struktur spasial **bisa kompleks** (barrier, anisotropi, klaster).

Sebagai gantinya:

> **Kita membiarkan bobot spasial dipelajari dari data**, tetapi **tetap menjaga lokalitas spasial** agar inferensi sah.

Dengan kata lain:

- **GWR tetap model inferensi utama**,
- **GNN hanya berperan sebagai *estimator kernel adaptif***.

1.2 Data dan Notasi

Kita mengamati n observasi independen:

$$\{(Y_i, X_i, U_i)\}_{i=1}^n$$

dengan:

- $Y_i \in \mathbb{R}$ — variabel respon (skalar)
- $X_i \in \mathbb{R}^p$ — vektor kovariat berdimensi p
- $U_i \in \mathbb{R}^d$ — lokasi spasial berdimensi d

1.3 Model Struktural (*data-generating process*)

Model sebenarnya diasumsikan:

$$\boxed{Y_i = X_i^\top \beta(U_i) + \varepsilon_i}$$

Penjelasan:

- $\beta(\cdot) : \mathbb{R}^d \to \mathbb{R}^p$ — fungsi koefisien spasial yang **halus**
- $\beta(U_i)$ — koefisien lokal sebenarnya di lokasi U_i
- ε_i — error acak dengan mean nol

Target inferensi **tetap**:

$$\boxed{\beta(u_0)}$$

untuk satu lokasi target tetap $u_0 \in \mathbb{R}^d$.

1.4 Prinsip Lokalitas (belum pakai kernel)

Agar $\beta(u_0)$ dapat diestimasi, kita perlu:

- hanya menggunakan observasi **cukup dekat** dengan u_0,
- tetapi **jumlahnya bertambah** saat n membesar.

Untuk itu, diperkenalkan **bandwidth** $h = h_n > 0$ dan neighborhood:

$$\mathcal{N}_h(u_0) = \{i : |U_i - u_0| \leq h\}$$

Dengan syarat asimtotik:

$$h \to 0, \qquad nh^d \to \infty$$

📌 **Sampai titik ini: belum ada kernel, belum ada GNN.** Ini murni struktur nonparametrik lokal.

---

2. Asumsi Klasik yang Digunakan

Asumsi berikut **bukan CLM global**, melainkan **versi lokal untuk GWR dengan kernel adaptif**.

(A1) Independensi

$$\{(X_i, U_i, \varepsilon_i)\}_{i=1}^n \quad \text{i.i.d.}$$

(A2) Eksogenitas Lokal

$$\boxed{\mathbb{E}[\varepsilon_i \mid X_i, U_i] = 0}$$

Ini menjamin bahwa **ketidakpastian hanya berasal dari error**, bukan dari desain pembobotan.

(A3) Variansi Terbatas

$$\mathbb{E}[\varepsilon_i^2 \mid X_i, U_i] = \sigma^2 < \infty$$

(A4) Momen Kovariat Terbatas

$$\mathbb{E}\left[|X_i|^2\right] < \infty$$

(A5) Kehalusan Koefisien Spasial

Fungsi $\beta(u)$ memiliki turunan kedua kontinu di sekitar u_0. Ini memungkinkan **ekspansi Taylor lokal**.

📌 **Perhatikan**: Belum ada satu pun asumsi tentang **kernel atau GNN**. Itu **sengaja**.

---

3. Estimator: Kernel sebagai Fungsi yang Diestimasi oleh GNN

Di sinilah perbedaan fundamental dengan GWR klasik dimulai.

3.1 Masalah yang Harus Diselesaikan

Dalam GWR klasik, bobot berbentuk:

$$w_i(u_0) = \frac{K\left(\frac{U_i - u_0}{h}\right)}{\sum_j K\left(\frac{U_j - u_0}{h}\right)}$$

Namun di sini:

- kita **tidak ingin menetapkan $K(\cdot)$ secara eksplisit**,
- kita ingin **mempelajari bentuk kernel dari data**.

Maka kita perlukan:

> suatu **fungsi pembobotan adaptif** yang **menggantikan kernel**, tetapi **tidak merusak struktur GWR**.

3.2 Objek yang Ingin Kita Estimasi (kernel adaptif)

Kita ingin bobot berbentuk:

$$w_i(u_0) = \frac{\exp\left(\ s(U_i, X_i, u_0)\ \right) \cdot \mathbf{1}\{|U_i - u_0| \le h\}}{\sum_{j:|U_j - u_0| \le h} \exp\left(\ s\left(U_j, X_j, u_0\right)\ \right)}$$

Penjelasan:

- $s(\cdot)$ adalah **fungsi skor**, belum ditentukan bentuknya
- indikator memastikan **lokalitas keras**
- softmax memastikan:
  - bobot positif
  - jumlah bobot = 1

3.3 Mengapa Fungsi Skor $s(\cdot)$ Harus Dipelajari?

Karena:

- hubungan spasial bisa **anisotropik**,
- jarak Euclidean saja tidak cukup,
- informasi graf (keterhubungan wilayah) penting.

Maka:

$$s(\cdot) \quad \text{harus mampu memproses struktur graf}$$

Di sinilah **GNN menjadi pilihan natural**.

3.4 Definisi Formal: GNN sebagai Estimator Kernel
Kita definisikan sebuah fungsi parametrik:

$$s_\theta(i, u_0) = \mathrm{GNN}_\theta \left(\mathcal{G}_h(u_0), \mathbf{f}_i(u_0) \right)$$

dengan:

- $\mathcal{G}_h(u_0)$: graf lokal dengan node $i \in \mathcal{N}_h(u_0)$
- $\mathbf{f}_i(u_0)$: fitur node, **minimal** berisi

$$\mathbf{f}_i(u_0) = \left(X_i, U_i - u_0 \right)$$

📌 **Belum ada embedding atau layer disebutkan.** Kita hanya butuh: GNN = fungsi kontinu, terbatas, dan invariant terhadap permutasi node.

3.5 Bobot Akhir (kernel hasil estimasi GNN)
Bobot lokal didefinisikan sebagai:

$$\boxed{w_i(u_0) = \frac{\exp\left(s_\theta(i, u_0) \right)}{\sum_{j \in \mathcal{N}_h(u_0)} \exp\left(s_\theta(j, u_0) \right)}}$$

Inilah **kernel adaptif yang diestimasi dari data**.

3.6 Estimator GWR dengan Kernel Hasil GNN
Definisikan matriks bobot:

$$W_\theta(u_0) = \mathrm{diag}\left(w_1(u_0), \ldots, w_n(u_0) \right)$$

Estimator koefisien lokal:

$$\boxed{\hat{\beta}(u_0) = \left(X^\top W_\theta(u_0) X \right)^{-1} X^\top W_\theta(u_0) Y}$$

📌 **Inilah estimator yang akan kita analisis.**

---

4. Penurunan Konsistensi Koefisien Lokal

(dengan kernel yang diestimasi oleh GNN)

4.1 Tujuan Formal
Yang ingin kita buktikan adalah:

$$\boxed{\hat{\beta}(u_0) \xrightarrow{p} \beta(u_0)}$$

artinya: estimator koefisien lokal berbasis GWR dengan **kernel hasil GNN** adalah **konsisten** untuk koefisien lokal sebenarnya.

4.2 Mulai dari Definisi Estimator
Estimator didefinisikan sebagai:

$$\hat{\beta}(u_0) = \left(X^\top W_\theta(u_0) X \right)^{-1} X^\top W_\theta(u_0) Y$$

dengan:

- $W_\theta(u_0)$ matriks diagonal berisi bobot
- bobot:

$$w_i(u_0) = \frac{\exp\left(s_\theta(i, u_0) \right)}{\sum_{j \in \mathcal{N}_h(u_0)} \exp\left(s_\theta(j, u_0) \right)}$$

4.3 Substitusi Model Struktural

Ingat model sebenarnya:

$$Y = X\beta(U) + \varepsilon$$

Substitusi ke estimator:

$$\hat{\beta}(u_0) = \left(X^\top W X\right)^{-1} X^\top W (X\beta(U) + \varepsilon)$$

Pisahkan:

$$= \left(X^\top W X\right)^{-1} X^\top W X \beta(U) + \left(X^\top W X\right)^{-1} X^\top W \varepsilon$$

Sampai sini **belum ada probabilitas**. Ini **identitas aljabar murni**.

4.4 Trik Fundamental: Tambah–Kurang $\beta(u_0)$

Tuliskan untuk setiap observasi:

$$\beta(U_i) = \beta(u_0) + \left(\beta(U_i) - \beta(u_0) \right)$$

Maka:

$$X^\top W X \beta(U) = X^\top W X \beta(u_0) + X^\top W X \left(\beta(U) - \beta(u_0) \right)$$

Substitusi kembali:

$$\hat{\beta}(u_0) = \beta(u_0) + \underbrace{\left(X^\top W X\right)^{-1} X^\top W X \left(\beta(U) - \beta(u_0) \right)}_{\text{Bias term}} + \underbrace{\left(X^\top W X\right)^{-1} X^\top W \varepsilon}_{\text{Noise term}}$$

📌 **Inilah dekomposisi utama.** Semua analisis konsistensi bergantung pada dua suku ini.

4.5 Strategi Pembuktian Konsistensi

Untuk menunjukkan:

$$\hat{\beta}(u_0) \xrightarrow{p} \beta(u_0)$$

cukup menunjukkan bahwa:

1. **Bias term**

$$(X^\top W X)^{-1} X^\top W X(\beta(U) - \beta(u_0)) \xrightarrow{p} 0$$

2. **Noise term**

$$(X^\top W X)^{-1} X^\top W \varepsilon \xrightarrow{p} 0$$

4.6 Analisis Bias Term (bagian deterministik)

4.6.1 Gunakan Kehalusan $\beta(\cdot)$

Dari Asumsi (A5), $\beta(\cdot)$ dua kali terdiferensiasi. Maka, untuk U_i dekat u_0:

$$\beta(U_i) = \beta(u_0) + \nabla\beta(u_0)^\top (U_i - u_0) + R_i$$

dengan remainder:

$$|R_i| \le C|U_i - u_0|^2$$

4.6.2 Substitusi ke Bias Term

Bias term menjadi:

$$(X^\top W X)^{-1} \sum_{i=1}^{n} w_i X_i X_i^\top \left[\nabla\beta(u_0)^\top (U_i - u_0) + R_i \right]$$

Pisahkan dua bagian: 1. **Suku linear** 2. **Suku remainder kuadrat**

4.6.3 Kenapa Suku Linear Hilang?

Kita ingin:

$$\sum_i w_i (U_i - u_0) \approx 0$$

Agar ini benar, **kita mentok** dan harus membuat **asumsi/desain**.

🔴 **TITIK KEBUNTUAN PERTAMA**

Jika bobot **tidak simetris** terhadap u_0, maka:

$$\sum_i w_i (U_i - u_0) \neq 0$$

➡️ Bias menjadi orde h, konsistensi tetap bisa, tapi sangat buruk.

Maka kita BUTUH:
Asumsi (K1) — Local symmetry (implisit)

Secara formal:

$$\sum_{i \in \mathcal{N}_h(u_0)} w_i (U_i - u_0) \xrightarrow{p} 0$$

Bagaimana ini DICAPAI dengan GNN?

➡️ **Desain, bukan asumsi kosong**:

- input GNN memakai $(U_i - u_0)$ (koordinat relatif)
- graf dibangun simetris di sekitar u_0
- output di-softmax $\longrightarrow$ bobot ter-normalisasi

Ini membuat **ekspektasi bobot simetris** meskipun bentuk kernel fleksibel.

📌 **Catatan penting**: Kita TIDAK mengasumsikan GNN simetris sempurna, hanya bahwa **dalam limit**, bobot tidak bias arah.

4.6.4 Remainder Term

Karena: * $|U_i - u_0| \leq h$, * $|R_i| \leq Ch^2$,

maka:

$$\left| (X^\top W X)^{-1} \sum_i w_i X_i X_i^\top R_i \right| = O_p(h^2)$$

Dan karena $h \to 0$:

$$O_p(h^2) \to 0$$

4.6.5 Kesimpulan Bias Term

$$\boxed{\text{Bias term} \xrightarrow{p} 0}$$

4.7 Analisis Noise Term (bagian stokastik)

Sekarang bagian **paling berbahaya**:

$$(X^\top W X)^{-1} X^\top W \varepsilon$$

4.7.1 Masalah Utama

Bobot W **dihasilkan oleh GNN dari data**. Maka secara umum:

$$\mathbb{E}[X^\top W \varepsilon] \neq 0$$

➡️ LLN gagal, konsistensi gagal.

🔴 **TITIK KEBUNTUAN KEDUA**

Tanpa intervensi, estimator **TIDAK konsisten**.

4.7.2 Solusi: Cross-fitting (desain, bukan asumsi)

Konstruksi

Bagi data:

$$\{1, \ldots, n\} = I_1 \cup I_2$$

- Gunakan I_1 untuk **melatih GNN**

- Gunakan I_2 untuk **menghitung $\hat{\beta}(u_0)$**

Bobot menjadi:

$$W_\theta^{(1)}(u_0) \quad \text{independen dari} \quad \varepsilon_i, i \in I_2$$

Konsekuensi kunci

Kondisional pada $W^{(1)}$:

$$\mathbb{E}\left[\, X_i w_i \varepsilon_i \mid W^{(1)} \,\right] = 0$$

➡️ **LLN berlaku kembali**

4.7.3 Konvergensi Noise Term

Karena: * w_i bounded dan ter-normalisasi, * $X_i \varepsilon_i$ punya mean nol dan variansi terbatas,

maka:

$$X^\top W \varepsilon = O_p\left(\sqrt{nh^d}\right)$$

dan:

$$(X^\top W X)^{-1} = O_p\left(\left(nh^d\right)^{-1}\right)$$

Sehingga:

$$(X^\top W X)^{-1} X^\top W \varepsilon = O_p\left(\left(nh^d\right)^{-1/2}\right) \xrightarrow{p} 0$$

karena $nh^d \to \infty$.

4.8 Kesimpulan Konsistensi

Gabungkan: * bias term $\longrightarrow 0$ * noise term $\longrightarrow 0$

maka:

$$\boxed{\hat{\beta}(u_0) \xrightarrow{p} \beta(u_0)}$$

4.9 Ringkasan Logika Konsistensi

> Konsistensi koefisien lokal tercapai karena neighborhood menyempit, fungsi koefisien halus, bobot GNN bersifat lokal dan simetris secara limit, serta cross-fitting memulihkan eksogenitas antara bobot dan error.

---

5. Penurunan Variansi Koefisien Lokal

(GWR dengan kernel hasil GNN, cross-fitted)

5.1 Tujuan Formal

Kita ingin menghitung (atau membatasi orde dari):

$$\mathrm{Var}\left(\ \hat{\beta}(u_0)\ \right)$$

lebih tepatnya, **orde asimtotiknya** sebagai fungsi dari: * ukuran sampel n, * bandwidth h, * dimensi spasial d.

5.2 Ingat Kembali Dekomposisi

Dari Part 4, kita punya:

$$\hat{\beta}(u_0) = \beta(u_0) + \underbrace{B_n(u_0)}_{\text{bias}} + \underbrace{\left(X^\top W X\right)^{-1} X^\top W \varepsilon}_{\text{noise}}$$

Variansi hanya datang dari **noise term** karena: * $\beta(u_0)$ deterministik, * $B_n(u_0)$ deterministik bersyarat pada U (dan kecil).

Jadi:

$$\mathrm{Var}\left(\ \hat{\beta}(u_0)\ \right) = \mathrm{Var}\left(\ \left(X^\top W X\right)^{-1} X^\top W \varepsilon\ \right)$$

5.3 Notasi

- $W = W_\theta(u_0)$: matriks bobot diagonal dengan elemen $w_i = w_i(u_0)$
- X: matriks desain $n \times p$
- $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)^\top$
- Kita bekerja **kondisional pada bobot W** (ini sah karena cross-fitting)

5.4 Variansi Bersyarat

Karena cross-fitting membuat W **independen dari** ε pada fold estimasi, kita boleh menulis:

$$\mathrm{Var}\left(\ \hat{\beta}(u_0)\ \right) = \mathbb{E}\left[\mathrm{Var}\left(\ \left(X^\top W X\right)^{-1} X^\top W \varepsilon \middle| W, X\ \right)\right]$$

5.5 Hitung Variansi Kondisional (aljabar matriks)

Gunakan fakta umum: Jika A matriks deterministik dan $\mathrm{Var}(\varepsilon \mid X, U) = \sigma^2 I_n$, maka:

$$\mathrm{Var}(A\varepsilon) = \sigma^2 A A^\top$$

Di sini:

$$A = \left(X^\top W X\right)^{-1} X^\top W$$

Maka:

$$\mathrm{Var}\left(\ \left(X^\top W X\right)^{-1} X^\top W \varepsilon \middle| W, X\ \right) = \sigma^2 \left(X^\top W X\right)^{-1} X^\top W^2 X \left(X^\top W X\right)^{-1}$$

📌 **Ini ekspresi variansi eksak** (belum asimtotik).

5.6 Orde Asimtotik

5.6.1 Orde dari $X^\top WX$

Perhatikan:

$$X^\top WX = \sum_{i=1}^{n} w_i X_i X_i^\top$$

Karena: * bobot w_i hanya non-nol di neighborhood $\mathcal{N}_h(u_0)$, * ukuran neighborhood $\#\mathcal{N}_h(u_0) \asymp nh^d$, * dan bobot **ter-normalisasi** ($\sum w_i = 1$),

maka secara orde:

$$X^\top WX = O_p\left(nh^d\right)$$

Lebih tepatnya:

$$\frac{1}{nh^d} X^\top WX \xrightarrow{p} Q(u_0)$$

dengan $Q(u_0)$ matriks positif definit.

Akibatnya:

$$\left(X^\top WX\right)^{-1} = O_p\ \left(\ \left(nh^d\right)^{-1}\ \right)$$

5.6.2 Orde dari $X^\top W^2 X$

Sekarang:

$$X^\top W^2 X = \sum_{i=1}^{n} w_i^2 X_i X_i^\top$$

Karena: * $w_i = O\left(\left(nh^d\right)^{-1}\right)$ secara tipikal (bobot dibagi rata), * hanya nh^d observasi yang relevan,

maka:

$$X^\top W^2 X = O_p\ \left(\ \left(nh^d\right)^{-1}\ \right)$$

5.6.3 Gabungkan Semuanya

Substitusi orde ke ekspresi variansi:

$$\operatorname{Var}\ \left(\ \hat{\beta}(u_0)\ \right) = \sigma^2 \underbrace{O_p\left(\left(nh^d\right)^{-1}\right)}_{(X^\top WX)^{-1}} \cdot \underbrace{O_p\left(\left(nh^d\right)^{-1}\right)}_{X^\top W^2 X} \cdot \underbrace{O_p\left(\left(nh^d\right)^{-1}\right)}_{(X^\top WX)^{-1}}$$

Sehingga:

$$\boxed{\mathrm{Var}\left(\hat{\beta}(u_0)\right) = O\left(\frac{1}{nh^d}\right)}$$

5.7 Bentuk Limit Variansi (lebih presisi)
Dengan normalisasi yang tepat, kita bisa tulis:

$$\mathrm{Var}\left(\hat{\beta}(u_0)\right) \approx \frac{\sigma^2}{nh^d} Q(u_0)^{-1}\Omega(u_0)Q(u_0)^{-1}$$

dengan: * $Q(u_0) = \mathbb{E}[X_i X_i^\top \mid U_i \approx u_0]$ * $\Omega(u_0) = \mathbb{E}[w_i^2 X_i X_i^\top \mid U_i \approx u_0]$

📌 **Bentuk "sandwich" ini akan muncul lagi di CLT.**

5.8 Interpretasi

5.8.1 Mengapa Muncul nh^d?
Karena: * hanya observasi di radius h yang dipakai, * jumlahnya $\asymp nh^d$.

Ini adalah **effective local sample size**.

5.8.2 Trade-off Inti GWR
- h kecil $\longrightarrow$ bias kecil, tapi variansi besar
- h besar $\longrightarrow$ bias besar, tapi variansi kecil

📌 Variansi memberi **harga statistik** dari pelokalan.

5.9 Ringkasan

$$\boxed{\mathrm{Var}\left(\hat{\beta}(u_0)\right) = O\left(\frac{1}{nh^d}\right)}$$

Ini menyiratkan bahwa **skala fluktuasi alami** dari estimator adalah:

$$\sqrt{nh^d}\left(\hat{\beta}(u_0) - \beta(u_0)\right)$$

➡️ **Ini tepat skala yang akan kita pakai di CLT.**

---

6. Penurunan Distribusi Koefisien Lokal

(GWR dengan kernel hasil GNN, cross-fitted)

6.1 Tujuan Formal
Kita ingin mempelajari limit distribusi dari estimator $\hat{\beta}(u_0)$. Secara khusus, kita ingin mencari **skala normalisasi** a_n sedemikian sehingga:

$$a_n\left(\hat{\beta}(u_0) - \beta(u_0)\right) \xrightarrow{d} \text{distribusi non-degenerate}$$

Dari Part 5, kita sudah tahu:

$$\operatorname{Var}\left(\hat{\beta}(u_0)\right) = O\left(\frac{1}{nh^d}\right)$$

➡️ **Ini langsung memberi kandidat normalisasi**:

$$\boxed{a_n = \sqrt{nh^d}}$$

6.2 Kembali ke Dekomposisi Lengkap
Dari Part 4, kita punya:

$$\hat{\beta}(u_0) = \beta(u_0) + B_n(u_0) + V_n(u_0)$$

dengan:

- **Bias term**

$$B_n(u_0) = \left(X^\top W X\right)^{-1} X^\top W X \left(\beta(U) - \beta(u_0) \right)$$

- **Noise term**

$$V_n(u_0) = \left(X^\top W X\right)^{-1} X^\top W \varepsilon$$

Kalikan seluruh persamaan dengan $\sqrt{nh^d}$:

$$\sqrt{nh^d} \left(\hat{\beta}(u_0) - \beta(u_0) \right) = \underbrace{\sqrt{nh^d} B_n(u_0)}_{\text{bias terstandar}} + \underbrace{\sqrt{nh^d} V_n(u_0)}_{\text{noise terstandar}}$$

Distribusi limit **ditentukan oleh dua suku ini**.

6.3 Distribusi Noise Term (CLT inti)

6.3.1 Tulis Noise Term sebagai Jumlah Eksplisit
Ingat:

$$V_n(u_0) = \left(X^\top W X\right)^{-1} \sum_{i=1}^{n} w_i X_i \varepsilon_i$$

Kalikan dengan $\sqrt{nh^d}$:

$$\sqrt{nh^d} V_n(u_0) = \left(\frac{1}{nh^d} X^\top W X \right)^{-1} \left(\frac{1}{\sqrt{nh^d}} \sum_{i=1}^{n} w_i X_i \varepsilon_i \right)$$

Perhatikan bahwa ini adalah **produk dua faktor**: 1. satu yang akan konvergen **dalam probabilitas**, 2. satu yang akan konvergen **dalam distribusi**.

6.3.2 Limit Faktor Pertama (deterministik secara limit)
Dari Part 5:

$$\frac{1}{nh^d} X^\top W X \xrightarrow{p} Q(u_0)$$

dengan $Q(u_0)$ matriks positif definit.

Maka:

$$\left(\frac{1}{nh^d} X^\top W X \right)^{-1} \xrightarrow{p} Q(u_0)^{-1}$$

6.3.3 Limit Faktor Kedua (CLT sesungguhnya)
Sekarang perhatikan:

$$S_n = \frac{1}{\sqrt{nh^d}} \sum_{i=1}^{n} w_i X_i \varepsilon_i$$

Inilah objek yang akan kita CLT-kan.

Mengapa ini kandidat CLT?

Karena: 1. Berupa **jumlah dari banyak suku kecil** 2. Tiap suku punya **mean nol** 3. Variansi totalnya **tidak nol dan terbatas**

6.3.4 Mean Nol (di sinilah cross-fitting KRUSIAL)
Karena cross-fitting: * bobot w_i ditentukan dari fold lain, * sehingga w_i **independen dari ε_i** pada fold estimasi.

Maka:

$$\mathbb{E}[w_i X_i \varepsilon_i \mid X_i, U_i, W] = 0$$

dan:

$$\mathbb{E}[S_n \mid W] = 0$$

6.3.5 Variansi Terbatas dan Stabil
Variansi kondisional:

$$\mathrm{Var}(S_n \mid W) = \frac{1}{nh^d} \sum_{i=1}^{n} w_i^2 \mathbb{E}[X_i X_i^\top \varepsilon_i^2 \mid U_i \approx u_0]$$

Dari Part 5, jumlah ini konvergen ke:

$$\Omega(u_0) = \mathbb{E}[w_i^2 X_i X_i^\top \mid U_i \approx u_0]$$

yang **positif definit** dan terbatas.

6.3.6 Aplikasi CLT
Karena: * suku-suku independen, * mean nol, * variansi terbatas,

maka (CLT Lindeberg–Feller):

$$\boxed{S_n \xrightarrow{d} \mathcal{N}\left(0, \Omega(u_0)\right)}$$

6.3.7 Gabungkan Dua Faktor (Slutsky)

Karena satu faktor konvergen dalam probabilitas dan satu dalam distribusi:

$$\sqrt{nh^d}V_n(u_0) \xrightarrow{d} \mathcal{N}\left(0, Q(u_0)^{-1}\Omega(u_0)Q(u_0)^{-1}\right)$$

6.4 Peran Bias Term dalam Distribusi

Sekarang kita analisis:

$$\sqrt{nh^d}B_n(u_0)$$

Dari Part 4, kita tahu:

$$B_n(u_0) = O(h^2)$$

Maka:

$$\sqrt{nh^d}B_n(u_0) = O\left(\sqrt{nh^d}h^2\right)$$

6.4.1 Dua Kemungkinan Fundamental

(i) Bias tidak hilang

Jika:

$$\sqrt{nh^d}h^2 \to c \neq 0$$

maka:

$$\sqrt{nh^d}\left(\hat{\beta}(u_0) - \beta(u_0)\right) \xrightarrow{d} \mathcal{N}\left(\mu(u_0), Q^{-1}\Omega Q^{-1}\right)$$

dengan mean **tidak nol**: $\mu(u_0) \neq 0$

➡️ **Distribusi normal, tapi tidak terpusat** ➡️ Inferensi Wald standar **tidak valid**

Ini terjadi, misalnya, jika: * bandwidth dipilih optimal untuk prediksi (CV), * tidak ada undersmoothing.

(ii) Bias dihilangkan (kasus inferensi)

Jika kita memaksakan:

$$\boxed{\sqrt{nh^d}h^2 \to 0}$$

yang ekuivalen dengan **undersmoothing**, maka:

$$\sqrt{nh^d}B_n(u_0) \xrightarrow{p} 0$$

6.5 Distribusi Asimtotik Akhir (kasus inferensi sah)

Di bawah kondisi undersmoothing:

$$\boxed{\sqrt{nh^d}\ \left(\ \hat{\beta}(u_0) - \beta(u_0)\ \right) \xrightarrow{d} \mathcal{N}\left(0, Q(u_0)^{-1}\Omega(u_0)Q(u_0)^{-1}\right)}$$

Inilah **jawaban akhir untuk pertanyaan distribusi**.

6.6 Interpretasi

1. **Distribusi normal muncul karena CLT lokal**, bukan karena error normal
2. Skala $\sqrt{nh^d}$ = akar dari **effective local sample size**
3. GNN **tidak mengubah bentuk limit**, hanya memengaruhi:
   - matriks $Q(u_0)$,
   - matriks $\Omega(u_0)$

6.7 Ringkasan Part 6

> Setelah dinormalisasi oleh $\sqrt{nh^d}$, koefisien lokal GWR dengan kernel yang diestimasi oleh GNN berdistribusi normal asimtotik, asalkan bias dilenyapkan melalui undersmoothing dan bobot dipelajari secara cross-fitted.

---

7. Ringkasan Asumsi dan Desain Model yang Diperlukan

Bagian ini **sangat penting secara akademis**, karena menunjukkan bahwa asumsi dan desain GNN **bukan selera, tapi konsekuensi logis dari inferensi**.

7.1 Asumsi Data dan Model (tidak bergantung pada GNN)

(A1) Independensi

$$\{(Y_i, X_i, U_i)\}_{i=1}^{n} \quad \text{i.i.d.}$$

➡️ Dipakai untuk: LLN, CLT lokal.

(A2) Eksogenitas Lokal

$$\mathbb{E}[\varepsilon_i \mid X_i, U_i] = 0$$

➡️ Dipakai untuk: mean nol noise term, validitas inferensi.

(A3) Variansi Error Terbatas

$$\mathbb{E}[\varepsilon_i^2 \mid X_i, U_i] = \sigma^2 < \infty$$

➡️ Dipakai untuk: variansi terbatas, CLT.

(A4) Momen Kovariat Terbatas

$$\mathbb{E}\left[|X_i|^2\right] < \infty$$

➡️ Dipakai untuk: kontrol $X^{\top}WX$, inversibilitas matriks desain lokal.

(A5) Kehalusan Fungsi Koefisien

$$\beta(\cdot) \in C^2 \quad \text{di sekitar} \quad u_0$$

➡️ Dipakai untuk: Taylor expansion, analisis bias $O(h^2)$.

7.2 Asumsi Lokalitas (dipaksa oleh konsistensi)

(L1) Shrinking Neighborhood

$$h \to 0$$

➡️ Dipakai untuk: memastikan $\beta(U_i) \approx \beta(u_0)$, menghilangkan bias.

(L2) Local Sample Size Diverges

$$nh^d \to \infty$$

➡️ Dipakai untuk: LLN lokal, variansi $\to 0$.

📌 **Ini menjelaskan mengapa h TIDAK bisa dihilangkan**, meskipun pakai GNN.

7.3 Asumsi Kernel (yang dipenuhi oleh desain GNN)

(K1) Non-negativity dan Normalisasi

$$w_i(u_0) \geq 0, \qquad \sum_i w_i(u_0) = 1$$

➡️ Dipakai untuk: stabilitas estimator, interpretasi sebagai rata-rata lokal.

Dipenuhi oleh: softmax pada output GNN.

(K2) Boundedness

$$\sup_i w_i(u_0) \leq C\left(nh^d\right)^{-1}$$

➡️ Dipakai untuk: mencegah satu observasi mendominasi, CLT Lindeberg.

Dipenuhi oleh: softmax + neighborhood terbatas.

(K3) Regularity (Kontinuitas)

$$|w_i(u_0; Z) - w_i(u_0; Z')| \leq L|Z - Z'|$$

➡️ Dipakai untuk: uniform LLN, kontrol remainder.

Dipenuhi oleh: arsitektur GNN kontinu (aktivasi halus, tanpa threshold keras).

(K4) Simetri Lokal (implisit)

$$\sum_i w_i(u_0)(U_i - u_0) \xrightarrow{p} 0$$

➡️ Dipakai untuk: menghilangkan bias orde pertama.

Dipenuhi oleh desain: * input relatif $(U_i - u_0)$, * graf lokal simetris, * normalisasi bobot.

7.4 Asumsi Estimasi Bobot (khusus GNN)

(G1) Cross-fitting
Bobot w_i dipelajari dari data yang **tidak dipakai** untuk estimasi koefisien.

➡️ Dipakai untuk: memulihkan eksogenitas, memastikan $\mathbb{E}[w_i \varepsilon_i] = 0$.

Tanpa ini: konsistensi **gagal total**.

(G2) Bounded Complexity
Parameter GNN θ berada dalam himpunan kompak.

➡️ Dipakai untuk: stabilitas limit, mencegah overfitting ekstrem.

7.5 Asumsi Inferensi (untuk CLT terpusat)

(I1) Undersmoothing

$$\sqrt{nh^d}h^2 \to 0$$

➡️ Dipakai untuk: menghilangkan bias dalam limit, distribusi normal terpusat.

Tanpa ini: distribusi tetap normal, tapi **mean ≠ 0** (inferensi Wald gagal).

---

8. Rangkuman Akhir: Model Formal, Asumsi, dan Desain End-to-End

8.1 Model Formal

Data:

$$(Y_i, X_i, U_i), \quad i = 1, \ldots, n$$

Model:

$$Y_i = X_i^\top \beta(U_i) + \varepsilon_i$$

Target:

$$\beta(u_0)$$

8.2 Desain Estimator (end-to-end)

Step 1 — Neighborhood

$$\mathcal{N}_h(u_0) = \{i : |U_i - u_0| \leq h\}$$

Step 2 — Graph Lokal

Node: $i \in \mathcal{N}_h(u_0)$

Fitur node: $(X_i, U_i - u_0)$

Step 3 — GNN sebagai Estimator Kernel

GNN menghasilkan skor: $s_\theta(i, u_0)$

Bobot:

$$w_i(u_0) = \frac{\exp(s_\theta(i, u_0))}{\sum_{j \in \mathcal{N}_h(u_0)} \exp(s_\theta(j, u_0))}$$

Step 4 — Cross-fitting

- GNN dilatih di fold A
- Koefisien dihitung di fold B
- Dirata-ratakan antar fold

Step 5 — Estimasi Koefisien Lokal

$$\hat{\beta}(u_0) = \left(X^\top W_\theta X\right)^{-1} X^\top W_\theta Y$$

8.3 Jawaban Eksplisit atas Dua Pertanyaan Utama

❓ **Apakah koefisien lokal konsisten?**
✓ **YA**, jika: * neighborhood menyempit, * fungsi koefisien halus, * bobot GNN lokal, bounded, dan cross-fitted.

$$\hat{\beta}(u_0) \xrightarrow{p} \beta(u_0)$$

❓ **Apa distribusi koefisien lokal?**
✓ **Normal asimtotik**, jika additionally: * undersmoothing dilakukan.

$$\sqrt{nh^d}\left(\hat{\beta}(u_0) - \beta(u_0)\right) \xrightarrow{d} \mathcal{N}\left(0, Q(u_0)^{-1}\Omega(u_0)Q(u_0)^{-1}\right)$$

Jika tidak: * distribusi tetap normal, * tetapi **mean tidak nol**.

8.4 Ringkasan Satu Paragraf (siap pakai di tesis)

> *This study establishes the consistency and asymptotic normality of local coefficients in geographically weighted regression with data-adaptive kernels learned via graph neural networks. By preserving shrinking neighborhoods, enforcing kernel regularity through architectural constraints, and employing cross-fitting to restore exogeneity, the proposed estimator achieves valid local inference under standard smoothness and moment conditions.*

---

9. Kriteria Arsitektur GNN yang Valid untuk Inferensi

9.1 Jawaban Singkat

> ❌ **TIDAK semua backbone GNN bisa digunakan** ✓ **BANYAK backbone populer *harus dimodifikasi*** ✓ Yang penting **BUKAN namanya**, tapi **properti matematisnya**

GNN **boleh apa saja** *selama* ia **berperilaku seperti kernel yang sah** (bounded, smooth, lokal, eksogen secara asimtotik).

9.2 Kriteria Matematis yang HARUS Dipenuhi GNN

(C1) Output harus menghasilkan bobot kernel, bukan prediksi
Syarat: GNN **tidak boleh** memprediksi Y. Ia harus menghasilkan **skor bobot**:

$$s_\theta(i, u_0) \quad \Rightarrow \quad w_i(u_0) = \text{softmax}\ (s_\theta)$$

Kenapa? Kalau GNN memprediksi Y: * bobot jadi fungsi langsung dari noise, * $\mathbb{E}[w_i \varepsilon_i] \neq 0$, * **konsistensi gagal total**.

Implikasi desain: * ✓ GNN = *kernel learner* * ❌ GNN = *regressor*

(C2) Boundedness & Normalization
Syarat:

$$0 \leq w_i \leq 1, \qquad \sum_i w_i = 1$$

Kenapa? * Variansi terbatas * CLT berlaku * Tidak ada observasi dominan

Implikasi desain: * ✓ **Softmax WAJIB** * ❌ ReLU / linear output langsung

(C3) Kontinuitas / Smoothness terhadap Input
Syarat: Bobot harus berubah **halus** saat U_i atau X_i berubah kecil:

$$|w_i(Z) - w_i(Z')| \leq L|Z - Z'|$$

Kenapa? * Dipakai di Uniform LLN * Dipakai di kontrol remainder bias * Tanpa ini, bukti runtuh diam-diam

Implikasi desain: * ✓ Aktivasi smooth (tanh, GELU, softplus) * ❌ Hard attention, threshold, top-k selection

(C4) Locality TIDAK boleh dipelajari oleh GNN
Syarat: Neighborhood ditentukan **sebelum** GNN:

$$\mathbf{1}\{|U_i - u_0| \leq h\}$$

Kenapa? * Ini yang menjamin $h \to 0$ * Ini yang mendefinisikan $\beta(u_0)$

Implikasi desain: * ✓ GNN hanya bekerja *di dalam* neighborhood * ❌ GNN memilih node global

(C5) Permutation Invariance
Syarat: Urutan node **tidak memengaruhi output**.

Kenapa? * Bobot kernel tidak boleh tergantung indexing * Ini asumsi implisit semua GWR

Implikasi desain: * ✓ Message passing / aggregation * ❌ Sequence-based encoder (tanpa simetrisasi)

(C6) Eksogenitas via Cross-fitting
Syarat: Bobot dihitung dari data yang **tidak digunakan** untuk estimasi koefisien.

Kenapa? Tanpa ini: $\mathbb{E}[w_i \varepsilon_i] \neq 0$

Implikasi desain: * ✓ Train GNN di fold A, apply di fold B * ❌ End-to-end training tanpa sample splitting

9.3 Backbone yang Aman vs Berbahaya

✅ Backbone yang *AMAN secara default* (dengan desain tepat)
Message Passing GNN (GCN-like, GraphSAGE-like)

Karena: * aggregation simetris, * output kontinu, * mudah dikontrol.

✓ Tambahkan softmax $\rightarrow$ kernel sah

Attention-based GNN (GAT-style)

✓ **BOLEH**, *dengan syarat*: * attention score $\rightarrow$ softmax lokal, * **tidak hard masking**, * hanya di neighborhood h.

⚠️ Backbone yang *BERBAHAYA tanpa modifikasi*
Hard attention / Top-k pooling

Masalah: * tidak kontinu, * melanggar (C3), * bias sulit dikontrol.

Deep GNN sangat dalam

Masalah: * Lipschitz constant bisa meledak, * ULLN tidak stabil.

Solusi: * shallow GNN (1–3 layer), * weight decay / spectral norm.

Global Transformer Graph

Masalah: * tidak lokal, * kernel tidak shrink, * target bukan $\beta(u_0)$ lagi.

9.4 Arsitektur yang "Legal" dari Sudut Pandang Teori
Input

Untuk setiap lokasi target u_0: * Node features: $\mathbf{f}_i = (X_i, U_i - u_0)$ * Graph: $\mathcal{G}_h(u_0)$

GNN Body (contoh aman)

- 1–2 message passing layer
- Aktivasi smooth
- Tidak ada pooling keras

Output: $e_i = \mathrm{GNN}_\theta(\mathbf{f}_i)$ (ini embedding skalar atau vektor kecil)

Kernel Head (WAJIB eksplisit)

Mapping embedding ⟶ skor: $s_i = a^\top e_i$

Bobot:

$$w_i(u_0) = \frac{\exp(s_i)}{\sum_{j \in \mathcal{N}_h(u_0)} \exp(s_j)}$$

📌 **Inilah kernel hasil estimasi GNN**

Training (cross-fitted)

Loss:

$$\sum_{i \in \text{ val}} \left(Y_i - X_i^\top \hat{\beta}^{(-i)}(U_i) \right)^2$$

Bukan loss prediksi langsung GNN.

9.5 Checklist Praktis

Sebelum pakai backbone GNN, tanyakan:

- ❓ Output GNN jadi **bobot kernel** atau **prediksi Y**?
- ❓ Ada **softmax lokal**?
- ❓ Neighborhood ditentukan **sebelum** GNN?
- ❓ Aktivasi **kontinu**?
- ❓ Ada **cross-fitting**?
- ❓ GNN dangkal & stabil?

Kalau satu saja ❌ ⟶ **inferensi tidak sah**.

9.6 Ringkasan

> *Not all GNN backbones are suitable for inferential GWR. Only architectures that yield bounded, smooth, locally normalized weights and are trained via cross-fitting preserve the kernel regularity conditions required for consistency and asymptotic normality.*

Secara konsep, **GNN di sini bukan "model ML"**, tapi:

> **alat untuk mengestimasi kernel nonparametrik adaptif**.

---

10. Pemilihan Bandwidth h

10.1 Prinsip Fundamental

> **h BUKAN parameter yang harus "optimal" secara numerik. h adalah parameter yang harus "benar secara asimtotik".**

Artinya: * tidak mengejar h terbaik di satu dataset, * mengejar aturan pemilihan h_n yang:

$$h_n \to 0 \quad \text{dan} \quad nh_n^d \to \infty$$

Inferensi **hidup–mati** di sini.

10.2 Ringkasan Rekomendasi

> **Dalam praktik inferensial:**
>
> ✓ Jangan pilih h murni dari CV ✓ Tentukan h dengan **aturan deterministik atau kNN** ✓ Gunakan CV hanya sebagai **kalibrasi kasar**, bukan penentu akhir ✓ Biarkan GNN mengurus adaptivitas, bukan h

10.3 OPSI 1 — Deterministic Bandwidth (PALING AMAN UNTUK TESIS)

Bentuk:

$$\boxed{h_n = C \cdot n^{-\alpha}}$$

dengan: * $C > 0$ konstanta skala, * $\alpha > 0$ laju penyusutan.

Kenapa ini paling aman?

Karena dapat **secara eksplisit** menyatakan: * $h_n \to 0$ (karena $\alpha > 0$), * $nh_n^d \to \infty$ (asal $\alpha < 1/d$).

Untuk **inferensi GWR**, kita butuh **undersmoothing**:

$$\sqrt{nh_n^d} h_n^2 \to 0$$

Untuk data spasial 2D ($d = 2$), pilihan aman:

$$\alpha \in (0.25, 0.4)$$

Bagaimana memilih C?

Ini **bukan masalah teori**, tapi skala data.

Praktik aman: * hitung jarak median antar lokasi, * set C sebagai proporsi jarak itu (mis. 0.5×).

Yang penting: C **tetap**, tidak tergantung n.

10.4 OPSI 2 — k-Nearest Neighbors (SANGAT NATURAL UNTUK GNN)

Alih-alih radius h, dipilih **jumlah tetangga**.

Bentuk:

$$\boxed{k_n = \lceil n^\gamma \rceil, \quad 0 < \gamma < 1}$$

Neighborhood: * ambil k_n tetangga terdekat dari u_0, * $h_n(u_0)$ = jarak ke tetangga ke-k_n.

Kenapa ini sah secara teori?

Karena:

$$h_n(u_0) \asymp \left(\frac{k_n}{n}\right)^{1/d}$$

Jika: * $k_n \to \infty$, * $k_n/n \to 0$,

maka:

$$h_n \to 0 \quad \text{dan} \quad nh_n^d \asymp k_n \to \infty$$

📌 **Semua syarat inferensi terpenuhi.**

Pilihan γ yang aman (praktis)

Untuk $d = 2$: $\gamma \in (0.4, 0.7)$

Contoh: * $n = 2,000 \longrightarrow k \approx 60$ * $n = 10,000 \longrightarrow k \approx 150$

📌 Ini **sangat cocok** dengan graf GNN: * graf lokal = kNN graph, * tidak perlu radius eksplisit.

10.5 OPSI 3 — Cross-validation (HANYA SEBAGAI REFERENSI)

Fakta jujur tentang CV

CV memilih:

$$h_{CV} \asymp n^{-1/(d+4)}$$

Untuk $d = 2$:

$$h_{CV} \asymp n^{-1/6}$$

Masalah:

$$\sqrt{nh_{CV}^d}h_{CV}^2 \to c \neq 0$$

➡️ Bias **tidak hilang** dalam CLT ➡️ Inferensi Wald **tidak sah**

Cara CV masih boleh dipakai

Sebagai **starting point**, bukan keputusan akhir.

Praktik yang jujur: 1. Hitung h_{CV} 2. Lakukan **undersmoothing**:

$$\boxed{h = c \cdot h_{CV}, \quad c \in (0.5, 0.8)}$$

3. Jelaskan di teks: > CV digunakan untuk kalibrasi skala, bukan untuk inferensi.

10.6 Praktik yang TIDAK Boleh Dilakukan

- ❌ Memilih h yang meminimalkan MSE dan berhenti di situ
- ❌ Mengklaim "h kecil" tanpa aturan asimtotik
- ❌ Membiarkan GNN menentukan neighborhood
- ❌ Menggunakan k tetap (mis. 50) tanpa kaitan ke n

10.7 Diagnostic Praktis (sanity check)

1️⃣ Effective local sample size

Pastikan:

$$30 \le nh^d \le 0.2n$$

Kalau: * terlalu kecil ⟶ variance meledak, * terlalu besar ⟶ tidak lokal.

2️⃣ Stabilitas koefisien

Hitung $\hat{\beta}(u_0)$ untuk: h, $0.8h$, $0.6h$

Koefisien **harus stabil**, bukan meloncat.

10.8 Rekomendasi Akhir

Karena konteks: * ingin inferensi, * memakai GNN sebagai kernel adaptif,

maka **pilihan terbaik**:

> ✓ **kNN-based bandwidth dengan $k_n \propto n^\gamma$** ✓ atau deterministic $h_n = Cn^{-\alpha}$ ❌ CV murni

Kalimat tesis yang kuat:

> *"The bandwidth is chosen to guarantee shrinking neighborhoods and diverging local sample size, while adaptivity is achieved through learned neural weights rather than bandwidth optimization."*

10.9 Ringkasan

> **Dalam inferensi GWR–GNN, h dipilih untuk memenuhi teori, bukan untuk memaksimalkan prediksi.**