

# أسلوب بيز لاختيار المتغيرات في الانحدار اللامعلمي

خلود يوسف

د. ظافر حسين رشيد

## الخلاصة

تأتي أهمية البحث في دراسة طرائق متعلقة بتمهيد دوال الانحدار اللامعلمي بهدف إيجاد افضل الطرائق التي تلائم نماذج متنوعة لتوزيع الخطأ العشوائي بحالته الطبيعية. حيث أن أهم أهداف البحث هي بحث ما قدمته البحوث الحديثة في مجال الانحدار اللامعلمي. وأيضاً إيجاد طرائق بديلة أو محورة تكون كفوءة في معالجة حالات الإخفاق في جانب الطرائق المتناولة، فضلاً عن تقليل تعقيد بعض الطرائق خصوصاً ما يتعلق بأساليب بيز.

كما أن إحدى أهم وابرز أهداف البحث المهمة تتمركز في بحث الانحدار اللامعلمي باستخدام أسلوب بيز لاختيار المتغيرات. وتقديم اقتراح أسلوب محور يكون كفوءاً واقل تعقيداً من الأسلوب قبل التحوير. وقد تم صياغة أنموذج محاكاة ولنماذج مختلفة وللتحقق من أداء هذه الطرائق تم استخدام عدة مقاييس.

## *A Bayesian variable Selection Approach to Nonparametric Regression*

### **ABSTRACT**

the importance of study focuses on methods related to smoothing of Nonparametric Regression functions. This is for the purpose of producing the best methods convenient for various models. And for the Distribution Random error, in its Normal cases.

Thus, the most important purpose of the research, is to find what the studies so far, have offered in the field of Nonparametric Regression. Also to find alternative or modified methods; which are reliable for the treatment of conditions of failure regarding the methods in use, as well as to alleviate the some methods, especially those related to Bayesian procedures.

One of the most outstanding aims of the research focuses on the study of Nonparametric Regression using Bayesian variable selection. This suggests a modified technique to be reliable and of less complexity than the A simulation model has been performed for a number of models. To verify the performance of such methods, many criteria have been carried out.

## 1 . المقدمة

أن تقنتي الانحدار المعلمي والانحدار اللامعلمي تمثلان طريقتين مختلفتين في تحليل الانحدار لكن هذا لا يعني أن استخدام طريقة يمنع من استخدام الأخرى، فتقنيات الانحدار اللامعلمي يمكن أن تستخدم لتقييم شرعية الأنموذج المعلمي المفترض وبالعكس، وربما شكل مطابقة منحنى الانحدار والحاصل عليه من تقنيات الانحدار اللامعلمي يقترح الأنموذج المعلمي المناسب ليستخدم في الدراسات المستقبلية.

كما وان إجراء الانحدار اللامعلمي ربما يمثل المرحلة النهائية في تحليل البيانات أو خطوة استكشافية في عمليات النمذجة، ولانموذج الانحدار :

$$y_i = g(x_i) + \epsilon_i \quad i = 1, \dots, n \quad \dots \quad (1)$$

حيث الأخطاء العشوائية  $\epsilon_i$  تكون غير مرتبطة وتتوزع بوسط حسابي صفر وتباين  $\sigma^2$ ، والهدف هو تقدير  $g$  التي يشار إليها بدالة الانحدار أو منحنى الانحدار عند النقاط  $x_1, \dots, x_n$ . المعروف أن تقريب دالة الانحدار اللامعلمي التقريبية إلى دالة الانحدار اللامعلمي الحقيقية هو هدف أساليب التمهيد والتي لاقت الدراسة النظرية لها اهتماماً ودراسة واسعة في السنوات العشر الأخيرة لأسباب عديدة منها أن التفكير المعلمي الصرف المستخدم في تقدير منحنى الانحدار لا يتوافق مع الحاجة إلى المرونة في تحليل البيانات، ومع التطور المنجز في أجهزة الحواسيب من الناحية المادية وكذلك إنجاز ومعالجة البيانات اصبح تطوير العديد من طرائق الانحدار اللامعلمي ممكناً نظرياً.

لقد تم تناول العديد من طرائق الانحدار اللامعلمي من قبل الباحثين وذلك لتقدير منحنى الانحدار، بعض منها حقق الامثلية في جوانب نظرية في حين لم تحقق تلك الامثلية في الجوانب التجريبية، وبعضها الآخر ما زال العمل مستمراً باتجاه تطويرها واثبات كفاءتها أو ما زالت تفتقر للأدوات الاستدلالية لوضع الأساس النظري لها كما أن هناك الكثير من المسائل غير المحلولة والمثيرة للجدل، فضلاً عن الصعوبات التطبيقية الكبيرة لطرائق مهمة وذات كفاءة عالية كطرائق بيز للانحدار اللامعلمي.

### 1-1 هدف البحث

يهدف هذا البحث إلى بحث طرائق متعلقة بتمهيد دوال الانحدار اللامعلمي بهدف إيجاد افضل الطرائق التي تلائم نماذج متنوعة ولتوزيع الخطأ العشوائي بحالته الطبيعية.

كما أن إحدى أهم وابرز أهداف البحث تتمركز في بحث الانحدار اللامعلمي باستخدام أسلوب بيز لاختيار المتغيرات، وتقديم اقتراح أسلوب لبيز يكون كفوءاً ويتخطى بعض صعوبات مقَدري المتوسط والمنوال المتمثلة في الحاجة لبرمجة متقدمة جداً فضلاً عن عقبات الحصول على

تقارب للتوزيع اللاحق والحاجة لعشرات المرات من التكرارات من اجل الحصول على التقارب وأحياناً يتعذر الحصول عليه خاصة بالنسبة لتقدير المنوال.

وقد تم صياغة أنموذج محاكاة ولنماذج مختلفة ومستويات تباين مختلفة، وللتحقق من أداء هذه الطرائق تم استخدام عدة مقاييس فضلاً عن مقارنة تقارب الطرائق من المنحنى الحقيقي من خلال رسوم (Plots) التي تعتمد على نتائج مستخلصة من تجارب المحاكاة.

## الجانب النظري

### أسلوب بيز للانحدار اللامعلمي والانحدار اللامعلمي الحصين 2. المقدمة

إن الاهتمام بنماذج بيز للطرائق اللامعلمية زاد في السنوات الأخيرة وبسرعة هائلة والفضل في ذلك يعود إلى التطور في سرعة وخوارزميات الكمبيوتر مما جعل تطبيق طرائق بيز المعقدة ممكناً عملياً.

وعلى الرغم من وجود العديد [2] من الطرائق التي استخدمت لاختيار المتغيرات في الانحدار اللامعلمي وصولاً للأنموذج النهائي كأسلوب إضافة الخطوات المتسلسلة (Addition) (Stepwise) أو أسلوب الاختيار الأمامي (Forward Selection) لكن لوحظ بأنها تعاني سلبيات كون مجموعة جزئية من مجال النماذج الكاملة تحسب فضلاً عن انه قلة من النماذج تزود بوصف معقول للبيانات، في حين طرائق بيز لها عدة ميزات إذ تستخدم طريقة مونت كارلو لسلسلة ماركوف MCMC والتي تعتمد أسلوب البحث العشوائي للمرور بالنماذج المسموحة كافة وكذلك سطح الانحدار يقدر بواسطة المتوسط اللاحق، والحاصل عليه من أي مجموعة جزئية من المتغيرات الموزونة بالاحتمال اللاحق.

### 2-1 أسلوب بيز لاختيار المتغيرات [2، 4، 6]

إن اختيار المتغيرات في الانحدار الخطي يشكل الأساس لطريقة الانحدار اللامعلمي وبفرض أن  $\Gamma$  عائلة لنماذج الانحدار الخطي التي لها المتغير المعتمد نفسه والى  $\gamma \in \Gamma$  فان الأنموذج

$$\underline{y} = \underline{X}_\gamma \underline{\beta}_\gamma + \underline{\epsilon} \quad \dots(2)$$

إذ أن:  $\underline{y}$ : متجه (n×1) للمتغير المعتمد.

$\underline{\beta}$ : متجه (r×1) للمعالم المجهولة وان  $\underline{\beta}_\gamma$  تتضمن كل عناصر  $\beta_i$  بحيث أن  $\gamma_i = 1$ .

$\underline{\epsilon}$ : متجه (n×1) للأخطاء العشوائية وله متوسط صفر وتباين  $\sigma^2 I_n$ .

$\gamma$ : متجه (r×1) لدليل المتغيرات.

### 2-1-1 حساب الاحتمال اللاحق باستخدام معاين Gibbs

للأنموذج (2) وبفرض  $\gamma$  معطاة وان  $q_\gamma = \sum_{i=1}^r \gamma_i$  والتي تمثل عدد العناصر غير الصفرية إلى  $\underline{\beta}$

فان

$$p(\underline{y}|\underline{\gamma}) \propto \int_{\sigma^2} \left\{ \int_{\underline{\beta}} p(\underline{y}|\underline{\beta}, \sigma^2) p(\underline{\beta}|\sigma^2) d\underline{\beta} \right\} p(\sigma^2) d\sigma^2 \propto (1+c)^{-q(\gamma)^2} S(\gamma)^{-n/2} \dots (3)$$

$$S(\gamma) = \underline{y}'\underline{y} - (c/(1+c)) \times \underline{y}'\underline{X}_\gamma (\underline{X}'_\gamma \underline{X}_\gamma)^{-1} \underline{X}'_\gamma \underline{y} = \underline{y}'\underline{y} - (c/(1+c)) \times \underline{\mu}'_\gamma \Sigma_\gamma^{-1} \underline{\mu}_\gamma$$

بدمج دالة الإمكان مع الكثافات المسبقة فان الاحتمال اللاحق لأي أنموذج  $\gamma$

$$p(\underline{\gamma}|\underline{y}) \propto p(\underline{y}|\underline{\gamma}) p(\underline{\gamma})$$

$$\propto \int p(\underline{y}|\underline{\beta}_\gamma, \sigma^2, \underline{\gamma}) p(\underline{\beta}_\gamma | \sigma^2, \underline{\gamma}) p(\sigma^2 | \underline{\gamma}) d\underline{\beta}_\gamma d\sigma^2 p(\underline{\gamma}) \dots (4)$$

$$\begin{aligned} & \text{likelihood} \quad \text{given priors} \\ & \propto \int (\sigma^2)^{-(n+q(\gamma)+2)/2} (2\pi c)^{-q(\gamma)/2} \left| \underline{X}'_\gamma \underline{X}_\gamma \right|^{1/2} \times \\ & \exp \left\{ \frac{-1}{2\sigma^2} \left[ (\underline{\beta}_\gamma - \underline{\mu}_\gamma)' \Sigma_\gamma^{-1} (\underline{\beta}_\gamma - \underline{\mu}_\gamma) - \underline{\mu}'_\gamma \Sigma_\gamma^{-1} \underline{\mu}_\gamma + \underline{y}'\underline{y} \right] \right\} d\underline{\beta}_\gamma d\sigma^2 p(\underline{\gamma}) \end{aligned}$$

حيث  $\underline{\mu}_\gamma = (\underline{X}'_\gamma \underline{X}_\gamma)^{-1} \underline{X}'_\gamma \underline{y}$  ,  $\Sigma_\gamma = (c / (c+1)) (\underline{X}'_\gamma \underline{X}_\gamma)^{-1}$  وبتكامل كثافة  $N(\underline{\mu}_\gamma, \Sigma_\gamma)$  إلى  $\underline{\beta}_\gamma$  نحصل على

$$p(\underline{\gamma}|\underline{y}) \propto (1+c)^{-q(\gamma)^2} \int (\sigma^2)^{-(n/2+1)} \exp(-S(\gamma)/2\sigma^2) d\sigma^2 p(\underline{\gamma}) \dots (5)$$

علماً أن  $S(\gamma)$  هي مجموع مربعات البواقي لإنموذج  $\gamma$ .

ثم بتكامل (5) ذات دالة كثافة معكوس كما (Inverse Gamma) إلى  $\sigma^2$  نحصل على الكثافة اللاحقة إلى  $\underline{\gamma}$

$$\begin{aligned} p(\underline{\gamma}|\underline{y}) & \propto (1+c)^{-q(\gamma)^2} S(\gamma)^{-n/2} p(\underline{\gamma}) \dots (6) \\ & \propto (1+c)^{-q(\gamma)^2} S(\gamma)^{-n/2} \prod_{i=1}^r \pi_i^{\gamma_i} (1-\pi_i)^{1-\gamma_i} \end{aligned}$$

ويسبب إن  $\gamma$  تأخذ  $2^r$  من القيم المتنوعة فانه من غير الملائم أن نحصل على الاحتمال اللاحق بواسطة الحسابات المباشرة ما لم تكن  $r$  صغيرة، في حين  $r$  تكون كبيرة في معظم التطبيقات لذا تم اقتراح عدة إستراتيجيات لحل هذه المسألة مثل انحدار الخطوات المتسلسلة علماً إن معاين Gibbs المقترح من قبل Gelfand & Smith (1990) يزود ببدائل عشوائية للحصول على التوزيع اللاحق  $\underline{\gamma} | \underline{y}$  بواسطة عينة مونت كارلو لسلسلة ماركوف وفيما يلي توضيح خطوات معاين Gibbs كآلاتي :

i. اختيار قيمة أولية  $\gamma^{[0]} = (\gamma_1^{[0]}, \dots, \gamma_r^{[0]})$  والتي تولد من بعض التوزيعات.

ii. تولد  $\gamma_i$  وينجح من  $i = 1, \dots, r$   $\cdot p(\gamma_i | \underline{y}, \gamma_{j \neq i})$

وتتضمن الخطوة ii مرحلتين الأولى فترة التمهيد حيث في النهاية يفترض اقتراب المعاينة من التوزيع المشترك إلى  $p(\underline{y} | \underline{\gamma})$ ، المرحلة الثانية هي المعاينة المسبقة وان  $\gamma_i$  المولدة خلال فترة المعاينة تستخدم للاستدلال، من (6) الاحتمال الشرطي إلى  $\gamma_i$  هو

$$p(\gamma_i | \underline{y}, \gamma_{j \neq i}) \propto p(\underline{y} | \underline{\gamma}) p(\gamma_i) \\ \propto \pi^{\gamma_i} (1 - \pi_i)^{1 - \gamma_i} (1 + c)^{-q(\gamma)/2} S(\gamma)^{-n/2} \quad \dots (7)$$

وبسبب  $\gamma_i$  هي ثنائية فالاحتمال الشرطي أعلاه نحصل عليه بواسطة حساب (7) لكل من  $\gamma_i = 0$  ,  $\gamma_i = 1$  بحيث أن

$$p(\gamma_i = 1 | \underline{y}, \gamma_{j \neq i}) = 1 / (1 + h) \quad \dots (8)$$

$$h = ((1 - \pi_i) / \pi_i) (c + 1)^{1/2} (S(\gamma^1) / S(\gamma^0))^{n/2}$$

وبفرض  $\underline{\gamma} = \underline{\gamma}^{[0]}$  قبل توليد  $\gamma_i$  ستصبح  $S(\gamma^0)$  معلومة أي  $\underline{y}' \underline{X}_{\gamma^0} (\underline{X}'_{\gamma^0} \underline{X}_{\gamma^0})^{-1} \underline{X}_{\gamma^0} \underline{y}$

الصعوبة تكمن في حساب  $S(\gamma^1)$  أي حساب  $\underline{y}' \underline{X}_{\gamma^1} (\underline{X}'_{\gamma^1} \underline{X}_{\gamma^1})^{-1} \underline{X}_{\gamma^1} \underline{y}$

ولجعل الحسابات مستقرة فان Dongarra<sup>[3]</sup> وآخرين وضعوا التصميم المعياري للمصفوفة X حيث

$$\tilde{X} = XK^{-1} , \quad K = \text{diag} (\|x_1\|, \|x_2\|, \dots, \|x_r\|) , \quad \|x_i\| = \left( \sum_{j=1}^n x_{ij} \right)^{1/2}$$

حيث  $x_{ij}$  هو العنصر في الصف j والعمود i إلى X.\*

ولتلافي بعض معوقات المقدّر الأول يتم التركيز على المقدّر الثاني وهو المتوسط اللاحق

(Posterior Mean)  $E(\underline{\beta} | \underline{y}) = \sum_{\gamma \in \Gamma} E(\underline{\beta} | \underline{y}, \gamma) p(\gamma | \underline{y})$  والذي يُقرب باستخدام مقدّر

الخليط (Mixture Estimate) التالي:

$$\hat{\underline{\beta}} = K^{-1} \sum_{j=1}^K E(\underline{\beta} | \underline{y}, \underline{\gamma}^{[j]}) \quad \dots (9)$$

\* لمزيد من التفاصيل راجع المصدر(3).

## 2-1-2 الانحدار اللامعلمي (شرائح الانحدار Regression Splines) [5، 6]

بفرض الأنموذج (1) حيث الهدف تقدير  $g(x)$  بواسطة شريحة الانحدار التكعيبية (Cubic Regression Spline) كالاتي:

$$b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \sum_{k=1}^m \beta_k (x - \tilde{x}_k)_+^3 + \dots \quad (10)$$

علماء أن  $\tilde{x}_1, \dots, \tilde{x}_m$  تمثل العقد (Knots)  $m$  والتي توضع خلال مجال المتغير المستقل بحيث  $\min(x_i) < \tilde{x}_1 < \dots < \tilde{x}_m < \max(x_i)$  ;  $(z)_+ = \max(0, z)$

وبإبدال  $g(x)$  في (1) بالتقريب في (10) يمكن إعادة كتابة الانحدار اللامعلمي كأنموذج انحدار خطي  $y = X\beta + \epsilon$  وبفرض  $r = m+4$  فان:

$$\beta = (b_0 \ b_1 \ b_2 \ b_3 \ \beta_1 \ \dots \ \beta_m)' \quad (r \times 1) \text{ متجه}$$

$$X = \begin{pmatrix} 1 & x & x^2 & x^3 & (x - 1\tilde{x}_1)_+^3 & \dots & (x - 1\tilde{x}_m)_+^3 \end{pmatrix} \quad (n \times r) \text{ مصفوفة}$$

$$\underline{x} = (x_1 \ \dots \ x_n)' \quad (n \times 1) \text{ متجه} \quad \underline{1} = (1 \ \dots \ 1)' \quad (n \times 1) \text{ متجه}$$

السؤال المهم المرتبط مع مطابقة شرائح الانحدار هو اختيار عدد ومواقع العقد فإذا وضعت العقد بشكل سيئ فان تفاصيل من المنحنى ستفقد، كما أن العدد الكبير من العقد المتضمنة في مطابقة الشريحة سيكون لها تباين موضعي عال، وأحد طرائق حل المشكلة هو تقديم عدد كبير من العقد الممكنة ثم نُختار العقد المعنوية ونظراً للسلبيات التي تعاني منها الطريقة فان مشكلة العدد الكبير من العقد قد حل بواسطة أسلوب بيز لاختيار العقد المعنوية، اما تقدير  $g(x)$  بواسطة تقريب المتوسط اللاحق  $E(g | y)$  يكون

$$\hat{g}(x) = X' \sum_{j=1}^K E[\beta | \gamma^{[j]}, y] \quad \dots \quad (11)$$

### 2-2 الطريقة المقترحة لأسلوب بيز \*

من خلال استعراض طرائق الانحدار اللامعلمي لوحظ إن القليل جداً من البحوث تناول طرائق بيز للانحدار اللامعلمي بالرغم من كفاءتها العالية إذ أنها تلائم العديد من النماذج والمعقدة منها كدوال عدم التجانس ونماذج التغير الحيزي الصارم والواطيء في التذبذب، وعلى صعيد القطر لم يتطرق أي بحث لهذا الموضوع لأسباب يمكن أدرجها بالصعوبة النظرية والحاجة لبرمجة متقدمة جداً على المستوى العملي فضلاً عن معوقات أخرى كمشاكل التقارب، ولكون مقدر المنوال اللاحق (Posterior Mode) يتطلب عشرات المرات تكرارات اكثر من مقدر المتوسط اللاحق (Posterior Mean) للحصول على التقارب وأحياناً يتعذر الحصول عليه للحصول على مقدرات من النوعية نفسها لذا تم اقتراح تحويل مقدر المنوال لتقريب فترة الحصول على التقارب وهي أن

\* الطريقة مقترحة في البحث وهي محاولة لاجتياز المعوقات الكبيرة في الطريقة قبل التحوير.

لا توضع قيم البداية إلى  $\gamma^{[0]} = [\gamma_1, \dots, \gamma_r]$  اعتباطاً بل استناداً إلى عدد العقد في المطابقة التكميلية القطعية والتي تعتمد أسلوب إضافة الخطوات المتسلسلة في اختيار العقد\*، وقد كانت استراتيجية كفاءة للاختيار بين النماذج وان  $r = m+4$ ، وكما هو معلوم بان الاختيار السيئ لقيم البداية قد يسبب تقارباً بطيئاً أو لا يتحقق التقارب أصلاً وبعد الحصول على قيمة مقدّر المنوال اللاحق إلى  $p(\gamma | y)$  تقدر معاملات الانحدار بواسطة المربعات الصغرى OLS وفقاً إلى  $\hat{\gamma}_M$  وبذلك يتم الابتعاد عن التعقيد في تقدير المعاملات باعتماد الأسلوب الخليط والمعطى في (9)، إن التحوير المقدم استخدم معاين Gibbs في تحديد العقد المعنوية والتي استخدمت أسلوب البحث العشوائي للمرور بالنماذج المسموحة كافة وليس مجموعة جزئية من مجال النماذج الكاملة، وان  $g(x)$  تقدر باستخدام شريحة الانحدار التكميلية.

### 3. الجانب التجريبي

لغرض تطبيق ما ورد في الجانب النظري تم اللجوء إلى أسلوب المحاكاة (Simulation) والذي يحاكي عدداً كبيراً جداً من الحالات الافتراضية التي يمكن أن تظهر في الواقع العملي لتكون النتائج أكثر شمولية وتعميماً.

#### 3-1 توليد المتغيرات والعينة المستخدمة

تم تنفيذ تجارب المحاكاة على ثلاثة حجوم للعينات صغيرة ومتوسطة وكبيرة كالاتي  $n=50$  و  $n=100$  و  $n=200$  وبواقع 1350 تجربة وبتكرار 500 لكل تجربة محاكاة ولحالة التوزيع الطبيعي للخطأ العشوائي وكالاتي:

- i. توليد المتغير التوضيحي  $X$  بتوزيع منتظم  $X \sim U(0,1)$  بالاستناد إلى طريقة Box-Muller والتي تعتمد على أسلوب توليد متغيرين عشوائيين  $U_1, U_2$  يتبعان التوزيع المنتظم  $U(0,1)$ .
- ii. الخطأ العشوائي ويتم توليده ليتوزع توزيعاً طبيعياً بمتوسط صفر وتباين  $\sigma^2$  أي  $\epsilon \sim N(0, \sigma^2)$
- iii. المتغير المعتمد  $y$  ويتم توليده من خلال النماذج المستخدمة في تجارب المحاكاة وذلك باستخدام المتغيرات التوضيحية التي تم توليدها في الفقرة i، مضافاً إليها الأخطاء العشوائية المولدة في الفقرة ii، ولكل أنموذج من النماذج المدروسة.

#### 3-2 النماذج المستخدمة في تجارب المحاكاة

هناك عدد كبير من النماذج التي لا يمكن تناولها جميعها في آن واحد في تجارب المحاكاة نظراً لما تتطلبه من الجهد والمساحة الواسعة وقد حاولنا قدر المستطاع أن تكون النماذج المتناولة متنوعة لتلائم جميع الحالات وقد أخذت النماذج من بحوث منشورة وهي كالاتي:

$$1. \text{ الأنموذج غير الخطي وصيغته } y = x + 2\exp(-16x^2)$$

\* لمزيد من التفاصيل حول أسلوب اختيار العقد راجع المصدر المبحث 2-1-4-2 في المصدر (1).

2. الأنموذج الخاص (تغير حيزي واطئ في التذبذب) وصيغته

$$y = \sqrt{x(1-x)} \sin\left(\frac{2\pi(1+2^{(9-4j)/5})}{x+2^{(9-4j)/5}}\right) \quad j=3$$

3. الأنموذج الخاص بدالة الانحدار غير المتجانسة وصيغته

$$y = \exp(-400(x-0.6)^2) + (5/3) \exp(-500(x-0.75)^2) + 2\exp(-500(x-0.9)^2)$$

### 3-3 تنفيذ تجارب المحاكاة لطرائق الانحدار اللامعلمي

لكل أنموذج من النماذج المدروسة تم عمل ما يأتي :

1. توليد المتغير التوضيحي X ليتوزع توزيعاً منتظماً U(0,1) مع أخطاء عشوائية تتبع توزيعاً طبيعياً.

2. وفيما يخص أسلوب بيز اختيرت العقد للمتغير التوضيحي X واحد لكل 3 إلى 5 مشاهدات وكحد أقصى لحد 40 عقدة، إن عدد العقد تلك تضمن العقدة عند أو قريبة من الموقع المطلوب لمسك الانحناء في دالة الانحدار، وكانت m=24 وان r = 28 ، ومع أن عدد العقد المقدمة بواسطة m كبيرة بحيث أن r أيضاً كبيرة لكن عدد المتغيرات المعنوية المطلوبة للحصول على تقريب جيد عادة تكون غير كبيرة وهذا يساعد على تسريع الخوارزمية.

3. تنفيذ معايير Gibbs لفترة تمهيدية بواقع 100 تكرار وفترة المعاينة كانت بواقع 1500

تكرار مع قيم بداية اعتباطية فيما يخص مقدري المتوسط والمنوال لبيز وهي

$$\gamma^{[0]} = [1,0,\dots,1,0,1]'$$

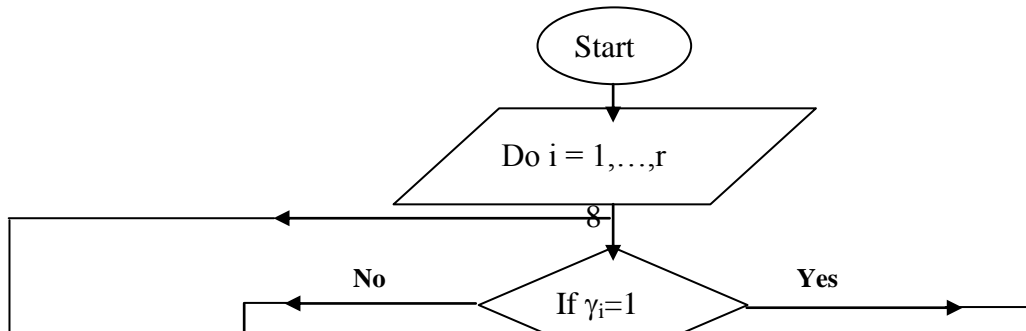
أما التقارب فتم الحصول عليه بواسطة الاحتمال اللاحق p(y | γ) لتكرار Gibbs وان المتغيرات المختارة تم اختبار معنويتها بالمستويين 5% ، 1% وباستخدام الإحصاء t. كما تم استخدام تقدير المنوال اللاحق لدالة الانحدار بواسطة دمج مطابقة المربعات الصغرى للأنموذج مع المعادلة (10) وأن مقدّر المتوسط اللاحق لدالة الانحدار تم الحصول عليه بواسطة دمج مقدّر المتوسط اللاحق إلى β مع المعادلة (10) ، ولتبسيط خوارزمية توليد γ بالنسبة لطرائق بيز تم عمل مخطط توضيحي هو المخطط (1).

4. تم استخدام معيار لقياس كمية تقدير الدالة اللامعلمية هو تكامل مربع الخطأ ISE حيث

يقرب التكامل بواسطة معدل مربع الخطأ حول 400 من نقاط المسافات المتساوية كالاتي:

$$ISE = (1/400) \sum_{i=1}^{400} (g(z_i) - \hat{g}(z_i))^2$$

مخطط (1): خوارزميه توليد  $\gamma_i$





### 4-3 تجارب المحاكاة \*

بعد تطبيق الأفكار الواردة في الجانب النظري تم الحصول على عدة نتائج ولكل أنموذج من

---

\* تم تنفيذ تجارب المحاكاة بكتابة برامج بلغة Visual Basic اضافة لبرامج بلغة S-Plus/Fortran.

النماذج المدروسة، علماً بأنه قد تم تفسير اغلب النماذج وكذلك الرسوم بالاعتماد على معيار ISE والذي تعتمده اغلب البحوث كونه يأخذ ما بين النقاط عند التقدير وكما يلي :

### الأنموذج الأول \* \*

للأنموذج غير الخطي لوحظ بالنسبة للممهدين NW,LLR أن معيار ISE لم يحقق نتائج جيدة وكان هناك بعض التقدم في إنجاز Double Exp. مقارنة مع بقية الدوال، وبالنسبة لطريقة LLSR يلاحظ تفوق الاختيار الفعال لعرض الحزمة  $\hat{h}_{DPI}$  وقد تطلب الأنموذج غير الخطي مقدر عرض حزمة عمومي (Global)، أما الاختيار الفعال المقترح  $\hat{h}_{DROT}$  فقد أعطى افضل النتائج لحالتي التباين المتوسط والواطي ومن الشكل (2) يمكن ملاحظة إخفاقات واضحة لطرائق Kernel في الاقتراب من المنحنى الحقيقي حيث يعاني المنحنى من ذبذبات كثيرة فيما كان إنجاز كل من طريقة Lowess وطريقة LLSR ويعرض حزمة  $\hat{h}_{DROT}$  اكثر قرباً من المنحنى الحقيقي ، بالنسبة إلى طرائق بيز أعطت نتائج كفاءة عالية مقارنة بجميع الطرائق اللامعلمية وكان إنجاز مقدر بيز المقترح هو الأفضل يليه مقدر المنوال ثم مقدر المتوسط، ولم نجد أي تذبذب في مقدرات بيز.

### الأنموذج الثاني

للأنموذج الخاص (التغير الحيزي الواطي في التذبذب) (Low Spatial Variability of Oscillations) لوحظ بالنسبة للممهدين NW, LLR أن معيار ISE لم يعط نتائج جيدة مع انه كان هناك بعض التقدم في إنجاز دالة Suggest2 مقارنة مع بقية الدوال، ولما يخص طريقة LLSR تفوق الاختيار الفعال لعرض الحزمة  $\hat{h}_{DPI}$  على الممهدين أعلاه وطريقة Lowess، ويلاحظ من الشكل (5) بعض التحسن في تقدير ممد LLR باعتماد الدالة المقترحة Suggest2 مقارنة مع بقية الدوال. وبالنسبة لطرائق بيز نجدها تمتلك كفاءة عالية في تبني الأنموذج الخاص ولم يعان المنحنى الخاص بأساليب بيز من أي تذبذب علماً أن إنجاز مقدر المتوسط هو الأفضل يليه المنوال ثم المقدر المقترح لبيز.

### الأنموذج الثالث

الانموذج هو دالة الانحدار غير المتجانسة (Heterogeneous Regression Function)،  
~~يلاحظ فيه أن معيار ISE لم يعط نتائج جيدة مع انه كان هذا~~  
Double Exp., Suggest1 وعلى التوالي مقارنة مع البقية، أما بالنسبة لطريقة LLSR فان

\* \* لمزيد من التفاصيل حول طرائق Kernel، Lowess، LLSR مع اختيارات عرض الحزمة اضافة الى المقترح لعرض الحزمة

$\hat{h}_{DROT}$  راجع المصدر (1) .

الاختيار الفعال للحزمة  $\hat{h}_{DPI}$  أعطى نتائج افضل يليه الاختيار الفعال المقترح  $\hat{h}_{DROT}$  علماً أن الحصول على التمهيد الخاص بهذا الأنموذج تطلب عرض حزمة مكيف، ومن الشكل (8) يمكن ملاحظة أن تقدير طريقة LLSR يعاني من تذبذب عال في الحد الأيسر كذلك فان تقدير Lowess يعاني من تذبذب عال في الذيل، وفيما يخص طرائق بيز فكانت كفاءة بدرجة عالية في تمثيلها لانموذج الانحدار غير المتجانس ولم يظهر أي تذبذب في تقديرات طرائق بيز باستثناء بعض التذبذب الطفيف في الحد الأيسر بالاعتماد على مقدر المنوال ويمكن ملاحظة الشكل (9) فقد كان مقدر المتوسط هو الأفضل ثم مقدر بيز المقترح واخيراً مقدر المنوال.

جدول (1): طرائق kernel الطرائق اللامعلمية الأخرى LLSR و Lowess ، طرائق بيز ولانموذج الأول

The	S.	$\sigma=1/2$	$\sigma=1/4$	$\sigma=1/8$
-----	----	--------------	--------------	--------------

Method	Size	MASE	ISE	MASE	ISE	MASE	ISE
NW	50	0.03579	0.77448	0.01313	0.70819	0.00439	0.65954
Kernel	100	0.01893	0.66297	0.00637	0.60911	0.00291	0.58961
Ep.*	200	0.01979	0.46829	0.00348	0.44674	0.00205	0.40515
NW	50	0.03240	0.66674	0.01239	0.66523	0.00358	0.56069
Kernel	100	0.01789	0.49581	0.00579	0.47549	0.00200	0.43102
Dou. Exp.	200	0.01658	0.39028	0.00325	0.37014	0.00142	0.32125
NW	50	0.03570	0.77105	0.01244	0.70432	0.00406	0.63954
Kernel	100	0.01891	0.64421	0.00599	0.59618	0.00278	0.55886
Suggest1	200	0.01765	0.46615	0.00456	0.44388	0.00189	0.38011
NW	50	0.01403	0.75861	0.01639	0.69523	0.00587	0.63326
Kernel	100	0.01999	0.58836	0.00792	0.55549	0.00326	0.53604
Suggest2	200	0.02022	0.44831	0.00558	0.41611	0.00259	0.37166
LLR	50	0.03124	0.79761	0.01244	0.72654	0.00381	0.67219
Kernel	100	0.01876	0.66939	0.00494	0.61152	0.00271	0.60113
Ep.	200	0.01524	0.46855	0.00281	0.45963	0.00193	0.41522
LLR	50	0.02806	0.68938	0.01228	0.67557	0.00331	0.58638
Kernel	100	0.01642	0.50152	0.00456	0.48107	0.00190	0.43931
Dou. Exp.	200	0.01385	0.41266	0.00167	0.37107	0.00134	0.34439
LLR	50	0.03023	0.79399	0.01234	0.72413	0.00345	0.65442
Kernel	100	0.01749	0.64563	0.00478	0.59854	0.00215	0.57247
Suggest1	200	0.01522	0.46711	0.00327	0.44733	0.00188	0.39522
LLR	50	0.03411	0.78088	0.01377	0.69557	0.00426	0.63499
Kernel	100	0.01890	0.59421	0.00780	0.56107	0.00322	0.54132
Suggest2	200	0.01586	0.45619	0.00339	0.42154	0.00241	0.37772
LLSR with	50	0.04796	0.27896	0.04389	0.23856	0.04156	0.20367
Bandw.	100	0.04387	0.18399	0.03568	0.17979	0.03158	0.13200
$\hat{h}_{ROT}$	200	0.02134	0.15854	0.01154	0.14524	0.00986	0.13537
LLSR with	50	0.03244	0.21521	0.03189	0.19529	0.03116	0.17651
Bandw.	100	0.03113	0.15575	0.02343	0.14727	0.02122	0.12755
$\hat{h}_{DPI}$	200	0.01120	0.11395	0.00902	0.11286	0.00713	0.11129
LLSR with	50	0.03301	0.23887	0.02425	0.18856	0.02312	0.17211
Su. Bandw.	100	0.03213	0.17581	0.02314	0.13671	0.01145	0.11978
$\hat{h}_{DROT}$	200	0.01129	0.12464	0.00711	0.11046	0.00510	0.10734
Lowess	50	0.06206	0.15244	0.04600	0.14381	0.04218	0.13504
	100	0.04546	0.13308	0.03644	0.13381	0.03448	0.11206
	200	0.01410	0.10233	0.01294	0.10445	0.01091	0.10258
Bayesian	50	0.07261	0.04452	0.05367	0.04397	0.04765	0.03633
with	100	0.03995	0.00689	0.02832	0.00489	0.02715	0.00349
Mean	200	0.00975	0.00588	0.00744	0.00403	0.00716	0.00256
estimate							
Bayesian	50	0.05496	0.04074	0.05197	0.03918	0.03124	0.02177
with	100	0.03513	0.00669	0.02621	0.00425	0.02413	0.00309
Mode	200	0.00845	0.00407	0.00733	0.00303	0.00635	0.00165
estimate							
Suggest	50	0.04424	0.03206	0.04132	0.03023	0.02939	0.01023
Bayesian	100	0.02849	0.00571	0.01528	0.00324	0.01037	0.00277
estimate	200	0.00778	0.00341	0.00622	0.00215	0.00592	0.00133

جدول (2): طرائق kernel الطرائق اللامعلمية الأخرى LLSR و Lowess ، طرائق بيز وللانموذج الثاني

The Method	S. Size	$\sigma=1/2$		$\sigma=1/4$		$\sigma=1/8$	
		MASE	ISE	MASE	ISE	MASE	ISE

\* لمزيد من التفاصيل حول دوال Kernel المستخدمة راجع المصدر (1) ص 23 .

NW	50	0.18573	0.75833	0.18384	0.71776	0.15636	0.53813
Kernel	100	0.14457	0.66982	0.13569	0.63693	0.14313	0.49416
Ep.	200	0.14203	0.55504	0.13492	0.52214	0.12737	0.41410
NW	50	0.10755	0.68296	0.10653	0.66123	0.09960	0.49128
Kernel	100	0.09998	0.60225	0.09654	0.55669	0.09260	0.44134
Dou. Exp.	200	0.08216	0.46337	0.08115	0.45544	0.07158	0.37316
NW	50	0.16953	0.73903	0.15205	0.69303	0.15273	0.51860
Kernel	100	0.13739	0.64733	0.13486	0.60693	0.12377	0.48054
Suggest1	200	0.12515	0.52258	0.12599	0.49850	0.11621	0.40105
NW	50	0.08414	0.66905	0.08344	0.64338	0.06307	0.48973
Kernel	100	0.07155	0.57921	0.07109	0.55316	0.03211	0.43274
Suggest2	200	0.07105	0.44682	0.07080	0.40203	0.01132	0.35136
LLR	50	0.18790	0.68094	0.18589	0.66235	0.16734	0.51362
Kernel	100	0.16729	0.63992	0.15816	0.57612	0.14425	0.45341
Ep.	200	0.14762	0.51153	0.14394	0.49467	0.12913	0.38712
LLR	50	0.11578	0.65912	0.10791	0.64361	0.10350	0.45698
Kernel	100	0.10358	0.58107	0.09921	0.53333	0.09421	0.42113
Dou. Exp.	200	0.10267	0.44170	0.08189	0.43421	0.07328	0.35125
NW	50	0.17845	0.67804	0.16821	0.67235	0.15670	0.50489
Kernel	100	0.15796	0.61386	0.14607	0.56781	0.12466	0.45204
Suggest1	200	0.13490	0.48722	0.13226	0.47023	0.12378	0.37663
LLR	50	0.09648	0.63914	0.09496	0.60404	0.06611	0.44142
Kernel	100	0.08468	0.56673	0.07253	0.53022	0.03219	0.40212
Suggest2	200	0.08516	0.38152	0.07146	0.32616	0.01235	0.29824
LLSR with	50	0.11977	0.33196	0.11405	0.30914	0.10366	0.26543
Bandwidth	100	0.11196	0.29977	0.08785	0.25317	0.07733	0.23717
$\hat{h}_{ROT}$	200	0.07112	0.24112	0.07867	0.21766	0.07552	0.21373
LLSR with	50	0.09224	0.27224	0.09178	0.25144	0.07050	0.21942
Bandwidth	100	0.07115	0.23011	0.06367	0.20811	0.04617	0.19801
$\hat{h}_{DPI}$	200	0.06011	0.19115	0.04714	0.18424	0.01322	0.17413
LLSR with	50	0.10382	0.31125	0.10125	0.27758	0.09313	0.25378
Su. Bandwidth	100	0.09718	0.27718	0.08772	0.23019	0.07023	0.21254
$\hat{h}_{DROT}$	200	0.07109	0.23109	0.06855	0.20632	0.06472	0.20522
Lowess	50	0.09015	0.35697	0.04831	0.33531	0.03385	0.29125
	100	0.08619	0.30318	0.06658	0.27911	0.02355	0.25910
	200	0.05249	0.26250	0.02704	0.22532	0.00846	0.22389
Bayesian with	50	0.03184	0.00717	0.03203	0.00622	0.02727	0.00402
Mean	100	0.02205	0.00556	0.02622	0.00473	0.00863	0.00265
estimate	200	0.00343	0.00236	0.00440	0.00222	0.00393	0.00148
Bayesian with	50	0.03288	0.00786	0.03536	0.00677	0.03346	0.00465
Mode	100	0.02370	0.00756	0.02827	0.00652	0.00910	0.00364
estimate	200	0.00746	0.00327	0.00836	0.00266	0.00544	0.00257
Suggest	50	0.04370	0.00888	0.04442	0.00731	0.03707	0.00517
Bayesian	100	0.03052	0.00822	0.04135	0.00771	0.01405	0.00706
estimate	200	0.00849	0.00576	0.00877	0.00403	0.00792	0.00344

جدول (3): طرائق kernel اللامعتمدة الأخرى LLSR و Lowess ، طرائق بيز وللا نموذج الثالث

The Method	S. Size	$\sigma=1/2$		$\sigma=1/4$		$\sigma=1/8$	
		MASE	ISE	MASE	ISE	MASE	ISE

NW	50	0.19731	1.33453	0.19365	1.32573	0.18511	1.29138
Kernel	100	0.18622	1.37331	0.18449	1.35515	0.16883	1.33422
Ep.	200	0.17461	1.39214	0.17226	1.38482	0.14418	1.35138
NW	50	0.18366	1.24821	0.18351	1.22353	0.17184	1.20597
Kernel	100	0.17315	1.22161	0.17211	1.20042	0.14531	1.19581
Dou. Exp.	200	0.16733	1.18595	0.16143	1.17377	0.13375	1.16987
NW	50	0.19296	0.84393	0.19150	0.79578	0.18433	0.77887
Kernel	100	0.18573	0.79958	0.18422	0.76095	0.14651	0.73141
Suggest1	200	0.17252	0.76696	0.16651	0.73245	0.13813	0.62167
NW	50	0.26471	1.36381	0.26222	1.37737	0.23601	1.32986
Kernel	100	0.26216	1.37236	0.25803	1.34542	0.20611	1.28283
Suggest2	200	0.25631	1.39194	0.24988	1.31413	0.19322	1.26174
LLR	50	0.25506	1.31349	0.24025	1.29870	0.23593	1.27507
Kernel	100	0.23495	1.35606	0.21683	1.34660	0.17277	1.29416
Ep.	200	0.18487	1.37455	0.18121	1.37131	0.17135	1.32507
LLR	50	0.22156	1.22769	0.19147	1.21409	0.18940	1.19610
Kernel	100	0.19132	1.19864	0.18621	1.18398	0.16398	1.18138
Dou. Exp.	200	0.17259	1.16992	0.16751	1.13994	0.14415	1.13415
LLR	50	0.22557	0.81282	0.22898	0.76129	0.22429	0.74114
Kernel	100	0.19241	0.77161	0.19203	0.68121	0.16216	0.66662
Suggest1	200	0.18385	0.75256	0.17762	0.64581	0.15274	0.56143
LLR	50	0.27102	1.34734	0.26701	1.34693	0.23853	1.31141
Kernel	100	0.26870	1.35616	0.26616	1.32873	0.20767	1.26335
Suggest2	200	0.25697	1.37214	0.25622	1.30162	0.19944	1.25431
LLSR with Bandwidth $\hat{h}_{ROT}$	50	0.12526	0.44736	0.12231	0.41153	0.09505	0.39169
	100	0.11289	0.43979	0.11268	0.40760	0.07198	0.38846
	200	0.10211	0.40588	0.09418	0.36966	0.07212	0.35082
LLSR with Bandwidth $\hat{h}_{DPI}$	50	0.09231	0.40564	0.08135	0.37895	0.05097	0.35845
	100	0.07211	0.40335	0.07023	0.37238	0.03166	0.34104
	200	0.07053	0.37168	0.04005	0.34577	0.00908	0.31193
LLSR with Su. Bandwidth $\hat{h}_{DROT}$	50	0.10363	0.41726	0.10237	0.38862	0.07122	0.36945
	100	0.09373	0.41151	0.08159	0.38315	0.04786	0.34948
	200	0.07184	0.38116	0.07127	0.35236	0.02174	0.32241
Lowess	50	0.16572	0.46572	0.14304	0.43492	0.13163	0.40904
	100	0.14005	0.44232	0.11770	0.41294	0.10113	0.38855
	200	0.11276	0.42447	0.10261	0.37542	0.09706	0.37023
Bayesian with Mean Estimate	50	0.07705	0.00957	0.08566	0.00331	0.06227	0.00293
	100	0.05240	0.00607	0.06003	0.00306	0.04074	0.00265
	200	0.00593	0.00316	0.00635	0.00209	0.00436	0.00114
Bayesian with Mode estimate	50	0.09664	0.04429	0.09755	0.00712	0.07616	0.00823
	100	0.06195	0.02351	0.07396	0.00509	0.04595	0.00776
	200	0.00820	0.00765	0.00833	0.00344	0.00599	0.00321
Suggest Bayesian estimate	50	0.08393	0.02321	0.09322	0.00343	0.07505	0.00353
	100	0.05843	0.00721	0.06021	0.00310	0.04517	0.00281
	200	0.00622	0.00411	0.00718	0.00218	0.00443	0.00298

#### 4 . الاستنتاجات

1. إن طريقتي LLSR و Lowess هي اكثر كفاءة وفعالية من بقية طرائق Kernel لكن تبقى طرائق Kernel غير ذي فعالية مقارنة بأساليب بيز كونها لا تقدر أنموذج الانحدار اللامعلمي.

2. أظهرت نتائج أساليب بيز للانحدار اللامعلمي تفوقاً عالياً لمعظم النماذج حتى الخاصة إذ لم تعان مقدّرات تلك الأساليب من أي تذبذب في الذبول ما عدا بعض التذبذب الطفيف في الحد الأيسر لحالة أنموذج الانحدار غير المتجانس وبالاعتماد على مقدّر المنوال ، واضح بان أسلوب بيز للانحدار اللامعلمي مكيف بشكل ناجح لجميع النماذج ولا يمكن تعميم حالة التذبذب الطفيف في الحد الأيسر لجميع نماذج الانحدار غير المتجانسة والتفوق ناتج من استخدام معايير Gibbs حيث اعتماد أسلوب البحث العشوائي بطريقة مونت كارلو لسلسلة ماركوف MCMC للمرور بالنماذج المسموحة كافة.

3. إن التحوير المقترح لأسلوب بيز للانحدار اللامعلمي أعطى نتائج جيدة مقارنة مع أسلوب بيز قبل التحوير فضلاً عن اختزله لبعض تعقيدات أسلوب بيز الكثيرة والتي تطلبت برمجة متقدمة جداً.

## 5 . التوصيات

1. كما هو معروف بان أسلوب بيز للانحدار اللامعلمي يستخدم طريقة مونت كارلو لسلسلة ماركوف MCMC وبالرغم من تحقيقه كفاءة عالية إلا إن الصعوبات النظرية والبرمجية أدت إلى ابتعاد العديد من الباحثين عن استخدامه، لذا نوصي باستخدام تحوير بيز المقترح الذي قلل من بعض الصعوبات البرمجية لأسلوب بيز فضلاً عن إعطائه نتائج جيدة وتقارب من المنحنى الحقيقي للعديد من النماذج والخاصة منها.
2. لقد تطرق البحث إلى اعتماد معايير Gibbs في توليد الاحتمال اللاحق والتي حقق نتائج فعالة لكن صعوبة الحصول على التقارب تزداد مع زيادة عدد المتغيرات لذا نوصي باستخدام أساليب أخرى كالمعينة المركزة (Focused Sampling) كونها تعالج عدداً كبيراً من المتغيرات في الوقت نفسه والتركيز يكون على المتغيرات الفعالة.
3. التوسع في دراسة أساليب بيز للانحدار اللامعلمي للحالة ثنائية ومتعددة الأبعاد.

## 6. المصادر

1. خمو، خلود يوسف ، (2004) " مقارنة أساليب بيز مع طرائق أخرى لتقدير منحنى

### الأنحدار

اللامعلمي " أطروحة دكتوراه في الاحصاء ،كلية الادارة والاقتصاد، جامعة بغداد.

2. Denison, D.G.T., Holmes, C.C., Mallick, B.K. & Smith, A. F.M., 2003, "**Bayesian Methods for Nonlinear Classification and Regression**" Wiley Series in probability and Statistics, John Wiley & Sons.
3. Dongarra, J.J., Moler, C.B., Bunch, J.R. & Stewart, G.W., "**Linpack User's Guide**", Society for Industrial and Application Mathematics, Philadelphia, PA.
4. Schimek, M.G., (2000), "**Smoothing and Regression Approaches, Computation and Application**", John Wiley & Sons.
5. Smith, M. & Kohn, R., (1996), " Nonparametric Regression using Bayesian variable Selection ", *Journal of Econometric*, Vol. 75, 317-343.
6. Smith, M. & Kohn, R., (1996), "A Bayesian Approach to Nonparametric Bivariate Regression " *J.Amer.Statist.Assoc.*, Vol. 92, 1522-1533.