

PENELITIAN BIDANG REGRESI SPLINE MENUJU TERWUJUDNYA PENELITIAN STATISTIKA YANG MANDIRI DAN BERKARAKTER

I Nyoman Budiantara

Jurusan Statistika FMIPA Institut Teknologi Surabaya
e-mail: i_nvoman_b@statistika.its.ac.id

PENDAHULUAN

Statistika merupakan suatu ilmu yang memuat Statistika Deskriptif dan Statistika Inferensial. Didalam Statistika Inferensial kesimpulan yang diperoleh memuat unsur ketidakpastian dan variasi. Kata Statistika mempunyai makna yang sangat berbeda dengan kata Statistik yang merupakan kumpulan nilai atau kuantitas (Bain & Engelhardt,1992). Ilmu Statistika dirancang dan diciptakan oleh manusia, sehingga kebenarannya adalah relatif dan diukur berdasarkan kebenaran manusia, sebatas kemampuan optimal yang dimiliki oleh manusia yang mendesain ilmu tersebut (Budiantara, 2009a). Dalam Statistika, dasar pengambilan keputusan adalah probabilitas. Seseorang Statistikawan akan melaksanakan keputusan Statistika apabila dia yakin probabilitas (peluang) keputusan tersebut akan benar adalah besar (misalnya 90%, atau 95%, atau 99%). Tetapi Statistikawan tidak akan yakin (tidak akan berani) melaksanakan keputusan Statistika, jika peluang keputusan tersebut akan benar adalah kecil (misalnya 10%, atau 5%, atau bahkan 1%).

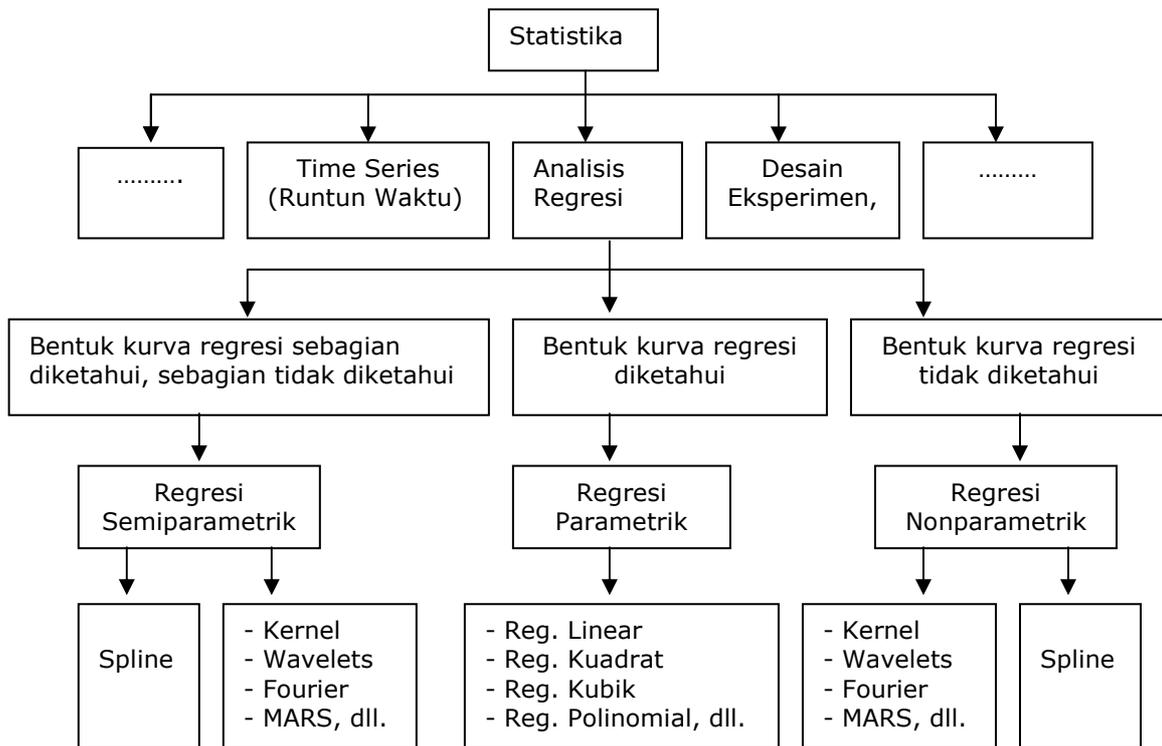
Tata cara pengambilan keputusan dalam Statistika mempunyai beberapa kelebihan diantaranya (Bain & Engelhardt,1992): (1) statistika mampu mengambil keputusan dalam situasi adanya variasi-variasi (keberagaman/perbedaan), (2) statistika mampu mengambil keputusan dalam situasi yang penuh dengan ketidakpastian, (3) statistika mampu mengambil keputusan menggunakan biaya yang relatif murah, (4) statistika mampu mengambil keputusan menggunakan tenaga yang relatif sedikit, dan (5) statistika mampu mengambil keputusan menggunakan waktu yang relatif singkat.

Dalam tulisan ini, akan dijelaskan tentang penelitian regresi Spline menuju terwujudnya penelitian Statistika yang mandiri dan berkarakter. Penelitian yang mandiri dan berkarakter yang dimaksudkan disini adalah penelitian yang memiliki ciri khas (spesifik) pada bidangnya dan mengikuti (tidak menyalahi) kaidah-kaidah serta norma-norma akademik dan non-akademik yang berlaku dalam masyarakat ilmiah. Beberapa contoh peneliti yang berkarakter adalah peneliti tersebut berperilaku jujur secara akademik, tidak melakukan *plagiat*, tidak mengakui karya orang lain sebagai karyanya sendiri, memiliki *road map* penelitian yang jelas, dan yang sejenisnya.

PEMBAHASAN

Penelitian Statistika Berkarakter dan Permasalahannya

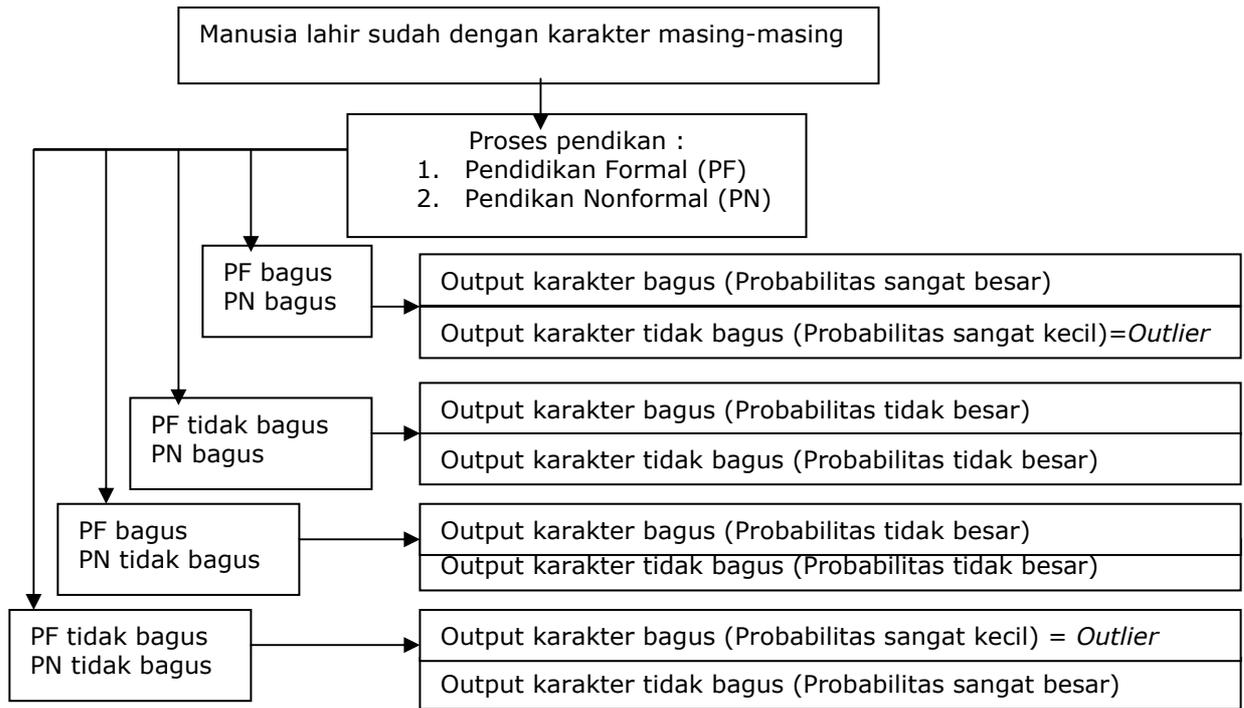
Statistika merupakan salah satu ilmu yang memuat sangat banyak cabang ilmu, diantaranya *time series* (runtun waktu), Proses Stokastik, Probabilitas, Rancangan Percobaan, Analisis regresi, Reliabilitas, dan lain sebagainya (Budiantara,2009b). Analisis regresi merupakan salah satu bidang Statistika yang memainkan peran sangat penting. Analisis regresi umumnya digunakan untuk menyelidiki model pola hubungan fungsional antara satu atau lebih variabel. Disamping itu, analisis regresi sangat bermanfaat untuk peramalan (*forecasting*). Untuk dapat memodelkan satu atau lebih variabel dalam regresi, hal pertama yang semestinya dilakukan adalah apakah variabel-variabel tersebut secara rasional berkorelasi atau tidak. Apabila terjadi korelasi, maka dapat dilakukan pemodelan Statistika dengan menggunakan analisis regresi. Seringkali dalam sebuah pemodelan regresi, seseorang tidak peduli tentang rasionalitas tersebut dan langsung memodelkan data dengan analisis regresi. Tentu cara seperti ini kurang tepat dan bijaksana (Budiantara, 2009b; 2001a; 2001b; 2004a) .



Gambar 1. Pengelompokan Analisis Regresi.

Dewasa ini terdapat 3 (tiga) jenis model pendekatan regresi yang banyak dikembangkan oleh para peneliti, yaitu pendekatan Regresi Parametrik, Regresi Nonparametrik, dan Regresi Semiparametrik (Budiantara, 2000a; 2000b; 2000c; 2004a; 2006a). Pembagian regresi ini dapat dilihat dalam Gambar 1. Penelitian bidang Statistika didasarkan pada bidang-bidang Statistika seperti dalam Gambar 1. Secara khusus, terdapat beberapa persoalan dalam penelitian bidang Statistika yang umumnya juga terjadi pada penelitian bidang lain (selain Statistika). Pada Gambar 2 diberikan ilustrasi sederhana terbentuknya karakter penelitian dari seorang peneliti. Terlihat dengan jelas pada Gambar 2, bahwa sejak manusia lahir telah membawa karakternya masing-masing. Selanjutnya, manusia memasuki proses pendidikan baik pendidikan formal (PF) (yaitu pendidikan yang diperoleh secara formal) maupun nonformal (PN) (yaitu pendidikan yang diperoleh secara nonformal, termasuk pengaruh lingkungan dan yang lainnya). Hasil proses pendidikan akan memberikan beberapa kemungkinan output karakter yang akan terjadi, diantaranya :

- Jika seseorang memperoleh PF yang bagus dan PN juga bagus, maka harapan akan memiliki output karakter yang baik, probabilitasnya sangat besar. Sebaliknya, kelompok ini akan memiliki output karakter yang tidak baik, probabilitasnya sangat kecil. Dalam bidang Statistika hal ini disebut pencilan (*outlier*).
- Jika seseorang memperoleh PF yang tidak bagus, tetapi PN bagus, maka harapan akan memiliki output karakter yang baik, probabilitasnya tidak terlalu besar/kecil. Sebaliknya, seseorang dalam kelompok ini akan memiliki output karakter yang tidak baik, probabilitasnya juga tidak besar/kecil.
- Jika seseorang memperoleh PF yang bagus, tetapi PN tidak bagus, maka harapan akan memiliki output karakter yang baik, probabilitasnya tidak terlalu besar/kecil. Sebaliknya, seseorang dalam kelompok ini akan memiliki karakter yang tidak baik, probabilitasnya juga tidak besar/kecil.
- Jika seseorang memperoleh PF yang tidak bagus dan PN juga tidak bagus, maka harapan akan memiliki output karakter yang baik, probabilitasnya sangat kecil (*outlier*). Sebaliknya, seseorang dalam kelompok ini memiliki output karakter yang tidak baik, probabilitasnya sangat besar.



Gambar 2. Proses Pembentukan Karakter dalam Penelitian.

Berdasarkan uraian di atas, tentu sangat diharapkan setiap orang semestinya mendapatkan pendidikan formal dan nonformal yang bagus, agar harapan output karakternya juga bagus. Secara Statistika, walaupun seseorang mendapatkan pendidikan formal dan nonformal yang bagus, tidak ada yang bisa menjamin secara pasti akan memiliki output karakter yang bagus. Apalagi kalau seseorang mendapatkan pendidikan formal atau nonformal yang tidak bagus, maka tidak akan ada jaminan secara pasti pula akan memiliki output karakter yang bagus, dan malah cenderung mempunyai karakter yang tidak bagus. Karakter yang tidak bagus semacam ini bisa terjadi dalam setiap tingkatan profesi (Mahasiswa S-1, S-2, S-3, Guru, Pejabat Publik dan bahkan Dosen), seperti misalnya :

- Mahasiswa Program Sarjana (S-1) dalam menyusun Skripsi, jika memiliki data yang tidak sesuai dengan metode yang dia rancang, maka dia akan memilih data yang baik-baik saja dan membuang data yang tidak baik, sehingga memperoleh hasil analisis data seperti yang diinginkan.
- Mahasiswa Program Magister (S-2) dalam menyusun Tesis, melakukan *Plagiat* terhadap karya Tesis mahasiswa lain dan diakui sebagai hasil penelitiannya sendiri.
- Mahasiswa Program Doktor (S-3) dalam menyusun Disertasi, menulis ulang kembali Tesis mahasiswa pada Perguruan Tinggi lain, dan diakui sebagai hasil penelitiannya sendiri.
- Calon Pejabat Publik melakukan perbuatan tercela dengan membeli Ijazah untuk memenuhi syarat-syarat administratif dalam Jabatan tersebut.
- Guru yang ingin memperoleh Sertifikasi sebagai pendidik, meminta tolong kepada orang lain untuk membuat karya Ilmiah dan diakui sebagai hasil karyanya sendiri.
- Dosen yang mempunyai Jabatan Guru Besar (Profesor), mengajukan kenaikan Jabatannya menggunakan hasil karya orang lain dan diakui sebagai hasil penelitiannya sendiri, dan contoh-contoh yang lainnya yang sejenis.

Terdapat banyak faktor (sangat kompleks) yang mengakibatkan seseorang memiliki karakter yang tidak bagus, dan secara teoritis akan tergoda untuk menghalalkan segala cara untuk mencapai tujuannya, diantaranya :

- Tuntutan Masyarakat (baik masyarakat ilmiah maupun masyarakat umum) yang terlalu tinggi, atau keinginan seseorang terlalu tinggi, tidak seimbang dengan kemampuan akademik dan non-akademik yang dimilikinya.
- Pendidikan formal (profesional) yang telah diselesaikan mempunyai kualitas yang tidak baik, sehingga dia tidak mampu menyelesaikan pekerjaannya dengan baik.

- c). Seseorang ingin menghasilkan suatu karya (penelitian) yang berkualitas dengan cara cepat, tanpa melalui proses yang benar dan kerja keras.
- d). Penghasilan yang diperoleh oleh seseorang relatif kecil, yang tidak seimbang dengan kebutuhan yang harus dipenuhinya.
- e). Seseorang tidak mengenal dirinya atau Profesinya (bidang keahliannya) dengan baik dan komprehensif.
- f). Seseorang yang tidak percaya pada dirinya sendiri, dan tidak pula percaya pada kemampuan profesional yang dimilikinya.
- g). Seseorang tidak mencintai profesinya dengan sepenuh hati, dan contoh yang lainnya.

Banyak orang yang merasa pesimis dengan profesi yang dimilikinya saat ini, dan bahkan tidak mengenal dirinya dan profesinya dengan baik, sehingga tidak dapat mencapai tujuan yang diinginkan. Pada sisi lain, terdapat orang lain yang memiliki profesi lain, seperti (*seolah-olah*) memiliki kehidupan yang sangat mapan dan berkecukupan (Budiantara, 2011). Oleh karena itu, dia ingin beralih profesi seperti orang lain tersebut agar dapat memperoleh kesuksesan. Kita menyadari bahwa apa yang kita uraikan disini, mirip dengan pernyataan “seperti melihat gunung dari kejauhan yang kelihatannya mulus, tetapi setelah didekati ternyata tidak mulus”. Jika kita beralih profesi, apakah kita pasti akan mendapatkan kesuksesan?. Jawabannya adalah “*belum tentu*”. Jangankan kesuksesan yang akan didapat, bisa jadi yang akan diperoleh adalah kehancuran. Kenapa demikian?. Jawabannya adalah mengubah profesi (khususnya profesi keilmuan) tidaklah semudah membalikan telapak tangan. Kita harus menyadari setiap manusia dimuka bumi ini memiliki keterbatasan-keterbatasan baik keterbatasan umur, keterbatasan fisik maupun keterbatasan intelegensia. Mari kita yakini bersama bahwa “*tidak ada satupun manusia dimuka bumi ini yang menguasai semua ilmu (profesi) secara sempurna*”, *apapun latar belakang pangkat, jabatan dan gelar orang tersebut*. Melalui seminar ini, kami mengajak bapak/ibu dan saudara-saudara semuanya, marilah kita mulai mencintai, mengenal dan memahami dengan baik profesi kita saat ini (Budiantara, 2011). Untuk memperoleh kesuksesan tidak mutlak disebabkan oleh karena profesi. Dengan demikian, kita berharap penelitian yang berkarakter dan mandiri, mudah-mudahan akan mampu kita raih.

Penelitian Spline Menuju Penelitian Statistika yang Mandiri dan Berkarakter

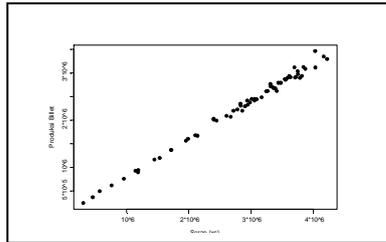
Pada bagian sebelumnya telah diuraikan penelitian yang berkarakter secara umum beserta persoalan-persoalannya. Pada bagian ini, kami mencoba memberikan uraian secara spesifik, bagaimana penelitian bidang Regresi Spline mampu menciptakan penelitian Statistika yang mandiri dan berkarakter. Penelitian bidang Regresi Spline merupakan penelitian dimana komponen utamanya adalah “Spline”. Spline merupakan salah satu bagian dari analisis regresi, khususnya regresi nonparametrik dan regresi semiparametrik. Untuk menghasilkan penelitian bidang spline yang mandiri dan berkarakter, diperlukan proses yang sangat panjang dan komprehensif, serta mengikuti tahapan-tahapan tertentu. Tahapan-tahapan tersebut diuraikan sebagai berikut.

Pemahaman Regresi Parametrik

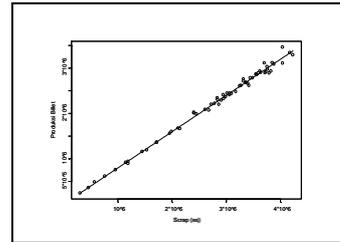
Tahap ini, dimulai dengan mempelajari dan memahami dengan sangat baik (sangat mendalam) konsep dasar dan cara berfikir filosofis dari pendekatan analisis regresi parametrik, seperti misalnya regresi linear sederhana, linear berganda, Polinomial, dan yang lainnya. Sebagai ilustrasi, jika kita memiliki data berpasangan (dengan suatu karakter tertentu) mengapa seseorang menggunakan pendekatan model regresi linear, dan kenapa bukan regresi yang lain. Sebaliknya pada karakter data yang lain, mengapa seseorang menggunakan pendekatan model regresi Polinomial, dan kenapa bukan regresi yang lain. Hal ini harus dapat dipahami dengan sangat baik dan utuh (tidak boleh setengah-setengah) oleh seorang peneliti yang ingin menggeneralisasikan regresi parametrik kedalam regresi nonparametrik dan semiparametrik, khususnya Spline. Pemahaman tentang proses inferensi dalam regresi parametrik harus pula diketahui dengan baik dan tidak setengah-setengah oleh para peneliti bidang Spline.

Diberikan sekumpulan data berpasangan (x_i, y_i) dan hubungan antara kedua variabel diasumsikan mengikuti model regresi $y_i = f(x_i) + \varepsilon_i, i = 1, 2, \dots, n$, dengan f kurva regresi dan

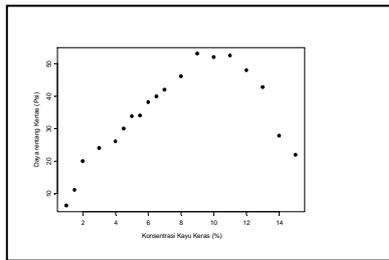
\mathcal{E}_i error random. Dalam regresi parametrik terdapat asumsi yang sangat kaku dan kuat yaitu bentuk kurva regresi diketahui, misalnya linear, kuadrat, kubik, polinomial derajat- p , eksponen, dan lain-lain. Untuk memodelkan data menggunakan regresi parametrik linear, kuadrat, kubik atau yang lain, umumnya dimulai dengan membuat *scatter* plot (Budiantara, 2006a). Apabila *scatter* plot ini terdapat kecenderungan data mengikuti pola linear maka digunakan model regresi (parametrik) linear, sebaliknya jika *scatter* plot data terdapat kecenderungan pola kuadrat maka digunakan model regresi (parametrik) kuadrat, dan seterusnya. Disamping memperhatikan pola kecenderungan data melalui *scatter* plot, kita juga dituntut dalam regresi parametrik memiliki informasi masa lalu yang detail tentang pola data agar diperoleh pemodelan yang baik (Wahba, 1990; Eubank, 1988; Antoniadis, 2001; Kayri, & Zirhhoglu, 2009; Wu & Zhang, 2006; Budiantara, 2009b).



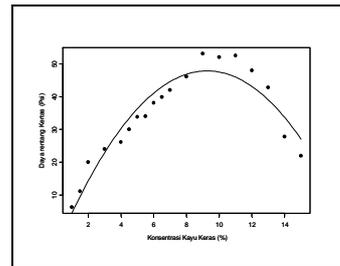
Gambar 3(a) : Plot Data Regresi Linear.



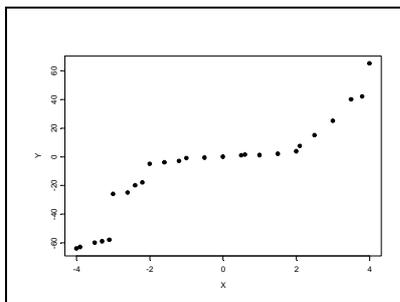
Gambar 3(b) : Estimasi Regresi Linear.



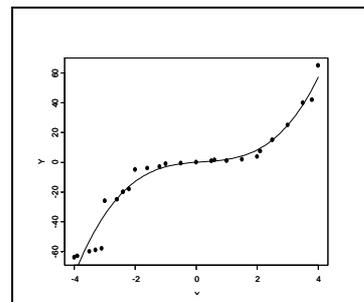
Gambar 4(a) : Plot Data Regresi Kuadrat.



Gambar 4(b) : Estimasi Regresi Kuadrat.



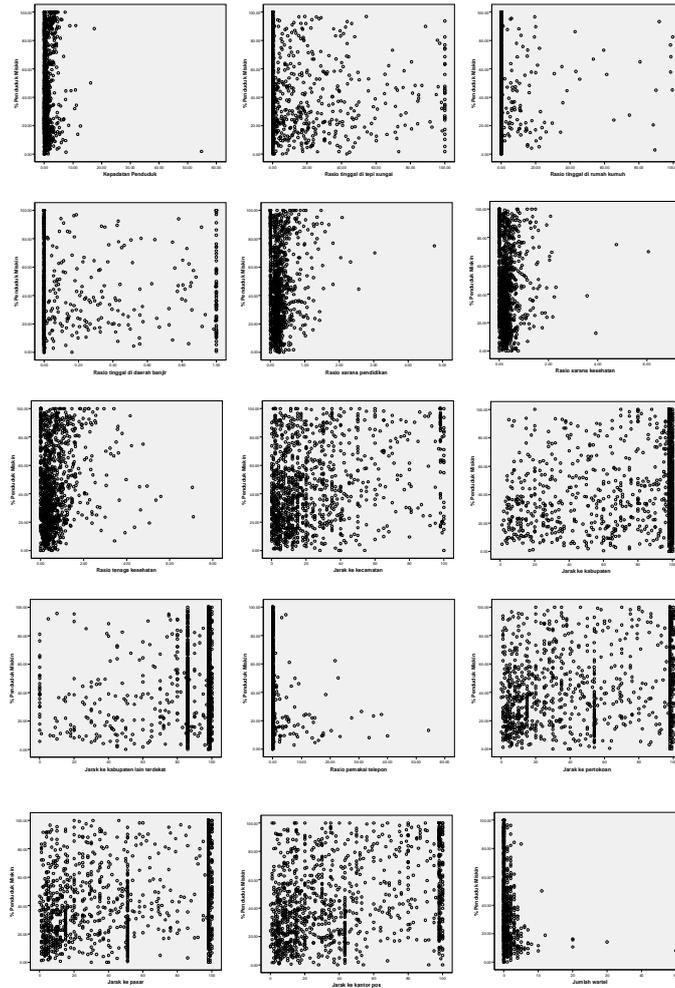
Gambar 5(a) : Plot Data Regresi Kubik.



Gambar 5(b) : Estimasi Regresi Kubik.

Sebagai ilustrasi tentang karakteristik data yang memiliki pola regresi parametrik diberikan dalam Gambar 3(a,b) (regresi parametrik linear), Gambar 4(a,b) (regresi parametrik kuadrat), dan Gambar 5(a,b) (regresi parametrik kubik). Pendekatan regresi parametrik memiliki sifat yang sangat baik dari pandangan Statistika inferensi (Budiantara, 2009b), seperti sederhana, mudah interpretasinya, parsimoni, estimatornya tidak bias, tergolong estimator linear, efisien, konsisten, BLUE (*Best Linear Unbiased Estimator*), yang sangat jarang dimiliki oleh pendekatan regresi lain seperti regresi nonparametrik dan regresi semiparametrik. Karena kebaikan (luar biasa) yang dimiliki oleh regresi parametrik inilah yang menyebabkan model regresi parametrik sangat populer dan sangat disukai oleh berbagai kalangan, baik dari golongan Statistika teoritis maupun golongan Statistika aplikasi (Becher, dkk., 2009; Huang & Liu, 2006).

Pemahaman Regresi Nonparametrik



Gambar 6. Data yang Tidak Mempunyai Pola Tertentu.

Mengingat perkembangan ilmu pengetahuan dan teknologi yang sangat pesat dan memperhatikan gejala-gejala alam yang dalam beberapa tahun terakhir mengarah pada pola yang tidak seperti biasanya (seolah-olah abnormal), maka sangat sulit bagi kita untuk menduga perilaku alam. Pada beberapa tahun yang lalu, kita masih mampu memperkirakan dengan tepat kapan mulai dan berakhirnya musim kemarau dan penghujan dengan sangat baik, sehingga para petani kita mampu mempersiapkan diri kapan memulai menanam padi dan kapan memanennya, tetapi sekarang hal tersebut seolah-olah sulit dilakukan. Beberapa tahun yang lalu kita dapat memprediksi dengantepat kapan akan terjadinya gelombang pasang, angin kencang dan lain sebagainya, sehingga para nelayan kita dapat merencanakan kapan mereka ke laut untuk mencari ikan. Tetapi sekarang hal tersebut seolah-olah sulit kita prediksi (Budiantara, 2009a; 2009b). Dalam era perubahan iklim global seperti sekarang ini, yang disebabkan oleh berbagai konflik kepentingan, baik Regional, Nasional, maupun Internasional, sangatlah mustahil jika kita masih menggunakan pendekatan standar untuk tujuan pemodelan dan prediksi yang kompleks tersebut. Diperlukan suatu metode baru, yang dapat digunakan dan diandalkan serta memberikan hasil pemodelan dan prediksi yang lebih baik.

Persoalan kemiskinan, keterbelakangan, kebodohan, dan krisis moral merupakan musuh kita bersama pada saat ini, dan mungkin musuh "abadi" kita pada masa yang akan datang. Persoalan ini, seolah-olah tidak pernah dapat diselesaikan dengan tuntas oleh Pemerintah dan kita semua. Persoalan seperti persentase penduduk miskin, beserta variabel-variabel penyebabnya merupakan salah satu contoh kejadian yang bentuk polanya tidak jelas (tidak mengikuti pola tertentu), dan seolah-olah tidak beraturan (lihat Gambar 6) (Budiantara, 2009b). Walaupun model regresi parametrik sederhana dan mudah, dalam situasi seperti ini, tidaklah bijaksana jika kita

menggunakan pendekatan regresi parametrik untuk tujuan pemodelan dan prediksi. Dalam pemodelan Statistika, memang sangat diharapkan sedapat mungkin menggunakan model yang parsimoni (sederhana), tetapi dalam keadaan dimana terdapat kondisi yang mengharuskan pemodelan menggunakan model yang lebih kompleks, maka model parsimoni tidak selayaknya dipaksakan, karena hasil yang diperoleh akan sangat bias dan memiliki *error* yang sangat besar (Budiantara, 2009a).

Dalam persoalan pemodelan, model yang baik tidaklah tunggal (tidak satu satunya). Model yang baik dapat dipandang dari berbagai aspek. Oleh karena itu, seorang pakar Statistika diharapkan memperlihatkan kearifannya, dan menghindari fanatisme bidang keilmuan yang berlebihan, serta dapat menempatkan sesuatu persoalan pemodelan tepat pada porsinya. Secara spesifik, apabila sekumpulan data dapat diselesaikan dengan baik menggunakan pendekatan regresi parametrik, ("karena fanatisme berlebihan yang melekat pada bidang keilmuannya"), maka harus diselesaikan dengan pendekatan regresi nonparametrik ataupun semiparametrik. Demikian pula sebaliknya. Walaupun hal tersebut benar secara Statistika teoritis, tetapi kurang bijaksana dilihat dari sudut pandang kekomplekan modelnya (Budiantara, 2009a; 2009b).

Berbeda dengan regresi parametrik yang tanpa disadari cenderung ada unsur pemaksaan dari peneliti dan tanpa disadari pula, peneliti ikut campur tangan dalam menentukan bentuk estimasi dari kurva regresi, maka dalam regresi nonparametrik, hal ini tidak akan terjadi. Dalam pandangan regresi nonparametrik, biarkan data sendiri yang akan mencari bentuk estimasi dari kurva regresinya, tanpa harus dipengaruhi oleh faktor subyektifitas sipeneliti (Eubank, 1988; Budiantara, 2001a). Ini berarti pendekatan model regresi nonparametrik sangatlah fleksibel dan sangat obyektif. Beberapa model regresi nonparametrik yang banyak digunakan diantaranya, Histogram, Kernel (Kayri, & Zirrhoglu, 2009; Budiantara & Mulianah, 2007), Spline (Budiantara, dkk., 1997; 2010a; 2010b; Becher, dkk., 2009; Huang dan Liu, 2006; Oehlert, 1992; Cox & O'Sullivan, 1996; Wahba, 1990; Lestari, dkk., 2010; Koenker, dkk., 1994), Polinomial Lokal (Eubank, 1988), Deret Ortogonal (Eubank, 1988), Deret Fourier (Bilaudio, 1992), k-NN (Härdle, 1990; 1991), Neural Network (NN), Wavelets (Antoniadis, 2001), MARS (Budiantara, dkk., 2006), dan yang lainnya. Semua model-model regresi nonparametrik ini mempunyai kelebihan dan kekurangan serta memiliki motivasi tersendiri dalam memodelkan pola data.

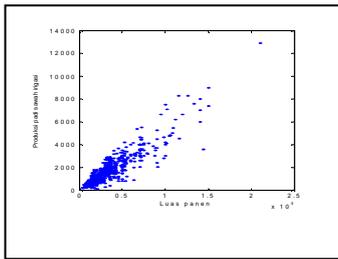
Pemahaman Regresi Semiparametrik

Disamping pendekatan regresi parametrik dan regresi nonparametrik, terdapat pula golongan Statistikawan, yang memandang kurva regresi dapat diklasifikasikan kedalam dua komponen, yaitu komponen parametrik (bentuk fungsinya diketahui) dan komponen nonparametrik (bentuk fungsinya tidak diketahui). Pandangan ini memberikan pendekatan regresi semiparametrik (Wahba, 1990; Budiantara, dkk., 2010; Budiantara, 2007; Wu & Zhang 2006). Apabila bagian parametriknya dapat dipolakan linear, maka regresi semiparametrik ini, disebut sebagai regresi linear parsial. Oleh karena itu, pada regresi semiparametrik estimasi untuk kurva regresi diperoleh secara ekuivalen dengan estimasi fungsi dan estimasi parameter dalam model. Beberapa model regresi semiparametrik yang populer adalah regresi semiparametrik Kernel, Spline, Polinomial Lokal, Deret Fourier, Wavelets, MARS (*Multivariate Addaptif Regression Spline*), dan yang lainnya.

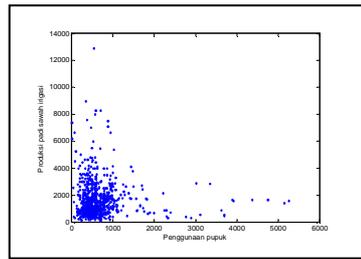
Swasembada beras merupakan cita-cita dan harapan dari pemerintah, guna memenuhi kebutuhan pangan nasional. Tetapi ironisnya hal tersebut sulit dicapai (walaupun negara ini pernah swasembada beras), padahal negara ini tercatat sebagai negara agraris. Produksi beras nasional dikuatirkan akan terancam, karena beralih fungsinya dan berkurangnya lahan pertanian dari tahun ketahun secara sistematis, untuk berbagai kepentingan seperti perumahan, industri, bisnis dan yang lainnya. Lebih dari separuh produksi beras nasional, disumbang dari produksi padi yang berasal dari padi sawah. Untuk mempertahankan dan meningkatkan produksi beras nasional, disamping memperluas lahan pertanian baru, salah satu hal yang perlu dilakukan adalah menyelidiki faktor-faktor yang mempengaruhi produksi padi sawah di Indonesia.

Pola hubungan antara besarnya produksi padi sawah dan variabel luas lahan memang berpola linear (parametrik), tetapi dengan variabel-variabel lain seperti penggunaan pupuk, penggunaan benih, tenaga kerja dan pestisida, tidak berpola linear, bahkan terlihat tidak ada pola tertentu (nonparametrik), (lihat Gambar 7(a-f)). Akibatnya, untuk memperkirakan produksi beras

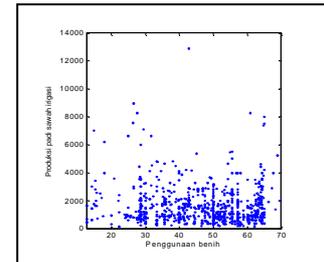
nasional pada periode waktu tertentu, selayaknya mempertimbangkan menggunakan model regresi semiparametrik sebagai salah satu alternatif (Budiantara, 2009b).



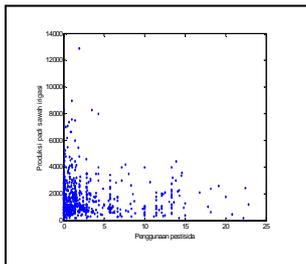
Gambar 7(a)
Pola Parametrik linear



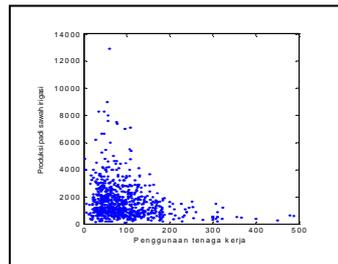
Gambar 7(b)
Pola Nonparametrik



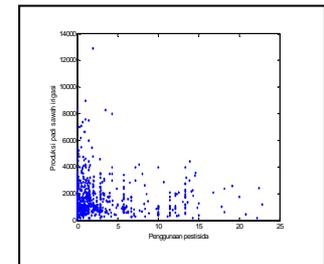
Gambar 7(c)
Pola Nonparametrik



Gambar 7(d)
Pola Nonparametrik



Gambar 7(e)
Pola Nonparametrik



Gambar 7(f)
Pola Nonparametrik

Pemahaman Tentang Spline

Diantara model-model regresi nonparametrik dan semiparametrik yang disebutkan di atas, Spline (Eubank, 1988; Budiantara, 2009a; 2009b; Budiantara dan Jerry, 2010; Wu dan Zhang, 2006) :

- Spline memiliki interpretasi Statistik dan interpretasi visual yang sangat khusus dan sangat baik. Model Spline diperoleh dari optimasi *Penalized Least Square* (PLS).
- Spline memiliki fleksibilitas yang tinggi.
- Spline mampu menangani data/fungsi yang mulus (*smooth*).
- Spline memiliki kemampuan yang sangat baik untuk menangani data yang perilakunya berubah-ubah pada sub-sub interval tertentu.
- Spline mempunyai kemampuan yang sangat baik untuk digeneralisasikan pada pemodelan Statistika yang kompleks dan rumit.

Diberikan model regresi nonparametrik $y_j = f(t_j) + \varepsilon_j$, $t_j \in [a, b]$, $j = 1, 2, \dots, n$. Bentuk kurva regresi f diasumsikan tidak diketahui dan termuat di dalam ruang *Sobolev* :

$$W_2^m[a, b] = \left\{ g ; \int_a^b (g^{(m)}(t))^2 dt < \infty \right\}.$$

Estimator Spline dalam regresi nonparametrik diperoleh dari optimasi PLS (Wahba, 1990; Eubank, 1988; Budiantara, et. al, 1997; Cox & O'Sullivan, 1996; Lestari, dkk., 2010) :

$$\text{Min}_{f \in W_2^m[a, b]} \{ R(f) + \alpha J(f) \}.$$

Kuantitas $R(f)$ dan $J(f)$ berturut-turut menyatakan *goodness of fit* dan ukuran kemulusan fungsi (*penalty*). Parameter penghalus α mengontrol antara $R(f)$ dan $J(f)$. Penelitian-penelitian tentang Spline dalam regresi nonparametrik dan semiparametrik ternyata diperoleh dengan cara mengambil variasi-variasi pada $R(f)$ dan $J(f)$. Wahba (1990) dan kawan-kawan dari *Wisconsin University* mengawali pengembangan Spline dengan mengambil $R(f)$ dan $J(f)$ dalam bentuk kuadrat dan menghasilkan Spline *Natural (original)*. Estimator Spline ini direkomendasikan untuk digunakan pada data lokal dan *smooth*. Cox & O'Sullivan (1996) mengikuti langkah Wahba ini

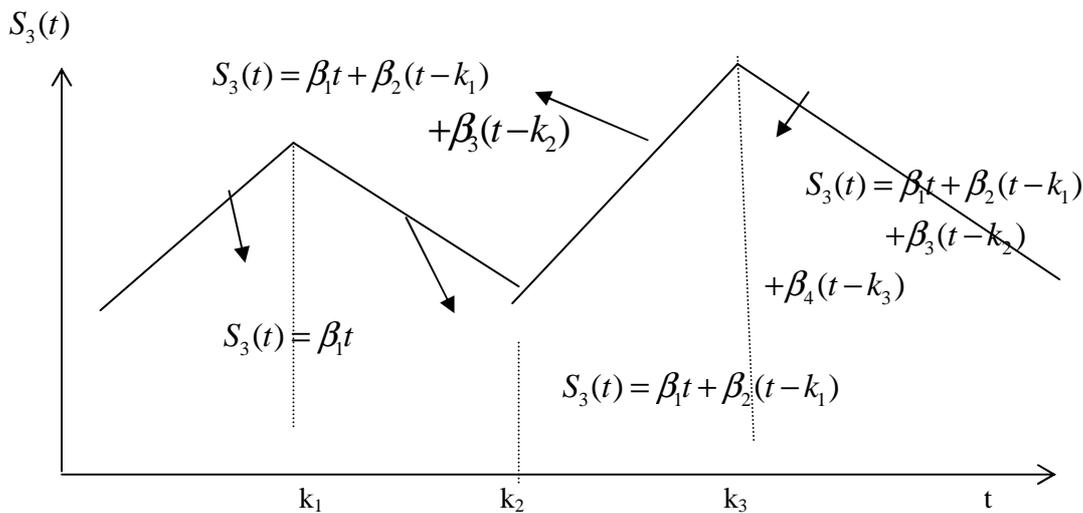
dan memperoleh estimator Spline tipe- M , dengan mengambil $R(f)$ merupakan fungsi *konveks* dan $J(f)$ dari bentuk kuadrat. Spline tipe- M digunakan untuk menangani *outlier* dalam regresi nonparametrik. Oehlert (1992) memberikan Spline *relaxed* dengan memberi bobot pada *penalty* $J(f)$ dan $R(f)$ berbentuk kuadrat. Pada sisi lain Koenker, dkk. (1994) memberikan Spline *quantile*, dengan memasang $J(f)$ dalam bentuk absolut dan $R(f)$ bentuk kuadrat.

Salah satu model Spline yang disukai dalam teori dan aplikasi adalah polinomial Spline *truncated*. Sebagai salah satu ilustrasi seserhana diberikan Spline linear *truncated* dengan tiga knots pada $t = k_1 \leq t = k_2 \leq t = k_3$ diberikan oleh :

$$S_3(t) = \beta_1 t + \beta_2 (t - k_1)_+^1 + \beta_3 (t - k_2)_+^1 + \beta_4 (t - k_3)_+^1.$$

Fungsi Spline $S_3(t)$ dapat pula disajikan dalam bentuk (lihat Gambar 10) :

$$S_3(t) = \begin{cases} \beta_1 t & , t < k_1 \\ \beta_1 t + \beta_2 (t - k_1) & , k_1 \leq t < k_2 \\ \beta_1 t + \beta_2 (t - k_1) + \beta_3 (t - k_2) & , k_2 \leq t < k_3 \\ \beta_1 t + \beta_2 (t - k_1) + \beta_3 (t - k_2) + \beta_4 (t - k_3) & , t \geq k_3 \end{cases}$$



Gambar 8. Fungsi Spline $S_3(t)$.

Fungsi $S_3(t)$ memberikan sifat berikut (Budiantara, 2007a):

- a. $S_3(t)$ merupakan fungsi kontinu pada setiap t , termasuk pada titik-titik knots :

$$t = k_1, t = k_2 \text{ dan } t = k_3.$$

- b. $S_3(t)$ merupakan potongan polinomial linear pada sub-sub interval :

$$t < k_1, k_1 \leq t < k_2, k_2 \leq t < k_3, k_3 \leq t.$$

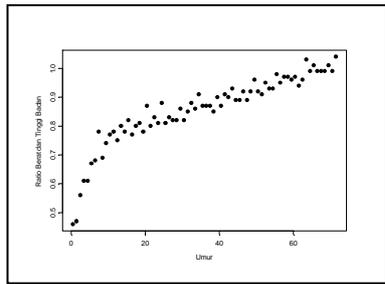
- c. Jika $S_3(t)$ diderivatikan terhadap t diperoleh fungsi tangga berikut :

$$\frac{d}{dt} S_3(t) = \begin{cases} \beta_1, & t < k_1 \\ \beta_1 + \beta_2, & k_1 \leq t < k_2 \\ \beta_1 + \beta_2 + \beta_3, & k_2 \leq t < k_3 \\ \beta_1 + \beta_2 + \beta_3 + \beta_4, & t \geq k_3 \end{cases}$$

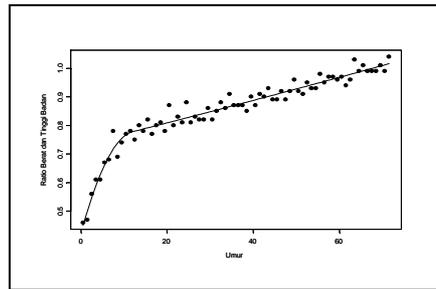
Berikut ini diberikan beberapa penggunaan Spline dalam pemodelan Statistika.

Data pertumbuhan balita cenderung memiliki perilaku yang khas. Eubank (1988) memodelkan umur balita dikaitkan dengan ratio berat dan tinggi badan negara bagian di Amerika Serikat dengan menggunakan Spline. Plot data dan estimasi pola data disajikan berturut-turut dalam Gambar 9(a) dan Gambar 9(b). Pola pertumbuhan balita yang diberikan oleh Eubank seperti pada Gambar 9(a,b), mengidentifikasi bahwa pertumbuhan balita sangat cepat sejak balita

dilahirkan sampai berumur 11,5 bulan, dan berpola kuadratik. Tetapi setelah balita berumur lebih dari 11,5 bulan, pertumbuhannya tetap naik tetapi secara perlahan-lahan, dan berpola linear.



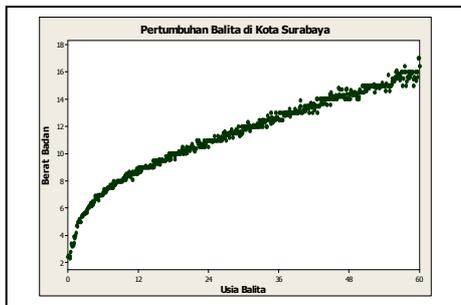
Gambar 9(a). Plot umur dan ratio.



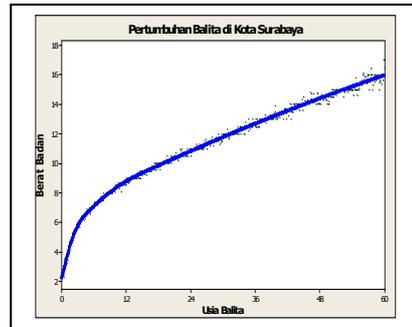
Gambar 9(b). Model Spline.

Generalisasi Spline (Spline terbobot)

Pola pertumbuhan balita seperti yang telah dikembangkan oleh Eubank di Amerika Serikat, mirip dengan data pertumbuhan balita yang ada di Kota Surabaya. Budiantara & Ratna (2009) dan Budiantara & Jerry (2010a; 2010b; 2011) memodelkan pertumbuhan balita di kota Surabaya yang variasinya meningkat seiring dengan meningkatnya umur balita dengan menggunakan Spline terbobot (Gambar 10(a,b)).

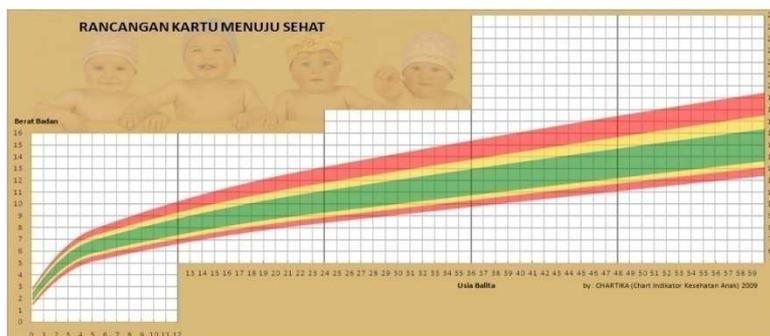


Gambar 10(a). Plot Pertumbuhan Balita di Surabaya.

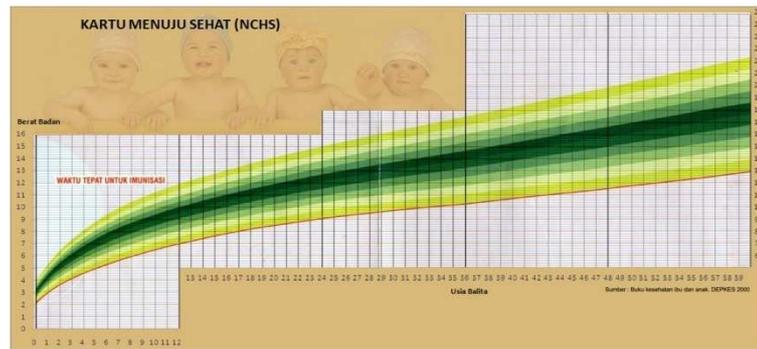


Gambar 10(b). Spline Terbobot Untuk Estimasi Pertumbuhan Balita di Surabaya.

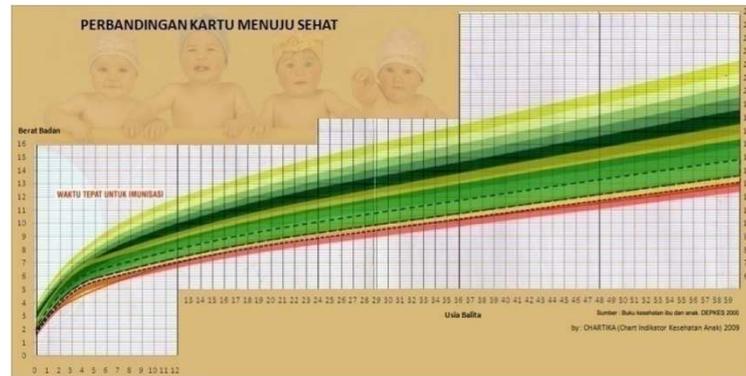
Spline yang telah dihasilkan oleh para peneliti dalam regresi nonparametrik dan semiparametrik, dapat dikembangkan dengan memberi kombinasi pada komponen $R(f)$ dan $J(f)$. Budiantara, dkk. (1997), memasang bobot pada *goodness of fit* $R(f)$ dan mempertahankan $J(f)$ berbentuk kuadrat, dalam optimasi PLS. Dengan menggunakan pendekatan *Reproducing Kernel Hilbert Space* (RKHS), Budiantara, dkk. (1997), memberikan penyelesaian optimasi PLS adalah Spline terbobot. Spline terbobot yang diperoleh ini bermanfaat untuk menangani ketaksamaan variansi (*heteroskedastik*) dalam analisis regresi nonparametrik.



Gambar 11. Rancangan Kartu Menuju Sehat (KMS) Balita Kota Surabaya.



Gambar 12. Kartu Menuju Sehat (KMS) yang Dikeluarkan Oleh NCHS Standar WHO.



Gambar 13. Perbandingan KMS yang Dikeluarkan Oleh NCHS dan Rancangan KMS Kota Surabaya.

Salah satu kegiatan yang dilakukan oleh Departemen Kesehatan RI dalam pemantauan pertumbuhan balita di Indonesia adalah dengan mencatat umur dan berat badan balita, dan menuliskannya kedalam sebuah kartu yang disebut dengan Kartu Menuju Sehat (KMS). Cara pemantauan pertumbuhan balita seperti ini dilakukan pada Rumah sakit, Posyandu, Klinik bersalin, Puskesmas di seluruh Indonesia. Dengan demikian, pertumbuhan balita dapat dipantau dari perkembangan berat badan balita yang ada pada KMS, apakah normal atau tidak. KMS yang digunakan di Indonesia saat ini adalah KMS standar *World Health Organization* (WHO) (lihat Gambar 14) yang dikeluarkan oleh *National Center for Health Statistics* (NCHS) Amerika Serikat. KMS standar WHO ini tidak menggunakan data pertumbuhan balita yang ada di Indonesia, sehingga KMS ini cenderung tidak menggambarkan perilaku pertumbuhan balita yang ada di Indonesia. Akibatnya, terdapat balita yang semestinya sehat, tetapi terdeteksi tidak sehat dalam KMS, dan sebaliknya.

Budiantara & Ratna (2009) dan Budiantara & Jerry (2010a; 2010b) merancang KMS untuk balita di kota Surabaya dimana Kurva pertumbuhan pada KMS diperoleh dengan menduga pola hubungan antara umur dan berat badan balita, menggunakan Spline terbobot. KMS yang diperoleh akan diusulkan ke-DEPKES RI untuk digunakan di Rumah sakit, Klinik bersalin, Puskesmas dan Posyandu di Surabaya (lihat Gambar 11). Jika dibandingkan, ternyata KMS balita di kota Surabaya memiliki berat badan yang lebih kecil dari KMS yang digunakan di Indonesia saat ini (termasuk digunakan di kota Surabaya) yang dikeluarkan oleh WHO, baik untuk semua warna dalam KMS (warna merah bawah, warna kuning bawah, warna hijau, warna kuning atas, warna merah atas), maupun semua umur dari 0 bulan sampai 60 bulan (Lihat Gambar 13).

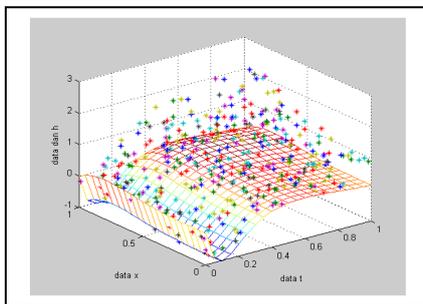
Generalisasi Spline (Regresi Spline Terbobot)

Estimator *smoothing spline* yang telah diperoleh oleh banyak peneliti, diperoleh dari optimasi *Penalized Least Square* (PLS) / *Penalized Likelihood* (PL) / *Penalized Log Likelihood* (PLL), dan penyelesaiannya menggunakan pendekatan *Reproducing Kernel Hilbert Space* (RKHS). Penelitian estimator *smoothing spline* ini, yang sangat teoritis dan sangat matematis, sehingga masih sulit untuk diaplikasikan untuk menyelesaikan persoalan praktis (real) dalam model regresi nonparametrik. Disamping itu, untuk dapat memahami dan menggunakan estimator *smoothing spline* yang diberikan oleh banyak peneliti, para pengguna dituntut memiliki

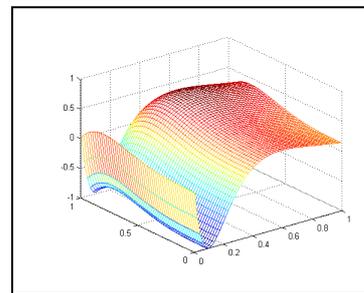
pengetahuan matematik yang relatif tinggi dan rumit, sehingga penggunaan estimator *smoothing spline* terbobot akan menjadi sangat terbatas hanya pada golongan (profesi) tertentu saja, padahal model spline terbobot sangat luas aplikasinya.

Berdasarkan latar belakang ini, kita dapat mencari suatu alternatif lain untuk estimasi kurva regresi nonparametrik. Budiantara (2006) telah memberikan pendekatan regresi spline yang tidak tergolong kedalam kelas PLS, PL, atau PLL, dan tidak juga menggunakan pendekatan RKHS, tetapi menggunakan kelas estimator yang sudah sangat dikenal dan sangat familier dalam Statistika inferensi, yaitu dari kelas optimasi *Weighted Least Square* (WLS), *Weighted Likelihood* (WL), atau *Weighted Log Likelihood* (WLL). Disamping itu, penyelesaian WLS/WL/WLL menggunakan derivatif parsial biasa. Dengan demikian, dihasilkan sebuah model spline terbobot yang relatif sederhana dan mudah secara matematik dan dapat digunakan oleh semua kalangan profesi untuk menyelesaikan persoalan pemodelan dalam regresi nonparametrik dan semiparametrik.

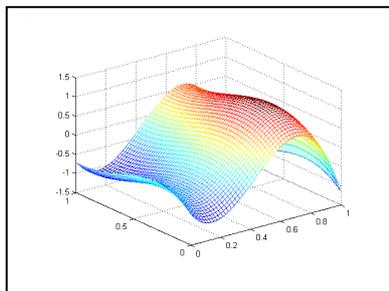
Generalisasi Spline (Spline Multivariabel)



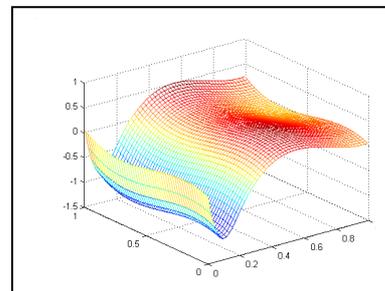
Gambar 14. Plot Data Tiga Dimensi.



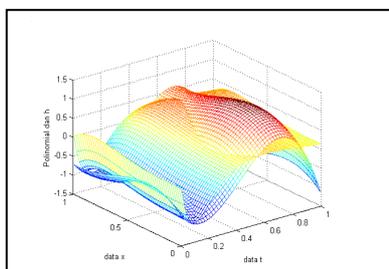
Gambar 15. Model Data Tiga Dimensi (Sebenarnya).



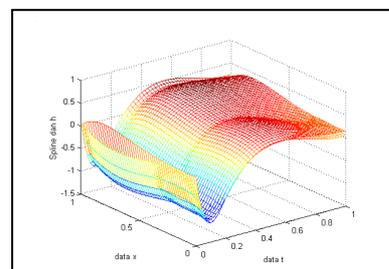
Gambar 16. Estimasi Polinomial Terbobot Tiga Dimensi.



Gambar 17. Estimasi Spline Terbobot Data Tiga Dimensi.



Gambar 18. Polinomial Terbobot dengan Kurva Regresi Tiga Dimensi.



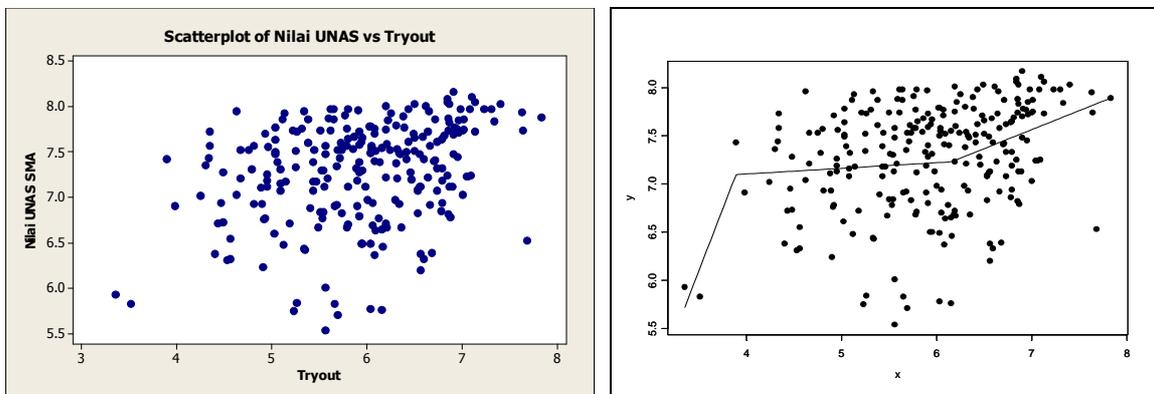
Gambar 19. Spline Terbobot dengan Kurva Regresi Tiga Dimensi.

Salah satu kelebihan pendekatan Spline adalah model ini cenderung mencari sendiri estimasi data kemanapun pola data tersebut bergerak. Kelebihan ini terjadi karena dalam Spline terdapat titik-titik knot (Eubank, 1988), yaitu titik perpaduan bersama yang menunjukkan terjadinya perubahan pola perilaku data. Sebagai ilustrasi numerik, Budiantara, dkk. (1997) merancang suatu studi

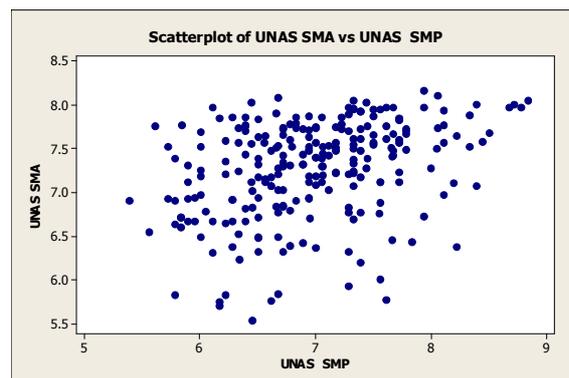
simulasi untuk memperlihatkan kemampuan Spline terbobot dalam mengestimasi pola data tiga dimensi, dan selanjutnya membuat suatu perbandingan visual tiga dimensi, dengan pendekatan polinomial.

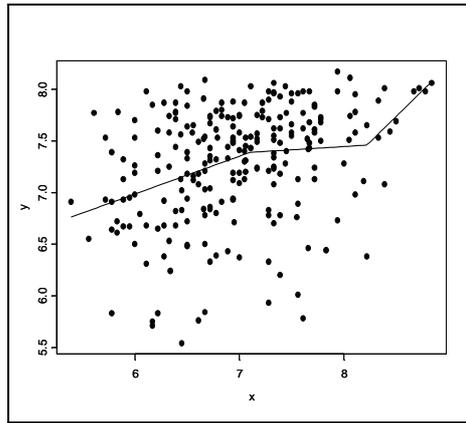
Plot pola data dan kurva regresi tiga dimensi yang sebenarnya, berturut-turut diberikan dalam Gambar 14 dan Gambar 15. Estimasi model polinomial kubik terbobot tiga dimensi untuk kurva regresi disajikan dalam Gambar 16. Sedangkan estimasi kurva regresi tiga dimensi dengan Spline kubik terbobot diberikan dalam Gambar 17. Gambar tiga dimensi antara estimasi Polinomial kubik terbobot bersama-sama dengan kurva regresi yang sebenarnya disajikan dalam Gambar 18. Selanjutnya, estimasi pola data dengan Spline kubik terbobot tiga dimensi bersama-sama dengan kurva regresi yang sebenarnya diberikan dalam Gambar 19. Terlihat dari gambaran visual tiga dimensi ini bahwa pendekatan model Spline terbobot cenderung mempunyai *error* estimasi yang relatif lebih kecil, jika dibandingkan dengan pendekatan polinomial kubik terbobot. Hal ini terlihat jelas dari gambaran visual tiga dimensi dari model Spline terbobot yang terlihat melekat (menyatu) dengan kurva regresi tiga dimensi yang sebenarnya. Sehingga dalam kurva pada Gambar 19 yang merupakan gambar dari dua kurva tetapi terlihat hanya satu kurva saja. Hasil ini, sangat berbeda dengan Gambar 18 dimana secara visual terlihat bahwa pendekatan polinomial kubik terbobot mempunyai *error* estimasi yang relatif besar.

Pola hubungan antara rata-rata nilai UNAS SMAN 11 Ambon dipengaruhi oleh Ambon tahun pelajaran 2007/2008. Variabel yang mempengaruhi rata-rata nilai UNAS adalah rata-rata nilai Tryout, Rata-rata nilai UNAS SMP, Rata-rata nilai UAS, Rata-rata nilai rapor, Pendapatan orang tua, dan jarak rumah ke sekolah (Henaulu,2009). Plot antara rata-rata nilai UNAS dengan rata-rata nilai tryout, Rata-rata nilai UNAS SMP, Rata-rata nilai UAS, Rata-rata nilai rapor, Pendapatan orang tua, dan jarak rumah ke sekolah, berturut-turut diberikan oleh Gambar 20(a), Gambar 21(a), Gambar 22(a), Gambar 23(a), Gambar 24(a) dan Gambar 25(a). Terlihat dari Plot-plot tersebut, polanya tidak menunjukkan pola tertentu. Oleh karena itu pola rata-rata nilai UNAS dimodelkan dengan Spline, seperti terlihat dalam Gambar 20(b), Gambar 21(b), Gambar 22(b), Gambar 23(b), Gambar 24(a) dan Gambar 25(b).

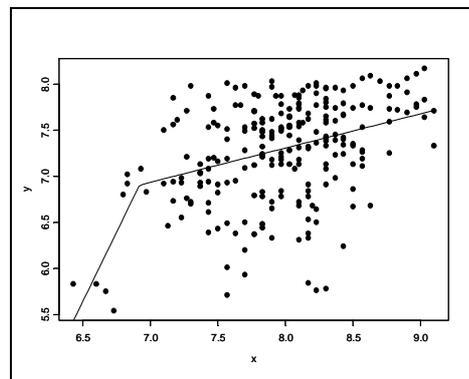
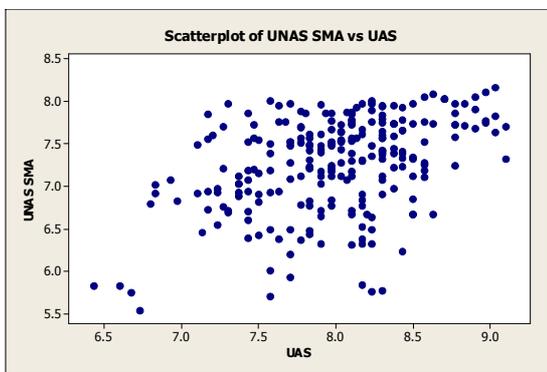


Gambar 20(a,b). Plot dan Estimasi nilai UNAS SMA dengan nilai *tryout* Untuk spline linear dua titik knot

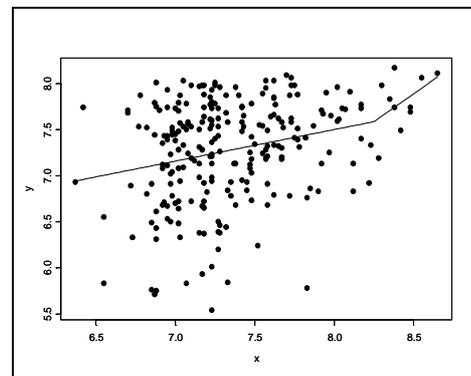
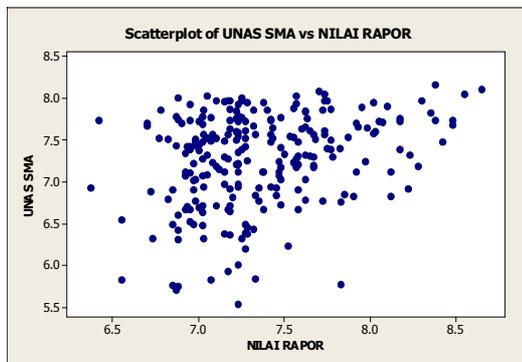




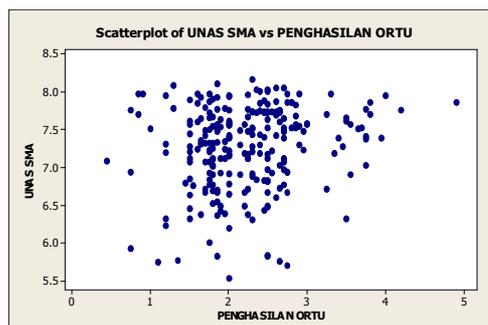
Gambar 21(a,b). Plot dan Estimasi nilai UNAS SMA dengan nilai UNAS SMP untuk spline linear dua titik knot

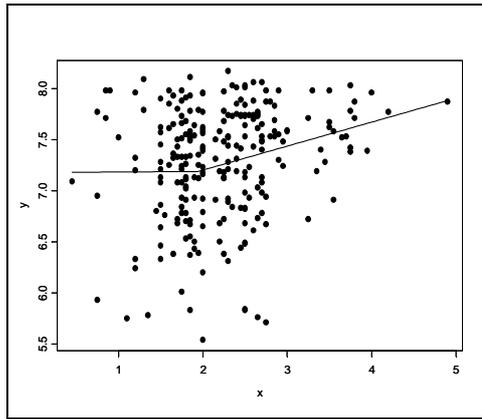


Gambar 22(a,b). Plot dan Estimasi nilai UNAS SMA dengan nilai UAS untuk spline linear satu titik knot

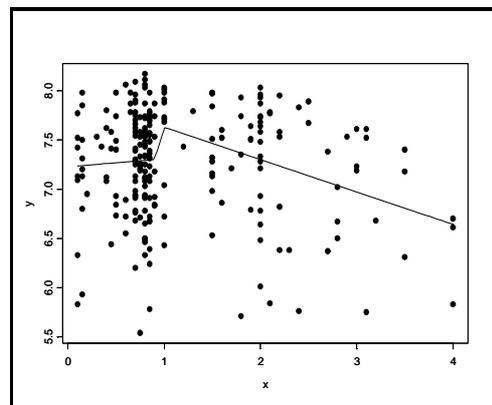
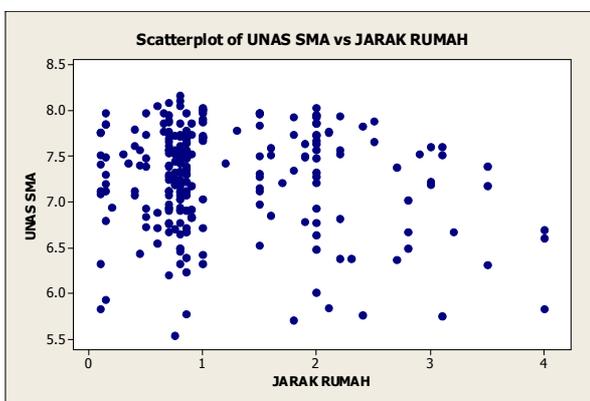


Gambar 23(a,b). Plot dan Estimasi nilai UNAS SMA dengan nilai Rapor Untuk spline linear satu titik knot.





Gambar 24(a,b). Plot dan Estimasi nilai UNAS SMA dengan Penghasilan Orang Tua untuk spline linear satu titik knot.



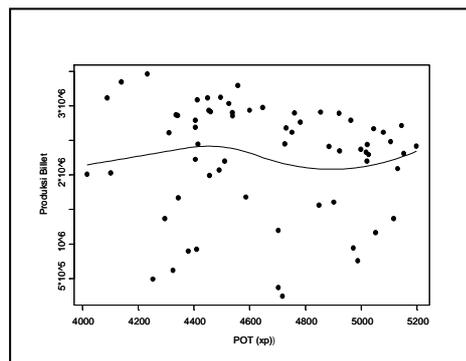
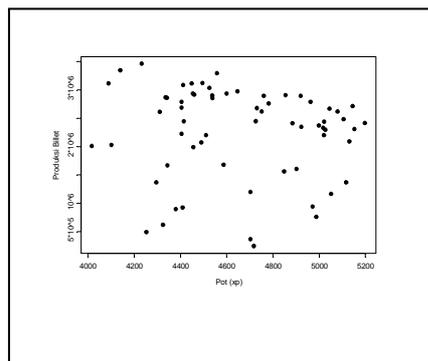
Gambar 25(a,b). Plot dan Estimasi nilai UNAS SMA dengan Jarak Rumah untuk spline linear dua titik knot.

Generalisasi Spline (Spline Parsial dalam Regresi Semiparametrik)

Disamping tipe-tipe pola data seperti di atas, kita sangat sering menemui tipe pola data yang merupakan campuran antara parametrik dan nonparametrik, seperti ilustrasi berikut (Budiantara, 2006a; 2006b; 2007b). Diberikan data tentang produksi Billet (dalam kg) merupakan produk besi batangan dari suatu perusahaan besi baja (y). Produksi Billet ini dipengaruhi oleh beberapa variabel, diantaranya HEAT (x_h) proses tertentu dalam produksi Billet, SCRAP (x_s) banyak bahan baku (dalam kg) untuk produksi Billet dan POT (x_p) lama waktu (dalam detik) pengerjaan Billet dengan menggunakan listrik. Ingin diestimasi model pola hubungan antara produksi Billet dengan variabel HEAT, SCRAP dan POT. Diambil sampel random dari hasil produksi Billet dan hubungan keempat variabel diberikan oleh model :

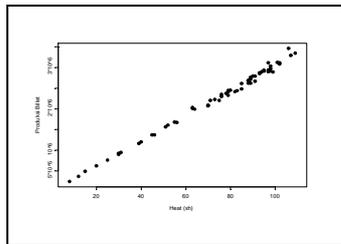
$$y_i = f(x_{hi}, x_{si}, x_{pi}) + \epsilon_i.$$

Plot antara variabel y_i dengan x_{pi} disajikan dalam Gambar 26(a).

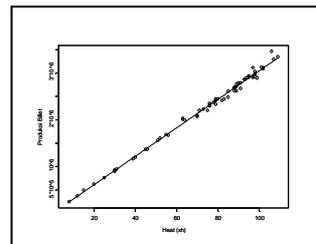


Gambar 26(a) : Plot antara y dengan x_p . Gambar 26(b) : Spline parsial dua knot.

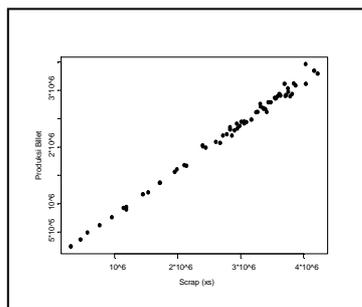
Sedangkan plot antara variabel y_i dengan x_{hi} dan variabel y_i dengan x_{si} , masing-masing disajikan dalam Gambar 27(a) dan Gambar 28(a). Telihat dari Gambar 27(a) dan Gambar 28(a) bahwa hubungan antara y dengan x_h cenderung linear. Demikian pula dengan hubungan antara y dengan x_s juga cenderung linear. Berbeda dengan Gambar 26(a), data memperlihatkan tidak adanya kecenderungan pola hubungan yang jelas antara y dengan x_p , sehingga sulit untuk dimodelkan menggunakan pendekatan parametrik.



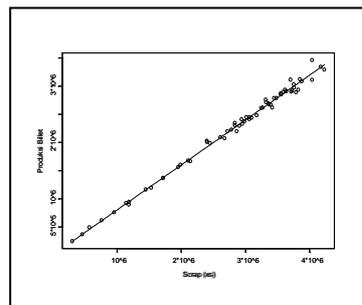
Gambar 27(a). Plot y dengan x_h .



Gambar 27(b). Regresi parametrik linear y dengan x_h .



Gambar 28(a). Plot y dengan x_s .



Gambar 28(b). Regresi parametrik linear y dengan x_s .

Berdasarkan plot ini, Budiantara (2006a; 2006b; 2007a) membuat model hubungan antara variabel respon y_i dengan x_{pi} , x_{hi} , x_{si} , menggunakan model semiparametrik, dimana variabel y_i dengan x_{hi} dan variabel y_i dengan x_{si} berhubungan parametrik linear, serta variabel y_i dengan x_{pi} berpola nonparametrik, menggunakan fungsi Spline. Kurva regresi semiparametrik dinyatakan sebagai :

$$f(x_{hi}, x_{si}, x_{pi}) = \beta_1 x_{hi} + \beta_2 x_{si} + g(x_{pi}).$$

Kurva g didekati dengan model spline kuadrat dengan 2(dua) titik knot K_1 dan K_2 :

$$g(x_p) = \sum_{j=1}^2 \theta_j x_p^j + \sum_{k=1}^2 \phi_k (x_p - K_k)^2 I(x_p \geq K_k).$$

Estimasi Spline parsial untuk kurva regresi g diberikan dalam Gambar 26(b). Sedangkan estimasi untuk komponen parametrik dalam regresi semiparametrik ini, berturut-turut disajikan dalam Gambar 27(b) dan Gambar 28(b) yang berpola linear.

Generalisasi Spline (Spline Terbobot dalam Regresi Nonparametrik dan Semiparametrik Multirespon)

Model spline dalam regresi nonparametrik dan model spline parsial dalam regresi semiparametrik yang dikembangkan oleh peneliti-peneliti, hanya untuk model regresi dengan satu variabel respon (respon tunggal). Dalam beberapa tahun terakhir, pengembangan model Spline untuk regresi nonparametrik multirespon telah dilakukan oleh Wang, dkk. (2000) tetapi dalam model regresi nonparametrik 2 (dua) respon. Holmes & Mallick (2001) mengembangkan pendekatan Bayesian untuk model spline linear, tetapi dalam regresi nonparametrik multirespon.

Hasil penelitian yang serupa juga diberikan oleh Gu (2000) yang mengembangkan model spline klasik dalam regresi nonparametrik multirespon. Model spline dan model spline parsial yang dikembangkan oleh para peneliti-peneliti di atas, membuat suatu asumsi yang sangat berat, yaitu variansi dari *error* random modelnya sama (*homogen*).

Dalam banyak kasus pada dunia real, sering ditemui kasus-kasus dimana pengukuran variabel dilakukan pada waktu yang bersamaan, sehingga akan melibatkan model regresi dengan variabel respon lebih dari satu dan antar variabel respon saling berkorelasi. Disamping itu, kesamaan variansi dari *error* random umumnya hanya ada secara teoritis, tetapi sulit diperoleh dalam dunia nyata. Akibatnya, persoalan regresi nonparametrik dan semiparametrik harus diselesaikan dengan model regresi multirespon dengan variansi tidak sama (*heteroskedastik*). Budiantara dan Ratna (2009) dan Lestari, dkk. (2010) memberikan model spline terbobot untuk menduga kurva regresi nonparametrik birespon :

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} f_1(t_1) \\ f_2(t_2) \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}$$

Estimator spline terbobot diperoleh dari optimasi *Weighted Penalized Least Square* (WPLS) :

$$\underset{f_k \in W_2^m[a_k, b_k], k=1,2}{\text{Min}} \left\{ \begin{aligned} & (n_1 + n_2)^{-1} (\underline{y} - \underline{f})' W_{NS}^T (\sigma_1^2, \sigma_2^2) (\underline{y} - \underline{f}) + \\ & + \alpha_1 \int_{a_1}^{b_1} [f_1^{(m)}(t_1)]^2 dt_1 + \alpha_2 \int_{a_2}^{b_2} [f_2^{(m)}(t_2)]^2 dt_2 \end{aligned} \right\}.$$

Budiantara & Ratna (2011a) memberikan estimator spline terbobot parsial untuk menduga kurva regresi semiparametrik birespon. Estimator Spline terbobot parsial diperoleh dari model regresi semiparametrik heteroskedastik dan menyelesaikan optimasi WPLS. Budiantara & Ratna (2011b) memberikan juga memberikan estimator Spline terbobot untuk menduga kurva regresi nonparametrik Multirespon.

Generalisasi Spline (Spline Terbobot dalam Regresi Nonparametrik dan Semiparametrik untuk Data Longitudinal)

Estimator spline dalam regresi nonparametrik dan estimator spline parsial dalam regresi semiparametrik yang dikembangkan oleh banyak peneliti hanya untuk data *cross-section*. Estimator spline dan spline parsial secara teoritis dan praktis tidak hanya mampu menangani data *cross section* saja, tetapi dapat digeneralisasikan untuk data longitudinal (Wu & Zhang, 2006). Dalam bidang kesehatan, terdapat banyak kasus yang merupakan persoalan data longitudinal (Demidenko, 2004; Diggle, dkk., 2002). Beberapa penelitian untuk data longitudinal menggunakan pendekatan regresi nonparametrik, telah banyak dipublikasikan diantaranya Wang (2003), Wu dan Chiang (2000) yang menggunakan estimator Kernel. Disamping itu, beberapa peneliti telah pula menggunakan estimator Spline untuk estimasi kurva regresi untuk data longitudinal, seperti Huang, Wu & Zhang (2006) yang menggunakan model Spline original, dan Zhang (1997) menggunakan regresi spline *generalized* (umum). Rice & Wu (2001), Wu & Zhang (2006) menggunakan pendekatan spline yang didasarkan pada suatu model efek campuran untuk estimasi kurva regresi untuk data longitudinal. Pada sisi lain, terdapat beberapa penelitian tentang data longitudinal yang didasarkan pada model semiparametrik, seperti diantaranya Fan & Zhang (2000), Guo (2002), dan Durban, dkk. (2005).

Model-model pendekatan regresi nonparametrik dan semiparametrik untuk data longitudinal yang dikembangkan oleh peneliti-peneliti di atas, jika ditelusuri secara mendalam, pada dasarnya terdapat dua asumsi yang sangat berat dan sangat mendasar pada modelnya, yaitu pertama variansi dari *error* random dalam model tersebut diasumsikan sama (*homogen*) untuk setiap pengukuran berulang didalam subjek. Asumsi kedua adalah matriks *variance-covariance* dari *error* random dalam model tersebut diasumsikan diketahui. Dua asumsi yang digunakan dalam model untuk data longitudinal ini, pada dasarnya hanya ada secara teoritis, dan dalam persoalan aplikasi sering dijumpai kasus-kasus dimana terjadi ketaksamaan variansi (*heteroskedastik*) dari *error* random dalam model tersebut. Demikian pula matriks *variance-covariance* dari *error* random dari model populasi, umumnya (hampir pasti) tidak diketahui. Berdasarkan pada realita ini, maka persoalan estimasi kurva regresi dalam regresi nonparametrik

spline untuk data longitudinal, harus diselesaikan dengan model regresi nonparametrik spline terbobot yang heteroskedastik.

Budiantara, dkk. (2009) memberikan estimator spline terbobot dalam regresi nonparametrik heteroskedastik untuk data longitudinal beserta permasalahan yang terkait, dimana untuk mendapatkan estimator kurva regresi menggunakan optimasi *Penalized Least Square* Terbobot. Sedangkan Budiantara, dkk. (2010) memberikan estimator spline terbobot parsial dalam regresi semiparametrik heteroskedastik untuk data longitudinal beserta permasalahan yang terkait dengan estimator tersebut. Model regresi spline terbobot untuk data longitudinal telah pula diberikan oleh Budiantara dan Ratna (2011).

DAFTAR RUJUKAN

- Antoniadis, A., Bigot, J. and Spatinas, T., 2001. Wavelet Estimators in Nonparametric Regression : A Comparative Simulation Study, *Journal of Statistical Software*, 6, 1-83.
- Bain, L. J., dan Engelhardt, M. 1992. *Introduction to Probability and Mathematical Statistics*, Duxbury Press, California.
- Becher, H., Kauermann, G., Khomski, P., dan Kouyate, B. 2009. Using Penalized Splines to Model Age and Season of Birth Dependent Effects of Childhood Mortality Risk Faktors in Rural Burkina Faso, *Biometrical Journal*, 51, 110-122.
- Budiantara, I. N., Subanar, dan Soejoeti, Z. 1997. Weighted Spline Estimator, *Bulletin of the International Statistical Insitute*, 51, 333-334.
- Budiantara, I. N. 2000a. Metode U, GML, CV dan GCV Dalam Regresi Nonparametrik Spline, *Majalah Ilmiah Himpunan Matematika Indonesia (MIHMI)*, 6, 41-45.
- Budiantara, I. N. 2000b. Optimasi dan Proyeksi Dalam Regresi Nonparametrik Spline, *Majalah Berkala Matematika dan Ilmu Pengetahuan Alam (BMIPA)*, Universitas Gadjah Mada, 10, 35-44.
- Budiantara, I. N. 2000c. Interval Konfidensi Bayes Untuk Kurva Regresi Nonparametrik Spline Terbobot, *Majalah Ilmiah Himpunan Matematika Indonesia (MIHMI)*, 6, 285-290.
- Budiantara, I. N. 2001a. *Regresi Nonparametrik dan Semiparametrik Serta Perkembangannya*, Makalah Pembicara Utama pada Seminar Nasional Alumni Pasca Sarjana Matematika Universitas Gadjah Mada, Yogyakarta.
- Budiantara, I. N. 2001b. *Estimasi Parametrik dan Nonparametrik untuk Pendekatan Kurva Regresi*, Makalah Pembicara Utama pada Seminar Nasional Statistika V, Jurusan Statistika, FMIPA, Institut Teknologi Sepuluh Nopember (ITS), Surabaya.
- Budiantara, I. N. 2004a. *Spline: Historis, Motivasi, dan Perannya Dalam Regresi Nonparametrik*, Makalah Pembicara Utama pada Konferensi Nasional Matematika XII, Jurusan Matematika, MIPA, Universitas Udayana (UNUD), Denpasar.
- Budiantara, I. N. 2004b. *Probabilitas Untuk Program S-2 Statistika*, Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, ITS, Surabaya.
- Budiantara, I. N. 2006a. *Regresi Nonparametrik Dalam Statistika*, Makalah Pembicara Utama pada Seminar Nasional Matematika, Jurusan Matematika, FMIPA, Universitas Negeri Makasar (UNM), Makasar.
- Budiantara, I. N. 2006b. Model Spline Dengan Knots Optimal, *Jurnal Ilmu Dasar, FMIPA Universitas Jember*, 7, 77-85.
- Budiantara, I. N., Suryadi, F., Otok, B. dan Guritno, S. 2006. Pemodelan B-Spline dan MARS pada Nilai Ujian Masuk Terhadap IPK Mahasiswa Jurusan Disain Komunikasi UK Petra, Surabaya, *Jurnal Teknik Industri*, 8, 1-13.
- Budiantara, I N., 2007a, Model Keluarga Spline Polinomial Truncated Dalam Regresi Semiparametrik, *Jurnal Matematika, Universitas Negeri Malang (UM), Malang*.
- Budiantara, I N. 2007b. Inferensi Statistik Untuk Model Spline, *Jurnal MAT STAT Universitas Bina Nusantara, Jakarta*.
- Budiantara, I N., dan Mulianah. 2007. Pemilihan Bandwith Optimal Dalam Regresi Semiparametrik Kernel dan Aplikasinya, *SIGMA : Jurnal Sains dan Teknologi, Universitas Sanata Dharma, Yogyakarta*.
- Budiantara, I. N., dan Ratna, M. 2008. *Model Kartu Menuju Sehat (KMS) Balita Kota Surabaya Dengan Pendekatan Spline Quantile*, Laporan Penelitian Hibah A-3, ITS, Surabaya.

- Budiantara, I N. 2009^a. “Meraih Kesuksesan Dengan Statistika”, Makalah Pembicara Utama pada Seminar Nasional Matematika, Jurusan Matematika, STKIP PGRI Jombang, Jawa Timur, 13 Februari 2009.
- Budiantara, I. N. 2009^b. *Spline Dalam Regresi Nonparametrik dan Semiparametrik : Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*, Pidato Pengukuhan Guru Besar Pada Jurusan Statistika, FMIPA-ITS, Surabaya, 25 April 2009.
- Budiantara, I. N., Lestari, B., dan Islamiyati, A. 2009^a. *Estimator Spline Terbobot dan Spline Parsial Terbobot Dalam Regresi Nonparametrik dan Semiparametrik Heteroskedastik Untuk Data Logitudinal*, Laporan Penelitian Hibah Kompetensi Tahun I, DP2M DIKTI, Jakarta.
- Budiantara, I. N., Lestari, B., dan Islamiyati, A. 2009^b. *Weighted Spline Estimator in Heteroscedastic Nonparametric Regression for Longitudinal Data*, Proc. IndoMS International Conference on Mathematics and Its Applications, October 12-13, 2009.
- Budiantara, I. N., dan Ratna, M. 2010. *Pendekatan Spline Bivariat Terbobot Dalam Regresi Nonparametrik Multirespon dan Semiparametrik Multirespon*, Laporan Penelitian Fundamental, Tahun I, DP2M DIKTI, Jakarta.
- Budiantara, I. N., dan Jerry, P. 2010^a. *Model Regresi Nonparametrik Spline Terbobot dan Aplikasinya Dalam Merancang Kartu KMS*, Laporan Penelitian Guru Besar, ITS, Surabaya.
- Budiantara, I. N., dan Jerry, P. 2010^b. *Kartu Menuju Sehat (KMS) of Babies in The Province of East Java by Using Weighted Spline Approach*, International Conference ITB, Bandung, 23-25 Nopember, 2010.
- Budiantara, I. N., Lestari, B., dan Islamiyati, A. 2010. *Estimator Spline Terbobot dan Spline Parsial Terbobot Dalam Regresi Nonparametrik dan Semiparametrik Heteroskedastik Untuk Data Logitudinal*, Laporan Penelitian Hibah Kompetensi Tahun II, Tahun 2010, DP2M DIKTI, Jakarta.
- Budiantara, I N. 2011. *Spline: Sebuah Pemodelan Statistika dan Perannya Dalam IPTEK*, Pembicara Utama pada Seminar Nasional Himpunan Mahasiswa Jurusan Pendidikan Matematika FPMIPA IKIP PGRI-BALI, Denpasar, 16 April 2011.
- Budiantara, I. N., dan Jerry, P. 2011. *Infants’ Weight Growth Model in Surabaya (Indonesia) by Using Weighted Spline Regression*, *International Journal of Basic & Applied Sciences*, 11, 151-156.
- Budiantara, I. N., dan Ratna, M. 2011^a. *Pendekatan Spline Multivariat Terbobot Dalam Regresi Nonparametrik Multirespon dan Semiparametrik Multirespon*, Laporan Kemajuan Penelitian Fundamental Tahun I, DP2M-DIKTI, Jakarta.
- Budiantara, I. N., dan Ratna, M. 2011^b. *Model Spline Terbobot Dalam Regresi Nonparametrik Heteroskedastik Untuk Data Longitudinal*, Laporan Kemajuan Penelitian Guru Besar, Lembaga Penelitian ITS, Surabaya.
- Budiantara, I. N., dan Ratna, M. 2011^b. *Pendekatan Spline Bivariat Terbobot Dalam Regresi Nonparametrik Multirespon dan Semiparametrik Multirespon*, Laporan Kemajuan Penelitian Fundamental Tahun II, DP2M-DIKTI, Jakarta.
- Cox, D. D. dan O’Sullivan, F. 1996. *Penalized Type Estimator for Generalized Nonparametric Regression*, 1983, *Journal of Multivariate Analysis*, 56, 185-206.
- Demidenko, E. 2004. *Mixed Models : Theory and Application*, John Wiley and Sons, Ney York.
- Diggle, P. J., Heagerty, P., Liang, K. Y., dan Zelger, S. L.. 2002. *Analysis of Longitudinal Data*, Oxford University Press, Oxford.
- Durban, M., Harezlak, J., Wan, M. P., dan Carroll, L. J. 2005. *Simple Fitting of Subject Specific Curve for Longitudinal Data*, *Statistic in Medicine*, 24, 1153-1167.
- Eubank, R.L. 1988. *Spline Smoothing and Nonparametric Regression*, Mercel Dekker, New York.
- Fan, J. dan Zhang, J. T. 2000. *Two Step Estimation on Functional Linear Models with Applications to Longitudioanal Data*, *Journal of the Royal Statistical Sociaty, Seies B*, 62, 303-322.
- Gu, C. 2000. *Multivariate Spline Regression*, In M. G. Schimeck (ed), *Smoothing and Regression : Approaches, Computation and Application*, New York.
- Guo, W., (2002), *Functional Mixed-Effect Models*, *Biometrics*, 58, 121-128.
- Holmes, C. C. dan Mallick, B. K. 2001. *Bayesian Regression with Multivariate Linear Splines*, *Journal of the Royal Statistical Society, Series B*, 63, 3-18.

- Huang, J. Z., dan Liu, L. 2006. Polynomial Spline Estimation and Inference of Proportional Hazards Regression Models with Flexible Relative Risk Form, *Biometrics*, 62, 793-802.
- Huang, J. Z., Wu, C. O., dan Zhou, L., 2002, Varying-coefficient Models and Basis Function Approximations for the Analysis of Repeated Measurements, *Biometrika*, 89, 111-128.
- Henaulu, M.H., 2009, Pemodelan Nilai UNAS Siswa SMA Negeri 11 Ambon Dengan Pendekatan Regresi Nonparametrik Spline, *Tesis Program S-2 Statistika ITS*.
- Kayri, M., dan Zirrhoglu, G. 2009. Kernel Smoothing Function and Choosing Bandwidth for Nonparametric Regression Methods, *Ozean Journal of Applied Sciences.*, 2, 49-60.
- Koenker, R., Ng, P. dan Portnoy, S. 1994. Quantile Smoothing Spline, *Biometrika*, 81, 673-680.
- Lestari, B., Budiantara, I. N., Sunaryo, S, dan Mashuri, M. 2010. Spline Estimator of Triple Response Nonparametric Regression Model, *Journal of Mathematics and Statistics*, 6, 327-332.
- Oehlert, G.W. 1992. Relaxed Boundary Smoothing Spline, *The Annals of Statistics*, 20, 1146-1160.
- Rice, J. A. dan Wu, C. O. 2001. Nonparametric Mixed Effects Models for Unequally Sampled Noisy Curve, *Biometrics*, 57, 253-259.
- Wahba G. 1990. *Spline Models For Observation Data*, SIAM Pennsylvania.
- Wang, N. 2003. Marginal Nonparametric Kernel Regression Accounting for Within-Subject Correlation, *Biometrika*, 90, 43-52.
- Wang, Y., Guo, W., dan Brown, M. B. 2000. Smoothing Spline for Bivariate Data with Applications to Association Between Hormones, *Statistica Sinica*, 10, 377-397.
- Wu, C. O., dan Chiang, C. T. 2000. Kernel Smoothing on Varying Coefficient Model with Longitudinal Dependent Variable, *Statistica Sinica*, 10, 433-456.
- Wu, H. dan Zhang, J. T. 2006. *Nonparametric Regression Method for Longitudinal Data Analysis : Mixed Effects Modeling Approaches*, John Wiley and Sons, New York.
- Zhang, H.P. 1997. Multivariate Addaptive Spline for the Analysis of Longitudinal Data, *Journal of Computational and Graphical Statistics*, 6, 74-91