

SINDRUM R^2 DALAM ANALISIS REGRESI LINIER RUNTUN WAKTU

Insukindro *)

Universitas Gadjah Mada

ABSTRACTS

This paper attempts to discuss the coefficient of determination (R^2) in time series econometric analysis. The coefficient is the most commonly used measure of the goodness of fit of a regression line. With time series data high R^2 values can be obtained if the linear regressions are estimated with the level. Therefore, there is a strong tendency to estimate the time series model in levels rather than, for example, in first differences. In general, researchers associate a high R^2 with a good fit and it can be considered as indicative of a strong ability of the independent variables to "explain " the dependent variable. In this case, investigators may face the "R~ syndrome " and the "spurious regression " problems.

However, the R^2 statistic is only valid if the proposed model must be linear and include a constant or intercept term and be estimated using ordinary least squares (OLS). Furthermore, it must be noted that the R^2 may not be directly comparable if the dependent variables of the models under consideration are not the same.

Key words: Coefficient of determination, goodness of fit, R^2 syndrome, stasionarity and spurious regression.

*) Penulis mengucapkan terima kasih kepada Prof. Dr. Sukanta Reksohadiprodjo; Prof. Dr. Dibylo Prabowo; Prof. Dr. Soelistyo; Dr. Arief Ramelan Karseno; Dr. Anggito Abimanyu; Drs. M. Adnan Had/am. MA; Drs. Bud; Purnomo, MA dan Drs. Achmad Makhfatih, MA alas komentar dan saran terhadap draft awal tuttsan ini, namun kekurangan dan keketiruan yang mungkin terjadi menjadi tanggung jawab penulis.

1. PENDAHULUAN

Pembahasan mengenai koefisien determinasi (R^2) dalam ekonometrika, khususnya, analisis regresi linier bukanlah hal yang baru, karena koefisien ini merupakan salah satu besaran yang selalu diperhatikan terutama oleh mereka yang masih "pemula" menggunakan pendekatan ekonometrika. Bahkan koefisien ini sering dipandang sebagai besaran yang sangat penting dalam estimasi dengan regresi linier dan tidak jarang mereka terlena oleh besaran itu atau mereka mengalami apa yang disebut dengan "sindrom R^2 " (Maddala, 1992: 235).

Sering terjadi peneliti merasa kecewa atau putus asa bila menemukan nilai R^2 yang rendah atau hasil studi empiriknya tidak menghasilkan nilai R yang mendekati 0,99. Padahal dalam analisis regresi linier klasik (classical linear regression=CR) tidak diharuskan bahwa koefisien R^2 tinggi, bahkan seperti yang diungkapkan oleh Goldberger tahun 1991 bahwa " a high R^2 is not evidence in favor of the model and a low R^2 is not evidence against if (lihat Gujarati, 1995:211).

Harus diakui bahwa nilai koefisien determinasi R^2 akan maksimum jika seorang peneliti menaksir model regresi linier dengan ordinary least squares (OLS), dan memang koefisien ini dapat dipakai sebagai salah satu besaran untuk memilih model empirik yang baik (untuk referensi lihat misalnya, Harvey,1991; Hendry dan Ericsson, 1991; Gujarati, 1995 dan Thomas, 1997). Harvey (1991: 5-6),misalnya, menyebutkan model yang baik adalah (1) model yang sederhana (parsimony), (2) mempunyai satu himpunan parameter yang konsisten dengan data (identifiability), (3)koheren dengan data (data coherency) yang antara lain dikaji melalui "goodness of fit" atau biasanya dicerminkan oleh koefisien determinasi R^2 , (4) admisibilitas data (data admisibility) yang antara lain menyatakan bahwa suatu model yang baik seharusnya tidak mampu memprediksi besaran-besaran ekonomi yang menyimpang dari kendala definisi ekonomika, misalnya, beberapa variabel tidak dapat negatif, (5) konsisten dengan teori (theoretical consistency) ekonomika yang dipilih, (6) mempunyai kemampuan untuk memprediksi (predictive power) di dalam sampel, dan (7) memiliki keunggulan (encompassing) dalam arti bahwa dia mampu

menjelaskan lebih baik studi empiris yang dihasilkan oleh model pesaingnya. Selaras dengan kriteria survey tersebut di atas, Thomas (1997: 361-363) menyebutkan bahwa suatu model yang baik seharusnya: (1) koheren dengan data (data-coherent), (2) mempunyai variabel bebas (independent variable) yang eksogen (exogenos), (3) mempunyai parameter yang konstan (constant parameters), (4) mempunyai admisibilitas terhadap data (data-admissible), (5) konsisten dengan teori ekonomika (consistent with economic theory), (6) mengungguli (encompass) model pesaingnya, dan (7) sederhana (parsimonious).

Berdasarkan dua pendapat di atas dapat dikatakan bahwa koefisien R^2 hanyalah salah satu dan bukan satu-satunya kriteria memilih model yang baik. Dengan demikian, bila suatu estimasi regresi linier menghasilkan koefisien yang tinggi, tetapi tidak konsisten dengan teori ekonomika yang dipilih oleh si peneliti atau tidak lolos dari uji asumsi regresi linier fisik, misalnya, maka model tersebut bukanlah model penaksir yang baik dan seharusnya tidak dipilih menjadi model empirik. Ha! semacam ini dalam analisis ekonometrika sering dikenal sebagai regresi lancung atau semrawut (spurious regressions) (Thomas, 1997).

Secara umum dapat dikatakan bahwa estimasi regresi linier untuk data lintas sektoral (cross section) akan menghasilkan koefisien R^2 yang relatif rendah karena adanya variasi yang besar antar masing-masing pengamatan (Ramanathan, 1992: 199). Hal ini akan berbeda dengan estimasi regresi linier untuk data runtun waktu (time series) yang biasanya menghasilkan koefisien R^2 yang tinggi (Pindyck dan Rubinfeld, 1991: 62, Maddala, 1992: 550). Namun perlu dikemukakan di sini, bahwa koefisien R^2 yang tinggi tidaklah berarti bahwa R^2 harus sama dengan satu, sebab bila $R^2 = 1$ berarti model tersebut memiliki variabel gangguan (residual) yang sama dengan nol untuk setiap pengamatan. Tentu saja bila seorang peneliti mengestimasi model regresi linier dan mendapatkan $R^2 = 1$, maka dalam analisis ekonometrika hasil tersebut merupakan peringatan (warning) awal bagi yang bersangkutan tentang kemungkinan adanya manipulasi data, misalnya, atau kesalahan spesifikasi.

Makalah ini bermaksud mengetengahkan kemungkinan adanya kerancuan pemberian makna statistika terhadap R^2 ataupun R^2 disesuaikan (adjusted R^2 atau \bar{R}^2) terutama dalam analisis regresi untuk data runtun waktu. Kerancuan ini sering menimbulkan adanya sindrum R^2 yang menganggap bahwa model yang baik adalah model yang memiliki R^2 yang tinggi tanpa memperhatikan kriteria-kriteria lain yang juga sangat penting dalam studi empiris.

Pembahasan dimulai dengan tinjauan teoritik mengenai konsep R^2 dan kemudian dilanjutkan dengan uraian mengenai validitas koefisien R^2 dan manfaat koefisien determinasi untuk memilih model yang baik. Guna mendukung tinjauan tersebut diungkapkan juga satu studi empirik mengenai pengeluaran konsumsi masyarakat di Indonesia. Makalah ini diakhiri dengan beberapa catatan sebagai penutup.

2. KONSEP KOEFISIEN R^2

Koefisien determinasi (R^2) dalam analisis regresi linier yang diestimasi dengan OLS menjelaskan proporsi variasi variabel tak bebas (dependent variable) yang mampu dijelaskan oleh variasi variabel bebasnya (Kennedy, 1992: 12 dan Thomas, 1997: 89). Dengan demikian jika suatu model yang terdiri dari, misalnya, satu variabel tak bebas (Y) dan satu himpunan variabel bebas (X) memiliki hasil estimasi R^2 sebesar **0,81** berarti **81 persen dari variasi variabel Y mampu dijelaskan oleh variasi himpunan variabel X**. Perlu dikemukakan di sini bahwa telah dijumpai di banyak buku ekonometrika terutama buku terjemahan dan juga peneliti yang mengartikan $R^2 = 0,81$ tersebut dengan mengatakan bahwa 81 persen perilaku variabel Y mampu dijelaskan oleh variabel X. **Konsep yang disebut terakhir ini adalah pemberian arti yang keliru** dan perlu mendapat perhatian yang serius bagi mereka yang menggunakan ekonometrika, khususnya analisis regresi linier dalam studinya. Uraian berikut ini dikemukakan untuk menghilangkan kerancuan dan kekeliruan yang masih saja sering terjadi mengenai koefisien R^2 tersebut.

Untuk mendapat gambaran yang lebih jelas mengenai konsep R^2 tersebut di atas, anggaplah terdapat model linier sederhana:

$$Y_i = \alpha_0 + \alpha_1 X_i + V_i \quad (1)$$

di mana parameter α_0 dan α_1 dianggap berpangkat satu, Y_i adalah variabel tak bebas, X_i merupakan himpunan variabel bebas dan V_i adalah variabel gangguan atau residu (residual).

Selanjutnya, dengan menggunakan analisis regresi linier persamaan (1) dapat diperoleh koefisien R^2 sebagai:

$$R^2 = 1 - \text{RSS} / \text{TSS} \quad (2)$$

di mana RSS adalah residual sum of squares

$$= \sum_{i=1}^k V_i^2$$

dan TSS yaitu total sum of squares yang merupakan penjumlahan RSS dan ESS (explained sum of squares) serta besarnya sama dengan:

$$\sum_{i=1}^k (Y_i - \bar{Y})^2$$

dan \bar{Y} adalah rata-rata Y_i

Melalui persamaan (2) dapat diamati bahwa besar kecilnya koefisien R^2 dipengaruhi dua faktor yaitu RSS dan sebaran data Y terhadap reratanya. Analisis parsial kedua faktor tersebut mengatakan bahwa jika TSS tertentu maka besar kecilnya R^2 tergantung RSS. Semakin kecil RSS maka koefisien R^2 semakin besar dan sebaliknya. Di sisi lain, untuk RSS yang tertentu maka semakin melebar sebaran data Y terhadap reratanya akan semakin besar pula R dan sebaliknya semakin

menyempit sebaran Y terhadap reratanya akan mengakibatkan semakin kecil nilai koefisien R^2 .

Secara matematika koefisien R^2 dapat sama dengan satu dan terjadi bila RSS sama dengan nol. Kondisi ini ($RSS = 0$) hanya dapat terjadi bila residu (V) untuk setiap titik pengamatan sama dengan nol. Fenomena semacam ini mungkin saja terjadi bila peneliti melakukan manipulasi data dengan menggunakan metode substitusi seperti yang sering digunakan dalam analisis keseimbangan atau analisis matematika ekonomika. Namun, dalam analisis ekonometrika RSS tidak mungkin sama dengan nol, sehingga R^2 juga tidak mungkin sama dengan satu. Hal ini karena dalam model ekonometrika residu harus ada, sebab konsep dasar yang dipakai dalam analisis ekonometrika runtun waktu, khususnya, adalah analisis ketidakseimbangan (disequilibrium) yang memungkinkan adanya biaya ketidakseimbangan (disequilibrium cost) dan biaya penyesuaian (adjustment cost) yang semuanya ini mencerminkan akan adanya residu (lihat misalnya, Domowitz dan Elbadawi, 1987; Cuthbertson, 1988 dan Insukindro, 1992a).

Secara umum dapat dikatakan bahwa residu terjadi dalam analisis regresi linier disebabkan oleh adanya kesalahan spesifikasi (specification error) terutama karena peneliti tidak memasukkan variabel yang relevan dan kesalahan membuat definisi operasional atau definisi kerja variabel yang diamati dan kesalahan cara mengukur variabel serta karena isu-isu statistika mengenai perilaku variabel yang sedang diamati (Kennedy, 1992: 2-4; Gujarati, 1995: 39-41, Johnston dan DiNardo, 1997: 110-111 dan Thomas, 1997: 339-350). Dengan kata lain, residu dalam analisis regresi pasti ada walaupun nilainya mungkin negatif atau positif atau mendekati nol untuk setiap titik pengamatan, sehingga RSS juga pasti ada dan lebih besar dari nol dan akibatnya koefisien R tidak mungkin sama dengan satu.

3. VALIDITAS KOEFISIEN R^2

Di muka telah dibicarakan konsep R^2 dan hubungannya dengan RSS yang merupakan bentuk lain dari OLS. Seperti yang akan dibahas di bawah ini, estimasi suatu model regresi linier dengan menggunakan OLS menghendaki agar RSS regresi tersebut minimum, sehingga dengan sendirinya akan diperoleh R^2 maksimum. Namun, seperti telah disinggung di muka, pengertian R^2 maksimum tidaklah berarti bahwa R^2 harus sama dengan satu atau mendekati satu.

Pertanyaan yang perlu diajukan di sini adalah apakah setiap nilai R^2 yang dihasilkan dari suatu estimasi regresi pasti layak atau sah (valid). Validitas nilai R ditentukan oleh beberapa faktor (Kennedy, 1992: 26-28): (1) model regresi yang akan ditaksir haruslah linier, (2) model tersebut harus memiliki intersep (intercept) dan (3) model regresi tersebut diestimasi dengan OLS.

Unsur suatu model yang akan ditaksir biasanya mencakup hubungan antar variabel yang dipilih, definisi operasional (kerja) variabel-variabel dalam model yang akan ditaksir, hipotesis dan asumsi yang digunakan. Umumnya model yang akan ditaksir didasari pada model teoritik dan dibangun berdasarkan suatu teori ekonomika yang dipilih oleh si pembuat model (Insukindro, 1992b). Pemilihan teori ekonomika sangat penting dalam analisis ekonometrika walaupun persoalan ini tidak selalu dapat dipecahkan dalam pembentukan model. Melalui teori ekonomika tertentu, seorang ekonom dapat memilih dan menjelaskan hubungan antar variabel bebas dan tak bebas. Setelah hubungan antara variabel-variabel itu diformulasikan dalam suatu model, maka langkah selanjutnya adalah menaksir model terkait.

Dalam ekonometrika dikenal adanya model regresi linier dan non linier. Namun perlu dicatat di sini bahwa pengertian linier dan non linier dalam ekonometrika tidak sama dengan konsep linier dan non linier dalam matematika. Menurut ekonometrika, suatu model dikatakan linier jika parameter yang diestimasi adalah linier atau berpangkat satu, sehingga bila parameternya berpangkat lebih dari satu maka model tersebut dikatakan model regresi non linier. Tentu saja konsep ini

berbeda dengan matematika yang mengatakan bahwa suatu fungsi atau model matematika dikatakan linier (atau non linier) jika variabel-variabel dalam model tersebut berpangkat satu (atau berpangkat lebih dari satu).

Konsep koefisien R^2 seperti yang dibicarakan di atas didasarkan pada anggapan bahwa model regresinya adalah linier dalam arti bahwa parameternya berpangkat satu dan memiliki intersep serta diestimasi dengan OLS atau dengan meminimumkan RSS. Dengan demikian jika modelnya tidak linier, maka untuk mengestimasi tidak dapat digunakan OLS tetapi harus menggunakan NLLS (nonlinear least squares) atau IV-NLLS (Instrumental variable-NLLS) (lihat misalnya, Price dan Insukindro, 1994). Jika model yang akan ditaksir tidak memiliki intersep, dengan sendiri persamaan (2) tidak dapat digunakan lagi dan dimungkinkan diperoleh R^2 negatif serta tidak mudah diberi arti. Untuk menaksir R^2 dalam kasus yang disebut terakhir ini, beberapa program komputer menggunakan formulasi sebagai berikut (Ramanathan, 1992: 166):

$$R^2 = 1 - \text{RSS} / Y_i^2$$

Estimasi OLS untuk model regresi runtun waktu hendaknya dilakukan dengan penuh kehati-hatian. Hal ini karena variabel bebas runtun waktu pada umumnya adalah variabel stokastik (stochastic variable) dan tentu saja menyimpang dari asumsi regresi linier klasik yang mengatakan bahwa variabel bebas adalah variabel deterministik atau non stokastik. Penyimpang-penyimpang tersebut akan lebih diperparah lagi bila variabel-variabel bebas yang dipilih dan digunakan dalam estimasi OLS berkorelasi semasa dengan residunya atau sering dikenal sebagai contemporaneously correlated case. Lebih lanjut, Kmenta (1990: 334-341) dan Thomas (1997: 206-211) menyebutkan bahwa jika variabel-variabel bebas runtun waktu adalah stokastik, maka terdapat 3 kemungkinan kasus yang dapat terjadi. Pertama, kasus bebas (independence case) terjadi jika variabel-variabel bebas tidak berkorelasi dengan residu baik untuk masa lalu, sekarang maupun yang akan datang. Dalam kasus ini

estimasi OLS tetap tidak bias dan konsisten Kedua, kasus tidak berkorelasi semasa (contemporaneously uncorrelated case) terjadi jika variabel-variabel bebas tidak berkorelasi semasa dengan residunya. Di sini estimasi OLS akan bias tetapi tetap konsisten. Ketiga, kasus berkorelasi semasa (contemporaneously correlated case) terjadi bila variabel-variabel bebas berkorelasi semasa dengan residunya, sehingga estimasi OLS akan bias dan tidak konsisten.

Dari ketiga kasus di atas terlihat bahwa estimasi OLS hanya tidak bias dan konsisten untuk kasus bebas (kasus pertama), sedang untuk kasus yang lain estimasi OLS akan bias. Padahal untuk data runtun waktu, koefisien R^2 pada umumnya adalah tinggi (sering mendekati 0,9), sehingga dalam hal ini peneliti dihadapkan pada regresi lancung (spurious regression).

Untuk mengatasi persoalan regresi lancung seperti tersebut di atas paling tidak ada dua cara yang dapat dikemukakan di sini (Insukindro, 1991). Pertama, peneliti tidak memperhatikan perilaku data atau tidak melakukan uji derajat integrasi (testing for degree of integration) terhadap variabel yang diamati. Dengan kata lain, pengamat mengabaikan akan adanya konsep stasioneritas (stationarity) dalam analisis data runtun waktu. Dalam kaitan ini untuk mengatasi kemungkinan adanya regresi lancung, maka peneliti harus membentuk model dinamik (dynamic model) yang meliputi variabel-variabel kelambanan tak bebas dan bebas (lagged dependent and independent variables). Salah satu model dinamik yang meliputi kedua komponen variabel kelambanan tersebut adalah model koreksi kesalahan (error correction model) yang akhir-akhir ini banyak diterapkan dalam studi empirik (untuk survei lihat misalnya, Domowitz dan Elbadawi, 1987; Wickens dan Breusch, 1988; Insukindro, 1992a; Thomas, 1993, 1997). Tentu saja agar hasil estimasi model dinamik tersebut layak, maka residu model koreksi kesalahan pun harus memenuhi asumsi regresi linier klasik, seperti terdapat otokorelasi dan homoskedastisitas atau residunya stasioner.

Kedua, peneliti memperhatikan pentingnya konsep stasioneritas dalam analisis runtun waktu. Dalam kaitan ini, pengujian mengenai derajat integrasi terhadap

variabel-variabel (baik variabel bebas maupun tak bebas) merupakan suatu kegiatan analisis statistika yang tidak dapat diabaikan. Langkah awal ini penting dan perlu dilakukan karena uji derajat integrasi merupakan prasyarat diberlakukannya pendekatan kointegrasi (cointegration approach) yang dalam satu dasa warsa terakhir ini merupakan topik penting dalam ekonometrika runtun waktu (untuk kepustakaan mengenai pendekatan kointegrasi, lihat misalnya: Engle dan Granger, 1991 dan Thomas, 1997, bab 15). Pendekatan kointegrasi menghendaki agar variabel-variabel runtun waktu yang biasa-nya stokastik mempunyai derajat integrasi yang sama dan selanjutnya akan diuji apakah residu regresi kointegrasi (cointegrating regression) akan stasioner atau tidak. Jika residu regresi kointegrasi stasioner diharapkan asumsi regresi linier klasik seperti asa residu sama dengan nol, homoskedastisitas dan tidak ada otokorelasi akan dapat dipenuhi (Thomas, 1997: 424-428). Jika yang disebut terakhir ini dapat tercapai, maka peneliti dapat terhindarkan dari kemungkinan munculnya masalah regresi lancung dalam studi empiriknya.

Dari dua cara penyelesaian di atas dapat dikatakan bahwa walaupun variabel-variabel yang digunakan dalam suatu model adalah variabel yang stokastik atau tidak stasioner, namun bila residu atau variabel gangguan untuk model dinamik (kasus pertama) maupun regresi kointegrasi (kasus kedua) stasioner, maka kemungkinan adanya penyimpangan dari asumsi regresi linier klasik dapat dihindarkan juga. Dengan kata lain, bila permasalahan regresi lancung dapat diatasi, maka koefisien R masih layak untuk dipakai sebagai salah satu kriteria memilih model seperti yang telah dibahas di muka.

Di samping itu perlu dikemukakan di sini bahwa koefisien R^2 juga sangat sensitif terhadap jumlah variabel bebas yang digunakan dalam suatu model yang akan ditaksir. Pada umumnya, setiap penambahan variabel baru ke dalam suatu model yang ditaksir akan secara otomatis meningkatkan R^2 walaupun belum tentu meningkatkan R^2 yang disesuaikan. Persoalannya adalah bahwa pemilihan variabel yang baru itu dapat dilakukan sesuka hati si peneliti dalam arti si peneliti dapat saja memilih suatu variabel tanpa ada dasar logika ataupun teori ekonomika. Dalam kasus

ini yang dipentingkan adalah jumlah pengamatan (N) variabel yang baru sama dengan jumlah pengamatan dalam model yang ditaksir. Di sisi lain, walaupun memasukkan variabel baru dapat meningkatkan nilai R^2 , tetapi kondisi ini dapat menyebabkan kemungkinan munculnya multi-kolinieritas (multicollinearity) dan biasanya dicirikan oleh banyaknya statistik t koefisien regresi rendah atau tidak signifikan (Ramanathan, 1992: 240 dan Gujarati, 1995: 315).

4. MEMILIH MODEL DENGAN KRITERIA KOEFISIEN DETERMINASI R^2

Seperti telah dibicarakan di atas, salah satu manfaat yang sering ditonjolkan dari koefisien determinasi R^2 adalah kemampuan besaran statistik deskriptif ini untuk dipakai sebagai salah satu kriteria memilih suatu model terutama dikaitkan dengan "goodness of fit" dari model tersebut. Namun perlu dicatat di sini bahwa agar suatu koefisien determinasi R^2 layak sebagai salah satu kriteria, dia harus dapat memenuhi syarat-syarat sebagai berikut (Kmenta, 1990: 241-243; Pindyck dan Rubin-feld, 1991: 79-81; Kennedy, 1992: 26-28; Ramanathan, 1992: 124; Gujarati, 1995: 265-266 dan Thomas, 1997: 86-90): (1).variabel tak bebas (dependent variable) dari model-model yang akan dipilih harus sama, (2) jumlah data atau pengamatan (N) dari model-model yang akan dipilih harus sama dan (3) model-model yang akan dipilh harus lolos uji asumsi regresi linier klasik atau residunya stasioner atau terhindar dari masalah regresi lancung.

Syarat pertama agar peneliti dapat membandingkan dua model dengan pendekatan koefisien determinasi R^2 adalah bahwa variabel tak bebas dari dua atau lebih model yang akan dipilih harus sama, sedangkan variabel bebas dalam model-model tersebut dapat berbeda. Misalnya, suatu model konsumsi C tentunya tidak dapat dibandingkan dengan model tabungan (S) walaupun mungkin kedua variabel itu (C dan S) dipengaruhi oleh variabel-variabel bebas yang sama, misalnya, pendapatan (Y) dan suku bunga (i). Hal ini karena dalam model konsumsi, koefisien

determinasinya mengukur besarnya proporsi variasi C yang mampu dijelaskan oleh variasi Y dan i . Di sisi lain, dalam model tabungan, koefisien determinasi di sini mengukur besarnya proporsi variasi S yang mampu dijelaskan oleh variasi Y dan i . Hal serupa juga terjadi bila ingin diamati pengaruh Y dan i terhadap LC (di mana $LC = \text{Log } C$). Koefisien determinasi model $LC = F(Y,i)$ mengukur besarnya proporsi variasi variabel LC yang mampu dijelaskan oleh variasi Y dan i . Tentu saja, koefisien determinasi model $C = C(Y,i)$ berbeda dengan R^2 untuk model $LC = LC(Y,i)$ dan tentu saja kedua koefisien determinasi tersebut tidak dapat dibandingkan karena memang yang diukur juga berbeda. Dengan demikian, koefisien R tidak layak dipakai untuk membandingkan model-model yang memiliki variabel tak bebas yang berbeda (lihat juga Kennedy, 1992: 102).

Untuk membandingkan model-model yang memiliki variabel tak bebas yang tidak sama dapat dilakukan dengan uji MWD atau uji MacKinnon, White dan Davidson (lihat, Gujarati, 1995: 265-266).

Jumlah sampel atau pengamatan (N) perlu pula diperhatikan jika ingin menggunakan koefisien R^2 sebagai kriteria pemilihan model. Hal ini karena besar kecilnya N akan menentukan besar kecilnya TSS {total sum of squares}. Dua model yang memiliki variabel tak bebas yang sama, tetapi memiliki N yang tidak sama akan mengakibatkan rerata variabel tak bebas yang berbeda, sehingga TSS juga berbeda dan dengan sendirinya koefisien determinasinya juga tidak layak dibandingkan.

Di samping itu dua model regresi lancung tentu saja tidak layak dibandingkan walaupun mungkin kedua model tersebut memiliki variabel tak bebas dan jumlah pengamatan yang sama. Hal ini karena hasil estimasi regresi tersebut mungkin akan bias atau tidak efisien. Dengan demikian sebelum membandingkan dua model, misalnya, perlu diberlakukan terlebih dulu uji diagnosis atau uji stasioneritas.

5. STUDI EMPIRIK: PENGELUARAN KONSUMSI MASYARAKAT DI INDONESIA

Sejauh ini telah dibicarakan konsep, validitas dan manfaat koefisien R^2 serta masalah-masalah yang muncul terutama bila dikaitkan dengan regresi lancung. Kini tiba saatnya dibicarakan lebih lanjut hal-hal yang sedang menjadi pusat perhatian dalam tulisan ini dengan satu studi empirik mengenai pengeluaran konsumsi masyarakat Indonesia untuk periode 1970-1994. Tentu saja sesuai dengan tujuan yang ingin dicapai dalam tulisan ini, pembahasan utama yang akan diketengahkan berkaitan dengan koefisien R^2 , sedangkan pembicaraan mengenai hasrat marginal mengkonsumsi (marginal propensity to consume), misalnya, hanya sebagai pelengkap.

Selanjutnya, untuk mendukung maksud tersebut akan dibahas pengaruh (positif) PDB atau Produksi Domestik Bruto (Y) terhadap pengeluaran konsumsi masyarakat (C) dan keduanya dihitung menurut harga berlaku. Pembicaraan akan diawali dengan uji derajat integrasi atau uji stasioneritas mengikuti metode yang dikembangkan oleh Dickey dan Fuller (1981). Hasil studi, seperti dilaporkan pada Tabel 1, menunjukkan bahwa variabel Produksi Domestik Bruto (baik yang dinyatakan oleh Y maupun LY) dan pengeluaran konsumsi masyarakat (dinyatakan oleh C atau LC) tidak stasioner. Dengan kata lain, kedua variabel adalah variabel-variabel stokastik.

Tabel 1. Estimasi OLS Statistik DF dan ADF untuk Uji Derajat Integrasi Pengeluaran Konsumsi Masyarakat dan PDB Indonesia: 1970-1994

Variabel	DF	ADF
Y	1,91	2,30
LY	-2,41	-1,97
C	-1,21	-1,97
LC	-1,82	-2,17

Keterangan:

- a. AD = Statistik Dickey - Fuller
ADF = Statistik Augmented Dickey-Fuller
- b. Tabel Statistik DF dan ADF dapat dilihat pada Dickey dan Fuller (1981: 1062) Tabel I dan II.

Selanjutnya ditaksir 4 model pengeluaran konsumsi masyarakat dan hasil estimasi dilaporkan pada Tabel 2. Model 1 dan 2 adalah model yang memiliki variabel tak bebas C, sedangkan model 3 dan 4 memiliki variabel tak bebas LC. Pertama, akan dibahas hubungan antara koefisien R dan regresi lancung, misalnya terlihat pada model 1 dan 3. Estimasi OLS model 1 dan 3 menghasilkan koefisien R masing-masing sebesar 0,996 dan 0,999, tetapi model 1 tidak lolos uji linieritas, normalitas dan homoskedastisitas dan hanya lolos uji tidak terdapat korelasi serial, sedangkan model 3 tidak lolos uji korelasi serial dan linieritas.

Dengan demikian, walaupun kedua model tersebut memiliki koefisien determinasi yang sangat tinggi, namun mereka merupakan regresi lancung, sehingga sudah selayaknya jika kedua model ini tidak dipilih sebagai model empirik yang baik.

Untuk memecahkan persoalan regresi lancung di atas dan juga selaras dengan perilaku data C dan Y (atau LC dan LY) yang stokastik (lihat juga Tabel 1), maka dibentuk model dinamik seperti tercermin pada model 2 dan 4. Estimasi model 2 dan 4 menghasilkan koefisien R^2 yang sangat tinggi yaitu masing-masing sebesar 0,997 dan 0,999 dan lolos dari uji asumsi regresi linier klasik (non-korelasi serial, homoskedastisitas, linieritas dan normalitas), sehingga terhindarkan dari persoalan regresi lancung dan kedua model ini layak untuk dipilih sebagai kandidat model empirik yang baik. Koefisien R^2 model 2 menunjukkan besarnya proporsi (0,997) variasi variabel C, yang mampu dijelaskan oleh variasi variabel bebas model 2 (yaitu Y_t , Y_{t-1} , dan C_{t-1}), sedangkan R model 4 menjelaskan proporsi (0,999) variasi variabel LC, yang mampu dijelaskan oleh variasi variabel LY_t , LY_{t-1} dan LC_{t-1} . Di samping itu dicatat pula bahwa rerata variabel tak bebas model 2 (yaitu sebesar 59329,1) lebih

besar dari pada rerata variabel tak bebas model 4 (yaitu sebesar 10,369). Dengan demikian, jika peneliti hanya mendasarkan analisisnya pada koefisien R maka model 2 dan 4 memang tidak layak dibandingkan, walaupun kedua model telah lolos dari berbagai uji diagnosis.

**Tabel 2. Estimasi OLS Pengeluaran Konsumsi Masyarakat di Indonesia:
1970-1994**

Variabel Bebas	Variabel Tak Bebas			
	Model 1 (C _t)	Model 2 (C _t)	Model 3 (L C _t)	Model 4 (L C _t)
Intersep	1174,6 (1,128)	201,523 (0,182)	0,323 (4,439)	0,183 (1,583)
Y _t	0,581 (77,846)	0,822 (6,775)		
Y _{t-1}		-0,607 (-2,529)		
C _{t-1}		0,542 (2,218)		
LY _t			0,925 (137,855)	0,783 (8,337)
LY _{t-1}				-0,261 (-1,172)
LC _{t-1}				0,438 (2,214)
N	25	24	25	24
R ²	0,996	0,997	0,999	0,999
\bar{R}^2	0,996	0,997	0,999	0,999
F	6060,0	22394,0	19003,9	7853,4
Rerata	57063,6	59329,1	10,270	10,639

Uji Diagnosis				
1. Korelasi Serial				
- DW	1,409	2,135	0,949	1,863
- $F^2(1)$	1,755	0,735	5,021	0,123
2. Linieritas				
- $F^2(1)$	4,250	2,594	7,530	3,024
3. Normalitas				
- $X^2(2)$	7,755	1,084	0,022	3,695
4. Heteroskedastisitas				
- $F^2(1)$	4,250	3,157	2,025	4,140*)

Catatan:

a. Angka dalam kurung di bawah koefisien regresi menunjukkan nilai statistik

b. Tabel Statistik

$$= 0,05 \quad F^2(1) = 3,841 \quad F^2(2) = 5,991$$

$$= 0,025 \quad F^2(1) = 5,024 \quad F^2(2) = 7,378$$

*) Signifikan untuk $\alpha = 0,025$

6. KESIMPULAN

Tinjauan teoritik dan empirik dalam makalah ini telah dipusatkan kepada apa yang sering disebut dengan sindrom R^2 terutama dalam analisis regresi linier runtun waktu. Koefisien R^2 yang tinggi merupakan harapan dari suatu penelitian karena dia memang dapat dipakai untuk mengetahui "goodness of fit" dari model yang sedang ditaksir. Namun penggunaan koefisien R^2 dalam ekonometrika runtun waktu harus dilakukan dengan hati-hati, terutama berkaitan dengan adanya masalah regresi lancung, terlebih lagi bila koefisien ini hendak digunakan untuk memilih model. Hal ini karena koefisien ini tidak layak dipakai untuk membandingkan dua model atau lebih yang memiliki variabel tak bebas yang berbeda. Hasil studi empirik yang dibahas dalam makalah ini juga mendukung argumen tersebut di atas dan menunjukkan bahwa koefisien R^2 tidak layak untuk dipakai sebagai salah satu kriteria pemilihan model yang memiliki variabel tak bebas yang berbeda.

Pada umumnya, mereka yang masih "pemula" menggunakan analisis ekonometrika, khususnya analisis regresi linier, sering terlena oleh harapan untuk selalu mendapatkan R^2 yang tinggi, walaupun pakar ekonometrika seharusnya lebih tertarik memperoleh penaksir yang baik dari parameter yang sedang menjadi pusat perhatiannya dari pada koefisien R^2 itu sendiri. Seperti yang diungkapkan oleh Kennedy (1992: 28) sebagai berikut:

"In general, econometricians are interested in obtaining "good" parameter estimates where "good" is not defined in terms of R^2 . Consequently the measure R^2 is not of much importance in econometrics". Ungkapan Kennedy ini nampaknya selaras dengan pendapat Golberger tahun 1991 seperti yang dikutip oleh Gujarati (1995: 211), " In fact the most important thing about R^2 is that it is not important in the CR (classical linear regression) model. The CR model is concerned with parameters in a population, not with goodness of fit in the sample".

7. DAFTAR PUSTAKA

- Cuthbertson, K. (1988), The Demand for M1: A Forward Looking Buffer Stock Model, *Oxford Economic Papers*, 40: 110-131.
- Dickey, D.A. and W.A. Fuller (1981), Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root, *Econometrica*, 49: 1057-1072.
- Domowitz, I and L. Elbadawi (1987), "An Error Correction Approach to Money Demand: The Case of the Sudan", *Journal of Development Economics*, 5: 26-46.
- Engle, R.F. and C.W.J. Granger (1991), *Long-Run Economic Relationships, Readings in Cointegration*, Oxford University Press.
- Gujarati, D.N. (1995), *Basic Econometrics*, McGraw-Hill.
- Harvey, A.C. (1991), *The Econometric Analysis of Time Series*, Philip Allan.
- Hendry, D.F. and N.R. Ericsson (1991), An Econometric Analysis of U.K. Money Demand in Monetary Trends in the United States and the United Kingdom by Milton Friedman and Anna J. Schwartz, *American Economic Review*, 81: 8-38.

- Insukindro (1991), "Regresi Linier Lancung dalam Analisis Ekonomi: Suatu Tinjauan dengan Satu Studi Kasus di Indonesia", *Jurnal Ekonomi dan Bisnis Indonesia*, 6: 75-88.
- Insukindro (1992a), "Dynamic Specification of Demand for Money: A Survey of Recent Development", *Jurnal Ekonomi Indonesia*, 1:8-23.
- Insukindro (1992b), "Pembentukan Model dalam Penelitian Ekonomi", *Jurnal Ekonomi dan Bisnis Indonesia*, 7 : 1-17.
- Johnston, J and J. DiNardo (1997), *Econometric Methods*, McGraw-Hill.
- Kennedy, P (1992), *A Guide to Econometrics*, The MIT Press
- Kmenta, J. (1990), *Elements of Econometrics*, Macmillan Publishing Company.
- Maddala, G.S. (1992), *Introduction to Econometrics*, Macmillan Publishing Company
- Pindyck, R.S. and D.L. Rubinfeld (1991), *Econometric Methods and Economic Forecasts*, McGraw-Hill.
- Price, S and Insukindro (1994), "The Demand for Indonesian Narrow Money: Long-run, Error Correction and Forward-Looking Behaviour", *Journal of International Trade and Economic Development*, 3: 147-163.
- Ramanathan, R. (1992), *Introductory Econometrics with Applications*, Harcourt Brace Javanovich, Inc.
- Thomas, (1993), *Introductory Econometrics, Theory and Applications*, Longman Group Limited.
- Thomas (1997), *Modern Econometrics*, Addison-Wesley.
- Wickens, M.R. and T.S. Breusch (1988), "Dynamic Specifications, the Long-Run and Estimation of Transformed Regression Model", *Economic Journal*, 98 (Supplement): 189-205