

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi bertujuan untuk menjelaskan atau memodelkan hubungan antar variabel. Variabel y disebut juga dengan variabel respon, output, variabel dependen atau variabel yang dijelaskan, variabel x adalah variabel independen, variabel input atau penjelas (Kurniawan dan Yuniarto, 2016).

Regresi terbagi menjadi dua persamaan matematika, yang bisa linier atau non-linier (Supranto, 2000). Semua model yang parameternya linier seperti model linier, terlepas dari bagaimana variabel independen dibentuk dalam model regresi. Sedangkan model regresi nonlinier merupakan model regresi yang parameternya tidak linier, seperti model regresi eksponensial (Walpole et al., 2011).

Regresi linier merupakan fungsi yang mengaitkan antara variabel bebas (*independent*) dan variabel terikat (*dependent*) dimana parameternya linier dengan variabel terikat (Walpole et al., 2011). Variabel yang mempengaruhi biasanya disebut variabel independen dan variabel yang terpengaruh biasanya disebut variabel terikat. Regresi linier juga membentuk hubungan antara variabel bebas dengan variabel terikat yang fungsinya linier atau polinomial. Regresi linier hanya dapat digunakan untuk skala interval dan proporsional (Supranto, 2000).

Penelitian ini mempertimbangkan variabel bebasnya merupakan order satu atau fungsi linier. Analisis fungsi linier yang hanya melibatkan sebuah variabel bebas disebut dengan analisis regresi linear sederhana. Sementara, apabila analisisnya melibatkan lebih dari satu variabel bebas, maka analisis yang digunakan adalah analisis regresi linear berganda (Pasaribu et al., 2015).

Sebagian besar masalah penelitian yang menerapkan analisis regresi memerlukan variabel dependen dan beberapa variabel independen dalam model regresi. Kerumitan dari sebagian besar mekanisme ilmiah yang sangat penting adalah untuk memprediksi variabel dependen, hal ini memerlukan regresi berganda. Apabila Koefisien-koefisien model adalah linier, maka model yang digunakan adalah model regresi linier berganda (Walpole et al., 2011).

Regresi linier berganda bertujuan untuk mengetahui hubungan fungsi linier antara variabel dependen dan beberapa variabel independen. Regresi berganda memungkinkan peneliti untuk secara bersamaan menguji efek dari beberapa variabel independen terhadap variabel dependen (Yan dan Su, 2007).

Pada analisis regresi diperlukan suatu metode untuk mengestimasi parameter agar dapat memenuhi *Best Linear Unbiased Estimator* (BLUE) salah satu metode yang paling sering digunakan adalah *Ordinary Least Square* (OLS) atau biasa disebut dengan Metode Kuadrat Terkecil (MKT) (Gujarati, 2007). Koefisien-koefisien regresi linier sebenarnya adalah nilai duga dari parameter model regresi. Parameter merupakan keadaan sesungguhnya untuk kasus yang diamati. Parameter regresi diduga melalui teknik perhitungan yang disebut *Ordinary Least Square* (OLS). Dalam menduga, mungkin tidak terlepas dari kesalahan, baik itu sedikit maupun banyak. Perhitungan dengan MKT, kesalahan pendugaan dijamin yang terkecil (dan merupakan yang terbaik) asal memenuhi beberapa asumsi. Asumsi-asumsi tersebut biasanya disebut asumsi klasik regresi linier (Kurniawan dan Yuniarto, 2016). Asumsi yang harus dipenuhi dalam asumsi klasik adalah asumsi normalitas, asumsi linieritas, tidak ada masalah multikolinieritas, tidak ada masalah autokorelasi dan tidak ada masalah heteroskedastisitas (Markidakis et al., 1999).

Koefisien regresi yang di dapatkan telah benar atau dapat diterima diperlukan melakukan pengujian terhadap kemungkinan adanya pelanggaran asumsi klasik tersebut. Secara manual, dalam melakukan uji asumsi klasik regresi linier, terlebih dahulu harus mendapatkan data *residual*. Pengujian asumsi klasik menggunakan data residual, bukan data pengamatan, kecuali uji asumsi multikolinieritas. Di antara asumsi-asumsi tersebut, ada asumsi penting yang harus dipenuhi, yaitu asumsi tidak adanya heteroskedastisitas. Heteroskedastisitas adalah adanya ketidaksamaan varians dari *residual (error)* dari pengamatan ke pengamatan pada model regresi (Kurniawan dan Yuniarto, 2016).

Heteroskedastisitas lebih sering terjadi pada data silang (*cross-section*). Dalam data *cross-sectional* biasanya dikaitkan dengan anggota populasi pada waktu tertentu, seperti konsumen orang ataupun keluarganya, industri, industri, ataupun subdivisi geografis semacam negeri bagian, negeri, kota, dll. Tidak hanya

itu, anggota ini bisa jadi dari dimensi yang berbeda, semacam industri kecil, menengah, ataupun besar ataupun berpenghasilan rendah, menengah, ataupun besar. sedangkan dalam *time series*, variabel cenderung mempunyai urutan yang seragam sebab seorang umumnya mengumpulkan informasi buat entitas yang sama sepanjang periode waktu tertentu. Contohnya merupakan produk nasional bruto (GNP), pengeluaran mengkonsumsi, tabungan, ataupun pekerjaan di Amerika Serikat, katakanlah, buat periode 1955- 2005 (Gujarati dan Porter, 2008).

Model regresi dengan variansi residual heteroskedastis memiliki istilah kesalahan terdistribusi normal, dan variansinya tidak konstan untuk semua pengamatan. Sebaliknya, model dengan variansi sisa dari *mean square error* memiliki istilah *error* terdistribusi normal, dan variansinya mencakup semua pengamatan (Sarwoko, 2016).

Asumsi yang melanggar homoskedastisitas disebut heteroskedastisitas, yang artinya variansi kesalahannya tidak konstan. Hasil heteroskedastisitas dapat menyebabkan estimasi MKT yang diperoleh masih memenuhi persyaratan yang tidak bias, namun varian yang diperoleh tidak efisien yang berarti variansi cenderung membesar sehingga tidak lagi merupakan varian yang kecil. Oleh karena itu, perlu dilakukan penyempurnaan model terlebih dahulu untuk menghindari hilangnya pengaruh heteroskedastisitas (Gujarati, 2007).

Heteroskedastisitas bisa muncul akibat adanya pencilan. Pengamatan terluar, atau pencilan merupakan pengamatan yang sangat berbeda dengan pengamatan dalam sampel (sangat kecil atau sangat besar). Lebih tepatnya, pencilan adalah pengamatan dari populasi yang berbeda untuk menghasilkan pengamatan untuk sampel yang tersisa. Memasukkan atau mengecualikan pengamatan semacam itu, terutama dalam kasus ukuran sampel yang kecil, dapat sangat mengubah hasil analisis regresi (Gujarati dan Porter, 2008).

Pencilan tidak dapat diabaikan begitu saja dari pengamatan. Draper dan Smith (1992), mengutarakan bahwa terkadang pencilan juga akan memberikan informasi yang tidak bisa diberikan oleh titik data lainnya yang mungkin saja sangat penting dan perlu diselidiki lebih jauh. Pencilan juga dapat diselesaikan dalam analisis regresi dengan menggunakan regresi kekar (Ryan, 1997). Regresi

robust merupakan metode regresi yang digunakan untuk menganalisis data yang mengandung pencilan yang berpengaruh pada model.

Metode *bootstrap* klasik adalah kumpulan teknik pengambilan sampel ulang yang dirancang untuk memperkirakan kesalahan standar dan interval kepercayaan. Memanfaatkan banyak sampel yang diambil dari pengamatan awal (Stein, 1990). Metode *bootstrap* klasik telah menjadi metode standar dalam statistik modern. Penelitian ini dapat ditelusuri kembali ke tahun 70-an dari konsep resampling. Penelitian yang dilakukan Efron (Efron, 1981) memberikan sintesis dari beberapa ide awal tentang pengambilan sampel ulang, dan tak terelakkan memberikan referensi baru untuk simulasi berdasarkan analisis statistik. Ide dasar *bootstrap* adalah membangun data palsu menggunakan informasi dari data mentah. Namun demikian kita tetap harus memperhatikan sifat-sifat data asli agar data citra memiliki fitur yang sedekat mungkin dengan data aslinya.

Seperti yang diketahui, metode *bootstrap* klasik adalah prosedur umum yang baik untuk memperkirakan distribusi sampling di bawah model distribusi seragam independen (i.i.d). Dalam analisis regresi, metode *boot* yang banyak digunakan adalah *bootstrap* klasik (Rana 2011).

Pada penelitian tentang metode *jackknife*, *bootstrap* dan metode resampling lainnya dalam analisis regresi, Wu (1986) mengatakan metode *bootstrap* klasik tidak sesuai untuk kemungkinan varians error yang heteroskedastisitas. Wu (1986) memberikan usulan bahwa metode *bootstrap* yang sesuai untuk kemungkinan varians *error* yang heteroskedastisitas. Metode *wild bootstrap* (*weighted bootstrap*) yang dapat digunakan untuk memperoleh standard *error* yang secara asimtotik benar di hadapan heteroskedastisitas. Hasil dari penelitian tersebut berupa gambaran yang berbeda dari metode *bootstrap* oleh Efron dan Tibshirani (1993), terutama yang berkaitan dengan masalah *independent identically distributed* (i.i.d). Penelitian selanjutnya mengenai *wild bootstrap* dengan judul "Bootstrap Procedures Under Some Non- I.I.D. Models" (Liu, 1988), dan dengan judul "Bootstrap and Wild Bootstrap for High Dimensional Linear Models" (Mammen, 1982).

Penelitian yang dilakukan oleh Adnan et al (2015) menggunakan data nyata dan studi simulasi, yang mana penelitian ini bertujuan melihat dari beberapa metode mana yang lebih efisiensi apabila dalam suatu data memiliki pencilan dan heteroskedesitas. Penelitian ini mendapatkan hasil bahwa ketika suatu data ada pencilan dan heteroskedesitas menunjukkan bahwa metode *bootstrap* yang efisien adalah RBootWu dan RBootLiu. Pada penelitian ini dapat disimpulkan pula *robust wild bootstrap* merupakan alternative yang sangat baik dibanding *wild bootstrap* terutama ketika data memiliki persentasi pencilan yang tinggi.

Astari (2014) telah melakukan penelitian untuk mengatasi bias pada penduga analisis regresi dengan menggunakan metode *bootstrap* klasik. Pada penelitiannya, Astari (2014) dengan menggunakan data simulasi berupa data yang dihasilkan oleh program R-2.15.3. Kemudian dilakukan simulasi data dengan menginput 5%, 10%, dan 15% pencilan pada data awal yang tidak mengandung pencilan. Hasil penelitian ini menunjukkan bahwa setelah menggunakan *bootstrap* klasik, hanya pencilan 5% dari total data yang dapat menghasilkan estimasi parameter dan interval kepercayaan 95% dan 99% yang mendekati estimasi awal, sehingga nilai estimasi tidak bias. Pada penduga parameter dimana nilai outlier 10% dan 15% tidak berada pada interval kepercayaan yang dikehendaki.

Penelitian yang dilakukan oleh Pakpahan et al (2018) memiliki konsep yang sama dengan penelitian Astria, dimana sama-sama mengatasi bias dari penduga parameter analisis regresi tetapi menggunakan metode *pairs bootstrap*. Penelitian ini menggunakan data yang dibangkitkan dari aplikasi Matlab 2009b. data yang telah dibangkitkan akan ditambah pencilan sebesar 5%, 10%, dan 15% dimana data awal yang didapat tidak terdapat pencilan. Hasil yang didapat dari penelitian adalah dengan menggunakan metode *bootstrap* berpasangan penduga parameter yang dihasilkan adalah *Best Linier Unbiased Estimaor* (BLUE) dengan interval kepercayaan 99%. Nilai bias yang diperoleh dengan menggunakan metode *bootstrap* data berpasangan pada setiap pengulangan mampu mengkoreksi sedikit nilai bias yang dihasilkan melalui penduga parameter dengan MKT, untuk tiap nilai parameter dan untuk setiap nilai pencilan yang diberikan (Pakpahan dan Simamora, 2018).

Penelitian yang dilakukan oleh Pakpahan dan Astria hanya menggunakan *bootstrap* klasik, yang mana pada penelitiannya dilakukan tanpa adanya kombinasi dengan metode lain. Penggunaan metode *bootstrap* klasik pada penelitian yang dilakukan oleh Astari dan Pakpahan sudah menghasilkan nilai terbaik dikarenakan kasus yang diangkat oleh Astria dan Pakpahan hanya data yang ditambahkan pencilan tanpa hadirnya heteroskedastisitas. Data yang digunakan pada kedua penelitian adalah data bangkitan. Kasus seperti ini metode *bootstrap* klasik sudah memberikan hasil yang paling baik.

Penelitian lain yang dilakukan oleh Rana (2011) mengenai *robust wild bootstrap* untuk menstabilkan variansi estimasi parameter pada model regresi heteroskedastik dengan adanya outlier, menunjukkan bahwa metode *robust wild bootstrap* memberikan hasil yang lebih baik dibandingkan dengan metode *wild bootstrap* jika terdapat pencilan (*outlier*) dalam data pengamatan yang heteroskedastisitas.

Metode *Bootstrap* klasik tidak akan menghasilkan nilai terbaik apabila hadirnya heteroskedastisitas dan pencilan, dengan itu dari penelitian yang dilakukan oleh Adnan dan Rana menunjukkan metode apa yang menghasilkan kesimpulan terbaik dengan data yang mengandung heteroskedastisitas dan pencilan. Hasil yang didapat adalah menggunakan metode *robust wild bootstrap* yang diperkenalkan oleh Wu dan Liu.

Keefesiesian suatu metode dapat diukur dengan melihat nilai standar error, semakin minimum suatu nilai standar error yang dihasilkan dari suatu penduga parameter, maka penduga parameter tersebut adalah yang terbaik dibanding penduga parameter lainnya.

Hal tersebut yang menjadi landasan dari penelitian ini, dimana penelitian ini akan menggunakan metode *robust wild bootstrap* untuk mengestimasi parameter regresi dengan adanya heteroskedastisitas dan pencilan. Bertujuan untuk memperoleh *estimator* (penduga) yang efisien. Berdasarkan penjelasan tersebut maka peneliti akan melakukan penelitian dengan judul "Penduga Parameter Model Regresi Linier Berganda Hadirnya Heteroskedastisitas Dan Pencilan Dengan Metode *Robust Wild Bootstrap*".

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana algoritma *bootstrap* klasik dan *robust wild bootstrap* untuk mengestimasi kesalahan baku (*standard error*) dari penduga parameter model regresi linier berganda dengan adanya heteroskedastisitas dan pencilan ?
2. Bagaimana efisiensi penduga parameter dari metode estimasi *bootstrap* klasik dan *robust wild bootstrap*?

1.3 Batasan Masalah

Agar penelitian yang dilakukan lebih terarah dan topik pembahasan tidak meluas, maka peneliti menyusun batasan masalahnya sebagai berikut:

1. Parameter yang diduga adalah parameter pada model regresi linier berganda dengan menggunakan pendekatan *robust wild bootstrap*.
2. Masalah yang diteliti adalah permasalahan proses estimasi yang tidak efisien sehingga menimbulkan variansi *error* yang tidak identik pada model regresi linier berganda dengan asumsi heteroskedastisitas dan penambahan pencilan.
3. Data awal yang digunakan sebanyak 20 sampel.
4. Data yang digunakan adalah data pembangkitan yang mengandung heteroskedastisitas dan pencilan.
5. Metode yang dilakukan untuk mengatasi masalah dalam penelitian ini adalah metode *robust wild bootstrap*. Sebagai pembandingan, dilakukan pula metode *bootstrap* klasik.
6. Menghitung estimasi pada data yang robust hanya menggunakan *S-estimator*.
7. Melihat keefesiensian suatu penduga dengan efisiensi relatif

1.4 Tujuan Penelitian

Berdasarkan rumusan dan batasan masalah yang dibuat, maka yang menjadi tujuan penelitian ini adalah sebagai berikut:

1. Menerapkan algoritma bootstrap klasik dan *robust wild bootstrap* untuk mengestimasi kesalahan baku (*standard error*) dari penduga parameter model regresi linier berganda dengan adanya heteroskedastisitas dan pecilan.
2. Menganalisis efisiensi penduga parameter dari metode estimasi *bootstrap* klasik dan *robust wild bootstrap*.

1.5 Manfaat Penelitian

Dengan dilakukannya penelitian ini diharapkan dapat memberi manfaat sebagai berikut:

1. Bagi peneliti, penelitian ini diharapkan menambah wawasan pengetahuan tentang estimasi *robust wild bootstrap* dalam mengatasi heteroskedastisitas dan pecilan pada estimasi parameter analisis regresi.
2. Bagi pembaca, memberikan informasi mengenai estimasi *robust wild bootstrap* dalam mengestimasi parameter regresi linier berganda dengan asumsi heteroskedastisitas dan memberikan masukan bagi penelitian selanjutnya yang berkaitan dengan masalah tersebut.
3. Bagi peneliti ataupun pembaca akan mengetahui metode yang lebih baik digunakan dengan kehadiran heteroskedastisitas dan pecilan, selain itu dengan adanya metode ini mempermudah peneliti mencari data, menghemat waktu, dan menghemat biaya penelitian.