

BAB III METODOLOGI PENELITIAN

3.1. Metode Penelitian

Penelitian ini menggunakan metode *Ordered Logit Model* (OLM) dalam menguji dan menjelaskan variabel-variabel dalam model ekonominya. OLM digunakan karena variabel dependen dalam model ini merupakan variabel diskret kualitatif bertingkat yang dibagi ke tiga tingkat yaitu: kejahatan tinggi-sedang-dan rendah. Sedangkan variabel independennya adalah satu variabel *dummy* dan enam variabel *continuous*.

3.2. Data yang Digunakan

3.2.1. Sampel dan Sumber Data

Studi ini dilakukan terhadap kabupaten dan kota di Pulau Jawa sehingga bisa didapatkan gambaran menyeluruh tentang kondisi ekonomi, sosial, demografi, dan tingkat kejahatan di masing-masing kota dan kabupaten tersebut. Namun, data-data faktual yang dibutuhkan dalam penelitian ini tidak tersedia. Oleh karenanya, peneliti menggunakan data Survey Sosial Ekonomi Nasional (Susenas) tahun 2007 yang menyediakan sampel dari sebagian besar data tersebut. Sisanya, variabel kekuatan kepolisian resor tahun 2007, menggunakan data dari Mabes Polri.

Data Susenas tahun 2007 didapatkan dari Badan Pusat Statistik (BPS) melalui Lembaga Demografi, sementara data kekuatan kepolisian resor didapatkan dari Deputi Sumber Daya Manusia Kapolri. Penelitian ini menggunakan data *cross section* kabupaten dan kota se-Jawa pada tahun 2007. Pemilihan tahun 2007 disebabkan data Susenas terakhir yang peneliti mampu dapatkan adalah data pada tahun 2007.

3.2.2. Jenis Data

Data yang digunakan dalam penelitian ini adalah data *cross section* yaitu data yang didapatkan dengan mengobservasi variabel-variabel dalam waktu yang sama

sehingga menggambarkan suatu kondisi dalam satu titik tertentu. Data *crosssection* yang digunakan adalah data sampel 110 kabupaten di Jawa (dengan mengeluarkan Jakarta) tahun 2007 yang didapatkan dari Susenas 2007 sedangkan data kekuatan polisi tingkat resor di seluruh Jawa didapatkan dari Mabes Polri. Software yang digunakan oleh penulis untuk mengolah data tersebut adalah Stata 10 dan Weka 3.4.15.

3.3 Spesifikasi Model

Model yang digunakan oleh penulis dalam penelitian ini berasal dari model Doyle, Ahmed & Horne (1999). Studi ini juga menggabungkan penelitian Doyle, Ahmed & Horne dan modifikasi model oleh Husnayain. Meskipun ekonomi kejahatan mengembangkan banyak model yang dapat menjelaskan korelasi antara variabel ekonomi-sosial dengan tingkat kejahatan, model Doyle, Ahmed & Horne (1999) dianggap merupakan model yang paling *feasible* sehingga dapat digunakan di Indonesia pada saat ini. Keterbatasan dan ketidakterediaan data belum memungkinkan model-model ekonomi kejahatan lainnya dapat digunakan pada penelitian ekonomi kejahatan di Indonesia.

Penulis mengganti variabel koefisien Gini dengan tingkat kemiskinan yang secara teori cukup representatif menunjukkan pengaruh terhadap tingkat kejahatan. Variabel penyelesaian kasus kejahatan ditiadakan karena tidak berhubungan dengan variabel dependen tindak kejahatan yang didapat dari Susenas 2007. Selain itu, penulis menambahkan variabel independen *dummy* kota atau bukan kota dan daerah dengan proporsi keluarga *single mother* tinggi atau rendah. Modifikasi tersebut menghasilkan model sebagai berikut:

$$PCrime = (\beta_1CTY + \beta_2SALi + \beta_3POVi + \beta_4URi + \beta_5YMENi + \beta_6SINGi + \beta_7POLi) + \alpha_i + v_i$$

Di mana:

PCrime = tingkat kejahatan properti di suatu daerah kabupaten/kota, didapat dari penduduk terkena tindak kejahatan per total penduduk

CTY = variabel *dummy*, dimana kota = 1, bukan kota = 0

- SAL_i = upah rata-rata per bulan pekerja di kabupaten/kota *i*
- POV_i = tingkat kemiskinan di kabupaten/kota *i*
- UR_i = tingkat pengangguran di kabupaten/kota *i*
- YMEN_i = proporsi pria usia 15-29 tahun dalam populasi penduduk di kabupaten/kota *i*
- SING_i = proporsi keluarga *single mother* di kabupaten/kota *i*, dibuat diskret di mana daerah dengan tingkat keluarga *single mother* yang rendah = 0 dan daerah dengan tingkat keluarga *single mother* yang tinggi = 1
- POL_i = jumlah polisi resor yang bertugas untuk 10.000 penduduk di kabupaten/kota *i*

3.3.1. Penjelasan Variabel dalam model

3.3.1.1. Variabel Independen

Variabel independen adalah variabel yang dalam model akan mempengaruhi variabel dependen. Dalam modifikasi model yang dilakukan, variabel independen yang mempengaruhi variabel dependen (tingkat kejahatan) adalah variabel upah rerata (SAL), variabel *dummy* berupa kota atau bukan kota (CTY), tingkat pengangguran (UR), tingkat kemiskinan (POV), proporsi pria usia 15-29 tahun dalam populasi (YMEN), jumlah polisi resor yang bertugas per 10 ribu penduduk (POL), dan proporsi keluarga *single mother* dalam populasi penduduk (SING).

Variabel *dummy* kota atau bukan kota dipilih untuk melihat seberapa besar pengaruh kota terhadap tingkat kejahatan di Jawa. Seperti yang telah diuraikan sebelumnya, kejahatan lebih banyak terjadi di kota karena daerah perkotaan menawarkan ekspektasi hasil kejahatan yang lebih besar. Oleh karenanya diharapkan terdapat hubungan positif antara daerah kota dengan tindak kejahatan. Variabel tingkat pengangguran digunakan untuk melihat seberapa besar ketiadaan pekerjaan yang berarti ketiadaan penghasilan mempengaruhi tingkat kejahatan. Menurut teori ketiadaan pekerjaan akan memperkecil *opportunity cost* dari pelaku kejahatan dan

meningkatkan ekspektasi harta rampasan kejahatan sehingga diperkirakan terdapat hubungan yang positif antara tingkat pengangguran dengan tingkat kejahatan.

Idealnya, variabel yang digunakan dalam penelitian ini adalah koefisien Gini yang menjelaskan tingkat kesenjangan pendapatan di daerah kota/kabupaten. Namun, ketidaktersediaan koefisien Gini pada tingkat kabupaten mengharuskan variabel tersebut diganti dengan variabel lain. Penulis menggunakan variabel yang paling mendekati dan cukup relevan terhadap penelitian ini yaitu tingkat kemiskinan di kota/kabupaten. Variabel kemiskinan diperkirakan berpengaruh positif terhadap tindak kejahatan karena kemiskinan memberikan ekspektasi harta rampasan yang lebih besar dengan *opportunity cost* dipenjara yang lebih kecil.

Terdapat dua variabel demografi yang digunakan dalam penelitian ini yaitu: proporsi pria usia 15-29 tahun di populasi penduduk dan proporsi keluarga *single mother*. Kedua variabel tersebut diperkirakan akan berpengaruh positif terhadap tingkat kejahatan di suatu kota/kabupaten. Idealnya variabel *single mother* dalam penelitian ini adalah data yang menggambarkan profil *single mother* yang bermasalah dalam membesarkan anak, sebagaimana peneliti di AS menggunakan data *single mother* ras kulit hitam dan data pengajuan aborsi yang ditolak di suatu daerah. Sayangnya penulis belum menemukan literatur yang menjelaskan profil keluarga *single mother* tersebut di Indonesia. Oleh karenanya, penulis membuat variabel *single mother* menjadi variabel diskret yang kategorik, yaitu tingkat proporsi *single mother* rendah dan tinggi. Pengkategorian tersebut didasarkan pada kedekatan antar data yang digunakan pada penelitian. Diskretisasi data dilakukan dengan *software* Weka 3.4.15 melalui metode *unsupervised* sehingga didapatkan:

- a. Proporsi *single mother* rendah apabila proporsi kurang dari 17.625 persen, terdapat 91 kabupaten/kota dan dalam model dinotasikan sebagai nol (0).
- b. Proporsi *single mother* tinggi apabila proporsi lebih dari 17.625 persen, terdapat 19 kabupaten/kota dan dalam model dinotasikan sebagai satu (1).

Variabel terakhir yang ingin diteliti adalah proporsi petugas kepolisian perkapita. Kekuatan kepolisian secara teori akan mempengaruhi permintaan kejahatan di suatu daerah, di mana semakin banyak jumlah polisi maka jumlah kejahatan akan semakin berkurang.

3.3.1.2. Variabel Dependen

Variabel dependen yang digunakan dalam penelitian ini adalah tingkat kejahatan properti pada tiap-tiap daerah kabupaten/kota di Jawa. Tingkat kejahatan properti tersebut diukur berdasarkan sampel yang diambil oleh Susenas tahun 2007 yaitu jumlah individu yang pernah menjadi korban kejahatan per total responden di tiap daerah kabupaten/kota. Pendekatan ini merupakan data dengan pendekatan terbaik yang dapat menghindari masalah *undervalue* yang umumnya terjadi pada data yang dimiliki oleh polisi.

Oleh karena penelitian ini menggunakan *ordered logit model* maka variabel dependen yang digunakan pun bersifat diskret yaitu tingkat kejahatan tinggi, sedang, atau rendah. Variabel dependen dalam model *ordered logit model* ini bersifat ordinal berurut dengan urutan kabupaten/kota dengan tingkat kejahatan yang rendah dengan $Y=0$, tingkat kejahatan sedang $Y=1$, dan kabupaten/kota tingkat kejahatan tinggi dengan $Y=2$. Oleh karena dalam ilmu kriminologi sendiri tidak terdapat teori yang mendiferensiasi tingkat kejahatan tinggi, sedang, dan rendah maka penentuannya dilakukan melalui kedekatan berdasarkan kedekatan antar nilai data yang ada. Penentuan rendah, sedang, dan tinggi ditentukan oleh software Weka 3.4.15 melalui metode *unsupervised discretiz* dan didapatkan hasil sebagai berikut:

- a. Kategori rendah, dengan tingkat kejahatan dibawah 4.45 persen terdapat 86 daerah kabupaten/kota.
- b. Kategori sedang, dengan tingkat kejahatan antara 4.45 hingga 8.34 persen terdapat 21 daerah kabupaten/kota.
- c. Kategori tinggi, dengan tingkat kejahatan diatas 8.34 persen terdapat 3 daerah kabupaten/kota.

Data kejahatan properti yang dimaksud pada variabel dependen ini adalah kejahatan pencurian, perampokan, dan penipuan yang diderita oleh responden Susenas 2007 pada setahun terakhir.

3.4. Pengolahan dan Pengujian Data

3.4.1. Qualitative Response Regression Model

Apabila regresi liner dipaksakan pada model dengan variabel dependen yang bersifat diskret, estimator yang dihasilkan tidak bersifat BLUE (*Best Linier Unbiased Estimator*). Hal ini disebabkan: varian error-nya tidak terdistribusi normal, estimator tidak efisien akibat heteroskedastis, dan R^2 tidak dapat digunakan sebagai pengukur *Goodness of Fit*. Oleh karenanya, untuk menghasilkan estimator persamaan yang BLUE, penelitian ini menggunakan *qualitative response regression model*.

Terdapat tiga pendekatan untuk mengembangkan model yang menjelaskan model regresi *binary response* yaitu:

- a. Linear Probability Model
- b. Logit Model
- c. Probit model

Linear Probability Model (LPM) merupakan metode regresi yang umum digunakan sebelum logit dan probit model dikembangkan. LPM bekerja dengan dasar bahwa variabel respon Y, yang merupakan probabilita terjadinya sesuatu, mengikuti *Bernoulli probability distribution* dimana:

Y_i	Probability
0	$1 - P_i$
1	P_i
Total	1

Namun, karena LPM bekerja berdasarkan metode OLS biasa maka timbul permasalahan yang telah diungkapkan sebelumnya: *non-normality of the disturbance*, *heteroscedastis*, tidak terpenuhinya ekspektasi nilai Y antara satu sampai dengan nol, dan tidak dapat digunakannya R^2 sebagai pengukur *Goodness of Fit*. Kebutuhan akan

model probabilita yang menghasilkan Y yang terletak antara interval satu sampai dengan nol dengan hubungan antara P_i dengan X_i yang tidak linear menyebabkan logit model dikembangkan. Model logit dimulai dari persamaan:

$$P_i = E(Y = 1 | X_i) = \frac{1}{1 + e^{-(\beta_1 + \beta_2 X_i)}} \dots\dots\dots(1)$$

Di mana $Z_i = \beta_1 + \beta_2 X_i$, persamaan di atas kemudian dapat disederhanakan menjadi

$$P_i = \frac{1}{1 + e^{-Z_i}} = \frac{e^z}{1 + e^z} \dots\dots\dots(2)$$

Persamaan dua lebih dikenal sebagai *logistic distribution function*. Persyaratan yang diminta sebelumnya, yaitu model probabilita yang menghasilkan Y antara interval satu sampai dengan nol dengan hubungan antara P_i dengan X_i yang tidak linear, dapat terpenuhi. Hal ini disebabkan, saat Z berkisar antara $-\infty$ sampai dengan ∞ , P_i berkisar antara 0 dan 1 sehingga P_i tidak berhubungan linear dengan Z. Meskipun begitu masih terdapat masalah estimasi karena P tidak hanya tidak linier pada X tetapi juga ke β . Namun, seperti dapat ditunjukkan pada persamaan berikut, masalah estimasi tersebut dapat diatasi.

$$1 - P_i = \frac{1}{1 + e^{Z_i}} \dots\dots\dots(3)$$

$$\frac{P_i}{1 - P_i} = \frac{1 + e^{Z_i}}{1 + e^{-Z_i}} = e^{Z_i} \dots\dots\dots(4)$$

Persamaan di atas yaitu $P_i/1-P_i$ dikenal dengan *odds ratio* yaitu rasio probabilita memiliki rumah terhadap rasio tidak memiliki rumah. Apabila persamaan di atas di *log naturalkan* maka akan didapatkan persamaan:

$$L_i = \ln\left(\frac{P_i}{1 - P_i}\right) = Z_i = \beta_1 + \beta_2 X_i \dots\dots\dots(5)$$

L, atau log dari *odds ratio* tidak hanya bersifat linear pada X tetapi juga bersifat linear terhadap parameter. Persamaan tersebut yang kemudian dikenal sebagai model logit. Kelebihan dari model logit tersebut adalah:

1. Saat P berpindah dari 0 ke 1, logit L akan berpindah dari $-\infty$ ke ∞ . Oleh karena itu, meskipun probabilitas terletak antara 0 hingga 1, logit sendiri tidak terbatas. Dan meski L linear terhadap X, probabilitasnya sendiri tidak.
2. L (logit) yang bernilai positif menandakan bahwa meningkatnya nilai regresor akan menyebabkan meningkatnya odds dari regresi yang setara dengan 1. Sebaliknya, L (logit) yang bernilai negatif menandakan bahwa menurunnya odds dari regresi yang setara dengan 1 akan menyebabkan meningkatnya nilai dari X.
3. Model logit yang diberikan pada persamaan lima dapat diinterpretasikan sebagai berikut: slope β_2 merupakan pengukur perubahan nilai L karena perubahan nilai X, sementara *Intercept* β_1 merupakan nilai dari log-odds apabila nilai suatu slope nol. Logit model juga mengasumsikan bahwa log sebuah *odds ratio* berhubungan linier terhadap Xi atau nilai sebuah slope.

Analisa regresi logistik bekerja dengan menggunakan variabel penjelas, kategorik atau numerik, untuk mengestimasi probabilitas terjadinya sebuah kejadian tertentu. Permodelannya dilakukan melalui formulasi transformasi logit berupa:

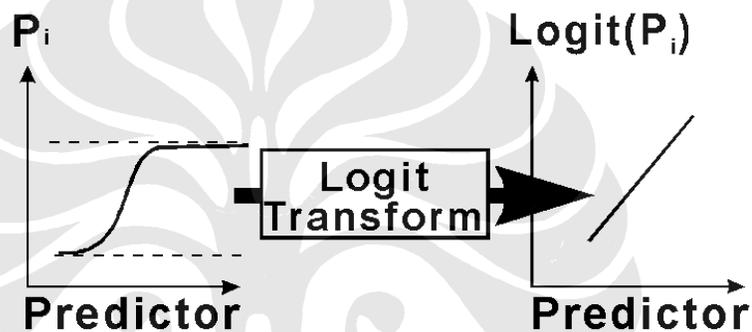
$$\text{Logit}(\pi_i) = \log_e \left(\frac{\pi_i}{1 - \pi_i} \right)$$

Persamaan diatas dapat diterangkan sebagai berikut: π_i merupakan probabilitas terjadinya kategori sukses dari sebuah kejadian. Probabilitas sukses ini berdasarkan pengaruh variabel penjelas terhadap variabel dependen pada orang ke-i. \log_e merupakan suatu logaritma dengan basis bilangan e. Dalam penelitian ini, model tersebut digunakan untuk meneliti kejadian kategori sukses terjadinya tingkat kejahatan yang tinggi di sebuah kabupaten/kota dibandingkan kejahatan rendah atau sedang. Cara kerja umum *conditional probability model* ini adalah (Greene, 1990)

$$Prob(\text{event } j \text{ occurs}) = prob(Y=f) = F[\text{relevant effects, parameters}]$$

Persamaan di atas dapat diartikan bahwa probabilitas terjadinya kejadian j dipengaruhi oleh variabel-variabel independen dalam model yang diperkirakan relevan mempengaruhi peristiwa tersebut. Proses transformasi logit diilustrasikan pada gambar 3-1

Gambar 3-1
Transformasi Logit



Sumber: Modul Regresi Logistik, Statistika IPB, 2007

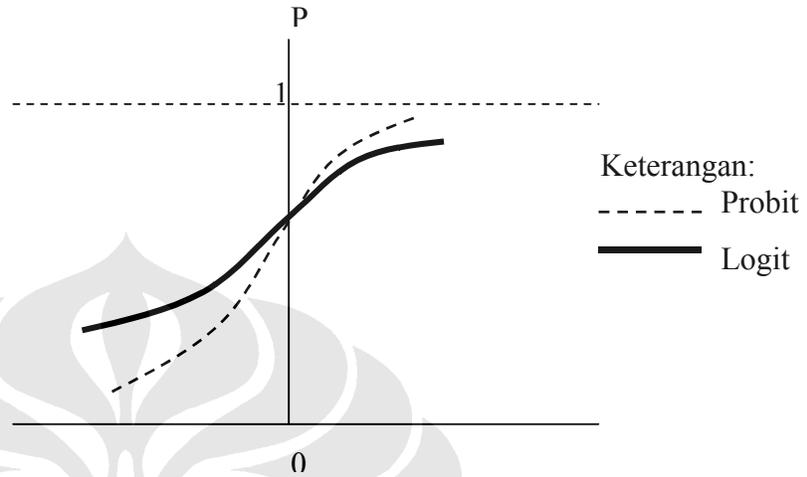
Apabila *error term*, selisih antara nilai variabel yang diestimasi dengan nilai variabel sebenarnya, terdistribusi secara normal maka regresi probabilitas dapat menggunakan model probit. Model probit dapat dituliskan sebagai berikut:

$$Y_i = x_i\beta + u_i$$

Di mana $u_i \sim N(0,1)$ atau *error* mengikuti distribusi normal dengan rerata 0, dan varians konstan antar variabel independen bernilai 1. Dalam model logit ini, distribusi Y bersifat normal dengan variannya yang konstan. Apabila *error term* tidak terdistribusi normal, atau dikenal dengan distribusi logistik, maka yang digunakan adalah model logit. Perbedaan antara keduanya dapat dilihat dari Gambar 3-2 dan Tabel 3-1

Gambar 3-2

Distribusi Kumulatif Logit dan Probit



Sumber: Gujarati, Basic Econometrics. 2003

Dari gambar di atas dapat dilihat bahwa *conditional probability* P_i pada logit lebih mendekati 0 dibandingkan probit, dan mendekati 1 sedikit lebih lambat daripada probit. Pada tabel yang diambil dari Gujarati, diperlihatkan bahwa model logit merupakan *cummulative logistik* sementara probit adalah *probit cummulative normal*.

Meskipun *cummulative normal* dan *cummulative logistik* bernilai tengah yang sama yaitu nol namun variannya memiliki nilai yang berbeda. *Cummulative normal* memiliki varian bernilai 1 dan *cummulative logistik* bernilai $\pi^2/3$. Nilai koefisien logit dapat diperoleh dengan mengalikan nilai koefisien probit dengan $\pi^2/3(1.81)$ sedangkan nilai koefisien probit dapat diperoleh dengan mengalikan nilai koefisien logit dengan $\pi^2/3(1.81)$.

Tabel 3-1
Nilai dari Cumulative Probability Function

Z	Cummulative normal	Cummulative Logistik
	$P_1(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-s^2/2} ds$	$P_2(Z) = \frac{1}{1 + e^{-z}}$
-2.0	0.0228	0.1192
-1.0	0.1587	0.2689
0	0.5000	0.5000
1.0	0.8413	0.7311
2.0	0.9772	0.8808

Sumber: Gujarati, Basic Econometrics. 2003

Pada prinsipnya, kedua model menggunakan jenis metode kerja yang sama. Namun, model logit lebih populer untuk digunakan dalam penelitian karena persamaan matematikanya yang lebih sederhana.

3.4.2. Klasifikasi dalam Model Logit

Apabila diklasifikasi berdasarkan jenis variabel independennya, terdapat dua tipe model logit yaitu: *unordered logit model* dan *ordered logit model*. Kedua model tersebut akan dijelaskan sebagai berikut:

A. *Unordered Logit Model*

Model logit jenis ini digunakan saat variabel dependennya merupakan pilihan yang tidak bertingkat dan tidak saling mempengaruhi satu sama lainnya. Contoh hari dalam satu minggu yaitu: Senin- Selasa- Rabu- Kamis, dst.

B. *Ordered Logit Model*

Model logit *ordered* adalah model yang variabel dependennya merupakan pilihan bertingkat di mana pilihan yang satu lebih baik atau lebih buruk terhadap pilihan lain. Contoh ,terdapat empat pilihan yaitu: tidak bagus – cukup bagus - bagus – bagus sekali, atau dalam ranking pendidikan yaitu: SD – SMP – SMA– Perguruan Tinggi, dan lain-lain.

Pada penelitian ini penulis menggunakan model *ordered logit*. Hal ini disebabkan, pada variabel dependen tingkat kejahatan properti di sebuah kabupaten/kota, kejahatan properti rendah dianggap lebih baik daripada tingkat kejahatan sedang dan tinggi. Apabila variabel dependen Y merupakan variabel dengan skala ordinal yang memiliki c kategori maka $P(Y \leq j)$ merupakan probabilita Y berada pada kategori j atau kategori di bawahnya. Penjumlahan antara probabilita Y berada pada kategori 1 sampai dengan Y berada pada kategori j disebut probabilita kumulatif yang secara matematis dapat dituliskan sebagai berikut:

$$P(Y \leq j) = P(Y = 1) + P(Y = 2) + \dots + P(Y = j)$$

Berdasarkan persamaan sebelumnya maka peluang kumulatif variabel dependen Y yang berada pada kategori 1 sampai c memiliki hubungan yang dapat direpresentasikan sebagai berikut:

$$P(Y \leq 1) \leq P(Y \leq 2) \leq \dots P(Y \leq c) = 1.$$

Dari itu, dapat dikemukakan konsep mengenai odds ratio yaitu rasio antara:

$$\frac{P(Y \leq j)}{P(Y > j)}$$

Odd ratio merupakan *odd* peubah respon pada kategori j atau di bawah j. Apabila rasio bernilai tiga maka diartikan probabilita variabel dependen Y, yang

berada pada kategori j atau di bawahnya, adalah sebesar 3 kali lebih besar dibandingkan peluang variabel dependen Y berada pada kategori di atas kategori j . Model yang variabel dependennya berskala ordinal dan menggunakan logit dari peluang kumulatif disebut logit kumulatif. Logit untuk $P(Y \leq j)$ merupakan log dari rasio yang dapat ditulis sebagai berikut:

$$\text{logit}[P(Y \leq j)] = \log \frac{P(Y \leq j)}{P(Y > j)} ; j = 1, 2, \dots, c-1$$

Apabila dihubungkan dengan variabel independen X maka persamaan di atas dapat dituliskan berupa:

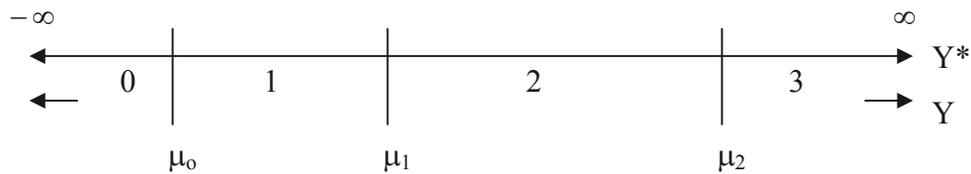
$$\text{Logit } [P(Y \leq j)] = \alpha_j + \beta X; j = 1, 2, \dots, c-1$$

Model tersebut dapat digunakan apabila asumsi pengaruh variabel dependen X untuk setiap peluang kumulatif adalah sama. Apabila tidak terpenuhi, maka dapat digunakan model regresi logistik nominal. Model logit kumulatif yang memiliki pengaruh bersama seperti ini dikenal sebagai model *odd* proporsional.

Model *ordered logit* digunakan untuk mengestimasi koefisien regresi yang dapat digunakan untuk memprediksi nilai dari *fitted probability* atau probabilitas variabel dependen Y untuk memilih nilai j dari setiap j yang mungkin. Probabilitas tertinggi didapatkan dengan adanya observasi yang masuk lebih banyak ke suatu kategori daripada kategori lainnya. Model ini dibangun melalui pendekatan laten dengan variabel tersembunyi di luar observasi penelitian, dari range $-\infty$ hingga ∞ , yang dapat digambarkan sebagai berikut:

Gambar 3-3

Pendekatan *Latent Variable* dari *Ordered Variable*



Sumber: Prafitria, 2006

Model struktural *latent* yang dimaksud adalah

$$Y^* = x\beta + \varepsilon$$

Di mana Y^* adalah *latent variabel* dengan $\mu_0 = 0$. Sedangkan *latent observations* dengan variabel respon ordinal adalah sebagai berikut:

$$Y = j \text{ jika } \mu_{j-1} < Y^* \leq \mu_j \text{ untuk } j = 0, 1, 2, \dots, J$$

Di sini, μ merupakan *unknown parameter* yang diperoleh dengan estimasi β sehingga dapat diperoleh tiga kelompok kabupaten dalam penelitian ini yaitu: kabupaten dengan peluang kejahatan properti rendah ($Y=0$), kabupaten dengan peluang kejahatan properti sedang ($Y=1$), dan kabupaten dengan peluang kejahatan properti tinggi ($Y=2$), aturan di atas kemudian menjadi:

$$Y = 0 \text{ jika } Y^* \leq \mu_0, Y = 1 \text{ jika } \mu_0 < Y^* \leq \mu_1, \text{ dan } Y = 2 \text{ jika } Y^* > \mu_1$$

Oleh karenanya, model *latent* dalam bentuk probabilitas adalah sebagai berikut:

Dimana $P(Y=j | x)$ dengan

$$\begin{aligned} &= P(\mu_{j-1} < Y^* \leq \mu_j | x) \\ &= P(\mu_{j-1} < x\beta + \varepsilon \leq \mu_j | x) \\ &= P(\mu_{j-1} - x\beta < \varepsilon \leq \mu_j - x\beta | x) \\ &= P(\varepsilon < \mu_j - x\beta | x) - P(\varepsilon \leq \mu_{j-1} - x\beta | x) \\ &= F(\mu_j - x\beta) - F(\mu_{j-1} - x\beta) \end{aligned}$$

Untuk mendapatkan probabilitas yang positif sehingga $0 < \mu_1 < \mu_2 < \dots < \mu_{j-1}$

F merupakan fungsi distribusi kumulatif (*cummulative distribution function*) dari ε dengan ε yang mengikuti distribusi logistik, di mana $e =$ bilangan logaritma natural = 2,7182 sehingga distribusi kumulatifnya menjadi

$$\begin{aligned} F(x) &= (1 + \exp(-x))^{-1} \\ &= \frac{1}{1 + e^{-x}} \\ &= \frac{e^{-x}}{(a + e^{-x})^2} \end{aligned}$$

Pada *ordered logit*, dikenal metode *cutoff points*. *Cutoff points* pada gambar 3-3 dinotasikan dengan μ . Apabila terdapat J pilihan dalam variabel dependen maka *ordered logit* akan menghasilkan J-1 *cutoff* yang dinotasikan dengan k_1 sampai dengan k_{J-1} dan bernilai dari $-\infty$ hingga ∞ .

Variabel dependen diestimasi bernilai 0 apabila *fitted value* terletak di bawah k_1 , diestimasi bernilai 1 apabila *fitted value of logit* terletak antara k_1 dan k_2 , dan seterusnya. Dengan kata lain, fungsi peluang tersebut diprediksi menggunakan fungsi linier dari variabel independen dan sekumpulan *cut points* yang dapat dinotasikan sebagai berikut:

$$\begin{aligned} P(Y = j | x) &= P(k_{j-1} < \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_j x_{ji} + \varepsilon \leq k_j) \\ &= P(k_{j-1} < \sum_m \beta_m X_m + \varepsilon \leq k_j) \\ &= \frac{1}{1 + \exp(-k_j + \sum_m \beta_m X_m)} - \frac{1}{1 + \exp(-k_{j-1} + \sum_m \beta_m X_m)} \end{aligned}$$

3.4.3. Penggunaan *Maximum Likelihood* pada Logit Model

Apabila metode regresi linier biasa digunakan dalam estimasi model distribusi logistik maka estimator tidak dapat memenuhi kriteria BLUE. Oleh karena itu, pada logit model, digunakan *maximum likelihood* untuk menggantikan fungsi *least square* yang meminimumkan *error*. Penggunaan *maximum likelihood* diharapkan akan

mendekatkan nilai variabel yang diestimasi dengan nilai variabel yang sebenarnya terjadi.

Metode *maximum likelihood* bekerja dengan membentuk suatu persamaan yang menunjukkan bahwa probabilitas dari data yang diobservasi merupakan fungsi dari parameter yang diestimasi. Persamaan tersebut merupakan *log likelihood function* yang dalam *ordered logit* ditunjukkan sebagai berikut:

$$L(\beta) = \sum_{j=0}^J \sum_{i=1}^N \ln(P(Y = j | X))$$

$$L(\beta) = \sum_{j=0}^J \sum_{i=1}^N \ln(F(\mu - x\beta) - x$$

3.4.4. Pengujian, Analisa, dan Interpretasi Model

Model ekonomi kejahatan yang digunakan dalam penelitian ini adalah model *ordered logit* yang diolah dengan program STATA 10 dan WEKA 3.4.1.5. Analisis dilakukan dengan merun observasi yang terdapat dalam model *ordered logit* kemudian mengevaluasi hasil run tersebut. Setelah hasil regresi model didapat dan evaluasi dilakukan maka, secara teoritis, asumsi statistik dan ekonometrik dapat dipertanggungjawabkan secara ilmiah. Setelah itu interpretasi model dapat dilakukan untuk kemudian diambil kesimpulan.

3.4.4.1. Statistik

Pengujian statistik dilakukan untuk: menguji apakah variabel independen pada model ekonomi benar mempengaruhi variabel dependen secara signifikan, mengetahui arah dari pengaruh, dan mencari koefisien besar signifikansi tersebut.

Pengujian statistik pada model logit berbeda dengan regresi linier biasa. Apabila pengujian statistik serentak pada regresi linier menggunakan uji F-stat, pada logit model metode yang digunakan adalah *Likelihood Ratio*. Pada uji parsial pun, model logit menggunakan uji Z-stat sementara regresi linier biasa menggunakan uji t-stat. Untuk uji *goodness of fit*, Logit model menggunakan *Count R-Square* dan *Mc Fadden R-Square*.

a. Uji Parsial dengan Z-stat

Uji parsial dilakukan dengan uji Z-Stat untuk melihat apakah masing-masing variabel independen secara terpisah mempengaruhi variabel dependen Y. Z stat dilakukan dengan hipotesa sebagai berikut:

Ho = Variabel independen (x) tidak mempengaruhi variabel dependen (Y)

dimana $\alpha_1 = \alpha_2 = \dots = \alpha_n = 0$ (tidak signifikan)

H1 = Variabel independen (x) mempengaruhi variabel dependen (Y)

dimana terdapat i yang merupakan $\alpha_i \neq 0$ (signifikan)

Dalam menentukan menerima atau menolak Ho, nilai Z-Stat pada masing-masing variabel independen dibandingkan dengan tingkat nyata (α). Ho akan ditolak apabila Z-stat $< \alpha$. Dan Ho tidak ditolak apabila Z-stat $> \alpha$. Nilai dari α dapat dilihat dari tabel berikut.

Tabel 3-2
Nilai Tingkat Keyakinan dan Tingkat Nyata

Tingkat Keyakinan	Tingkat Nyata
95%	5% = 0.05
90%	10% = 0.1
80%	20% = 0.2

Sumber: Kharisma, 2007

b. Uji serentak dengan *Likelihood Ratio*

Likelihood ratio digunakan untuk menguji apakah semua variabel independen dalam model serentak mempengaruhi variabel dependen. Hipotesis dalam pengujian *Likelihood Ratio* tersebut adalah:

H0 = semua variabel independen secara serentak tidak mempengaruhi variabel dependen.

H1 = semua variabel independen secara serentak mempengaruhi variabel dependen.

Hipotesa 0 akan ditolak apabila probabilita *Likelihood Ratio* $< \alpha$ dan

H_0 tidak akan ditolak apabila probabilita *Likelihood ratio* tersebut $> \alpha$.

c. *Goodness of fit* dengan *R-square*

Untuk melihat seberapa besar variasi dalam variabel dependen dapat dijelaskan secara bersama-sama oleh variabel dependen, dan untuk melihat seberapa baik model dapat menjelaskan variabel dependen, maka statistik menggunakan *R-square* (R^2). Semakin tinggi nilai *R-square* maka menunjukkan model semakin mampu menjelaskan variabel dependen. Oleh karena itu, nilai *R-square* yang tinggi sangat diharapkan dalam suatu penelitian.

Namun, logika ekonomi tetap diutamakan dalam melihat *R-square* tersebut. Apabila pada data *cross-section* didapatkan nilai *R-square* rendah namun hasil pengujian *Z-stat* signifikan dan arahnya sesuai dengan teori ekonomi maka model tersebut dapat digolongkan sebagai model yang layak secara statistik. (Gujarati, 2003). Pada model logit pun, penggunaan *R-square* masih diperdebatkan. Penggunaan metode *R-Square* pada model logit dinyatakan sebagai *pseudo R-square*. *Pseudo R-square* adalah tiruan *R-square* yang digunakan untuk menggantikan *R-square* biasa. Hal ini dilakukan karena tidak adanya padanan tepat yang dapat menggantikan *R-square* biasa. (UCLA Academic Technology Services, 2007. Dalam Kharisma, 2007).

Oleh karenanya untuk menentukan kelayakan model, penelitian ini menggunakan *Z-stat* sebagai parameter utama, meskipun pengujian *R-square* akan tetap dilakukan. Opsi pengujian *R-square* yang digunakan untuk logit dalam penelitian adalah *Mc-Fadden R-Square* [yang dikatakan analog dengan *R-square* pada OLS. (*E-Views*, 1999 pada Kharisma, 2007)], *maximum likelihood R2*, *Mc Fadden R2*, *Cragg & Uhler's R2*, *McKelvey and Zavoina's R2* dan *Adjusted R2* yang dapat digunakan pada STATA 10.

3.4.4.2. Teoritis

Secara teoritis, hasil estimasi telah teruji apabila hasil penelitian telah menunjukkan bahwa arah dan nilai koefisien dari pengujian regresi sesuai dengan teori yang ada. Contoh, pada hasil penelitian didapatkan bahwa korelasi antara tingkat kekuatan polisi dengan tingkat kejahatan negatif, di mana semakin banyak polisi maka tingkat kejahatan semakin rendah. Apabila hasil tersebut konsisten dengan berbagai teori yang mendukung hasil tersebut, maka dapat dikatakan bahwa koefisien dari variabel kekuatan polisi tersebut telah teruji secara teoritis.

3.4.4.3. Ekonometrik

Asumsi yang harus dipenuhi pada model regresi logistik adalah *error* pada hasil estimasi haruslah terdistribusi normal. Asumsi tersebut tidak memerlukan pengujian khusus sehingga hampir selalu dipenuhi dalam setiap data yang digunakan dalam penelitian. Namun, pengujian multikolinearitas akan dilakukan untuk mengetahui apakah ada korelasi kuat antar variabel independen.

Tabel 3-3
Perbandingan Pengujian OLS dengan Logit Model

	OLS	Logit Model
Uji serentak	F-Stat	<i>Likelihood Ratio</i>
Uji parsial	t-stat	Z-Stat
Godness of fit	<i>R Square</i>	<i>Count R-Square</i>

3.4.4.4. Interpretasi

Hasil koefisien variabel dependen yang didapatkan dari hasil penelitian melalui model logit tidak dapat langsung diinterpretasikan seperti pada model OLS. Diperlukan transformasi logit dengan cara mentransformasi koefisien estimasi tersebut ke dalam antilog natural untuk mendapatkan *odds ratio*. *Odds ratio* tersebut kemudian diinterpretasikan sebagai nilai yang menunjukkan pengaruh perubahan variabel dependen. Sebelum dilakukan langkah ini, hanya dapat diketahui arah dari pengaruh variabel dependen yang belum dapat diinterpretasikan.