DATA DAN METODOLOGI

Kami menggunakan beberapa metode sebagai berikut untuk melakukan proyeksi PDRB dan pertumbuhan ekonomi Kabupaten Nganjuk.

A. Metode Proyeksi PDRB Kabupaten Nganjuk

1. Panel VAR

Panel Vector Autoregressive (PVAR) merupakan pengembangan model VAR yang menggunakan pendekatan panel data, sehingga memperbolehkan adanya *unobserved individual heterogeneity.* Lebih jauh, Panel VAR juga mampu menangkap adanya keterkaitan *cross-sectional* antar unit dibandingkan model VAR standar. Hal ini disebabkan ketidakmampuan model VAR standar untuk estimasi data dengan dimensi tinggi (Canova & Ciccarelli, 2013). Oleh karena itu, persamaan pada model PVAR dengan unit i adalah sebagai berikut:

$$\boldsymbol{y_{it}} = \boldsymbol{\Phi}_{\mathrm{i}}\boldsymbol{x_{it}} + \boldsymbol{\Lambda}_{\boldsymbol{i}}\boldsymbol{z_{it}} + \boldsymbol{\epsilon_{it}} \tag{1.1}$$

$\boldsymbol{\Phi_i}$ dinotasikan sebagai matriks $(M \times MP)$ yaitu koefisien dari unit itu sendiri dengan $x_{it} = \left(y'_{it-1}, \dots, y'_{it-p}\right)'$dan Λ_i adalah matriks dimensi $\left(M \times MP(N-1)\right)$ berhubungan dengan lag dari y_{jt} (∀j $\neq$ i), dengan $z_{it} = \left(x'_{1t}, \dots, x'_{i-1}, x'_{i+1,t}, \dots, x'_{Nt}\right)'$. Persamaan di atas menjelaskan bahwa model PVAR tidak hanya menangkap koefisien dari variabel endogen di unit i melainkan juga ditambah dengan informasi lain dari variabel endogen di unit j.

Analisis PVAR mengakomodasi adanya *spillovers* antar unit dari $\boldsymbol{\Lambda}$ memungkinan terdapat *dynamic interdependencies*, bersumber dari efek yang merambat dari nilai lag dari unit lainnya. Model PVAR juga memungkinkan adanya dampak *contemporaneous* pada unit i di periode yang sama dari nilai eror di unit lain akibat *shock* atau disebut *static interdependencies.* Terakhir, model PVAR mensyaratkan bahwa dinamika domestik pada setiap unit sama dengan unit-unit lain di seluruh sampel, disebut *cross-sectional homogeneity.* Restriksi ini seringkali mengharuskan pemilihan sampel bersifat homogen atau dinamika pada fundamental makroekonominya cenderung mirip (Canova & Ciccarelli, 2013).

Dalam analisis ekonomi makro, model PVAR telah beberapa kali digunakan untuk berbagai analisis. (Afin, 2023) menganalisis hubungan antara ketidakpastian ekonomi terhadap pertumbuhan ekonomi di negara ASEAN. Lebih lanjut, (Dées & Güntner, 2017) menemukan bahwa model PVAR memiliki performa yang baik dalam melakukan proyeksi, dalam hal ini adalah tingkat inflasi negara-negara Eropa.

2. *Panel VAR*: Pendekatan GMM

Seiring perkembangan metode ekonometrika, terdapat varian dari model PVAR yaitu Generalized Method of Moments (GMM)-PVAR. Model yang dikembangkan oleh (Sigmund & Ferstl, 2021) menggunakan *framework* GMM sebab mampu mengakomodasi permasalahan endogenitas dengan cara memasukkan seluruh variabel sebagai variabel endogen. Selain itu, model ini mampu menangkap mengisolasi lonjakan dari 1 *shock* eksogen sehingga menyebabkan kovariat lain tidak dipengaruhi oleh *shock* tersebut (Sigmund & Ferstl, 2021).

GMM sendiri memiliki dua instrumen, yaitu *first difference* GMM dan *system* GMM yang mana *first difference* GMM menggunakan lag dari variabel endogen sebagai instrumen (Arellano & Bond, 1991) sedangkan *system* GMM menggunakan *moment conditions* tambahan berdasarkan informasi yang didapatkan di tingkat *"levels"* (Blundell & Bond, 1998a). Pada model *first difference* GMM-PVAR kita bisa memulai dari *first difference* dari persamaan PVAR

$$\Delta^* \boldsymbol{y_{it}} = \sum_{\boldsymbol{l=1}}^{\boldsymbol{P}} \boldsymbol{A_l} \Delta^* \boldsymbol{y_{it-1}} + \boldsymbol{B} \Delta^* \boldsymbol{x_{it}} + \boldsymbol{C} \Delta^* \boldsymbol{s_{it}} + \Delta \boldsymbol{\epsilon_{it}} \tag{1.2}$$

Δ^* digambarkan sebagai transformasi yang mana pada *first difference* adalah $t \in \{p+2, \dots, T\}$ dan *forward orthogonal transformation* adlaah $t \in \{p+1, \dots, T-1\}$, $\boldsymbol{y_{i,t}}$ merupakan vektor $m \times 1$ dari variabel endogen untuk unit i pada waktu t. $\boldsymbol{y_{it-1}}$ adalah variabel endogen dengan lag. $\boldsymbol{x_{it}}$ adalah vektor $k \times 1$ dari variabel yang telah ditentukan dan berpotensi berkorelasi dengan eror di periode sebelumnya (*predetermined exogenous*). $\boldsymbol{s_{it}}$ adalah vektor $n \times 1$ adalah variabel eksogen yang tidak berpengaruhi ϵ_t maupun ϵ_{t-s}. Sedangkan $\boldsymbol{A}, \boldsymbol{B}$, dan $\boldsymbol{C}$ merupakan parameter untuk setiap unit i.

Pada masing-masing variabel lag endogen, variabel *predetermined exogenous,* dan eksogen memiliki *moment conditions* sebagai berikut:

$\mathrm{E}\left[\Delta^* \epsilon_{it} \boldsymbol{y_{ij}^{\mathrm{T}}}\right] = 0$

$\mathrm{E}\left[\Delta^* \epsilon_{it} \boldsymbol{x_{ij}^{\mathrm{T}}}\right] = 0$

$\mathrm{E}\left[\Delta^* \epsilon_{it} \Delta^* \boldsymbol{s_{it}^{\mathrm{T}}}\right] = 0$

Selanjutnya, kita mendefinisikan q_{it} sebagai berikut:

$$\boldsymbol{q}_{it}^{\mathrm{T}} = (\boldsymbol{y}_{it-p-1}, \boldsymbol{y}_{it-p-2}^{\mathrm{T}}, \dots, \boldsymbol{y}_{ia}^{\mathrm{T}}, \boldsymbol{x}_{it-1}^{\mathrm{T}}, \boldsymbol{x}_{it-1}, \boldsymbol{x}_{it-2}^{\mathrm{T}}, \dots, \boldsymbol{x}_{i1}^{\mathrm{T}}, \Delta^* \boldsymbol{s}_{it}^{\mathrm{T}}) \tag{1.2}$$

Lalu, kita memasukkan $\boldsymbol{q}_{it}^{\mathrm{T}}$ ke dalam bentuk matriks menjadi $\boldsymbol{Q_i}$ untuk setiap hasil t. Sehingga ketika dimasukkan pada persamaan *first difference* menjadi:

$$\Delta^{*}\boldsymbol{Y_i} = \sum_{\boldsymbol{l-1}}^{\boldsymbol{p}} \Delta^{*}\boldsymbol{Y_{il}}\boldsymbol{A_l}^{\mathrm{T}} + \Delta^{*}\boldsymbol{X_i}\boldsymbol{B}^{\mathrm{T}} + \Delta^{*}\boldsymbol{S_i}\boldsymbol{C}^{\mathrm{T}} + \Delta^{*}\boldsymbol{E}_i \tag{1.3}$$

Yang mana masing-masing variabel memiliki dimensi matriks telah disesuaikan dengan jumlah lag. Setelah didapatkan persamaan di atas, kita dapat menyusun *moment conditions* untuk setiap i

$\mathrm{E}\left[\boldsymbol{Q_i}^{\mathrm{T}}(\Delta^{*}\mathrm{E}_i)\right] = 0$

Berdasarkan persamaan sebelumnya, kita dapat memformulasikan *value function* dari instrumen $\boldsymbol{Q_i}$ menjadi:

$$\prod(\boldsymbol{\Phi}) = (\sum_{\mathrm{i=1}}^{\mathrm{N}} \boldsymbol{Z}_i^{\mathrm{T}} vec\big(\Delta^{*}\boldsymbol{Y_i} - \left[\Delta^{*}\boldsymbol{Y}_{i,-1}\, \Delta^{*}\boldsymbol{X_i}\, \Delta^{*}\boldsymbol{S_i}\right]\boldsymbol{\Phi}\big)\big)^{\mathrm{T}}\ \boldsymbol{\Lambda_Z^{-1}} \tag{1.4}$$

$$(\sum_{\mathrm{i=1}}^{\mathrm{N}} \boldsymbol{Z}_i^{\mathrm{T}} vec\big(\Delta^{*}\boldsymbol{Y_i} - \left[\Delta^{*}\boldsymbol{Y}_{i,-1}\, \Delta^{*}\boldsymbol{X_i}\, \Delta^{*}\boldsymbol{S_i}\right]\boldsymbol{\Phi}\big)\big)^{\mathrm{T}}$$

$\boldsymbol{\Phi}$ merupakan matriks dari $[\boldsymbol{A}\ \boldsymbol{B}\ \boldsymbol{C}]$. Λ_Z merupakan matriks pembobotan GMM.

Pada *system* GMM dapat dikonstruksi menggunakan asumsi (Blundell & Bond, 1998b) menggunakan definisi sebagai berikut:

$$\mathrm{E}[(\epsilon_{it} + \boldsymbol{\mu_i})(\boldsymbol{y_{it-1}} - \boldsymbol{y_{it-2}})^{\mathrm{T}}] = 0 \tag{1.5}$$

$$\mathrm{E}[(\epsilon_{it} + \boldsymbol{\mu}_i)(\boldsymbol{x_{it}} - \boldsymbol{x}_{it-1})^{\mathrm{T}}] = 0 \tag{1.6}$$

$$\mathrm{E}[\epsilon_{it} + \boldsymbol{\mu_i})\boldsymbol{s_{it}^{\mathrm{T}}}] = 0 \tag{1.7}$$

Dengan definisi di atas, maka perubahan atas y_{it} secara sistematis tidak berhubungan dengan μ_i. Maka, asumsi tersebut mampu memenuhi mode PVAR yang stasioner (Blundell & Bond, 1998b). (Blundell & Bond, 1998b) juga berargumen bahwa *system* GMM memiliki performma estimasi lebih baik daripada estimasi *first difference* GMM sebab adanya tambahan instrumen tetap menjadi prediktor tetap baik untuk variabel endogen.

Pada notasi matriks, *moment conditions* P_i untuk $p = 1$ adalah:

$$\mathbf{P}_i := \begin{pmatrix} \mathbf{0} & \Delta\mathbf{y}_{i,2} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Delta\mathbf{y}_{i,3} & & \mathbf{0} \\ \vdots & & & \ddots & \\ \mathbf{0} & \mathbf{0} & \cdots & & \Delta\mathbf{y}_{i,T-1} \\ \Delta\mathbf{x}_{i,2} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Delta\mathbf{x}_{i,3} & \mathbf{0} & & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Delta\mathbf{x}_{i,4} & & \mathbf{0} \\ \vdots & & & \ddots & \\ \mathbf{0} & \mathbf{0} & \cdots & & \Delta\mathbf{x}_{i,T} \\ \mathbf{s}_{i,2} & \mathbf{s}_{i,3} & \cdots & & \mathbf{s}_{i,T} \end{pmatrix} \tag{1.8}$$

Maka, matriks instrumen tersebut didefinisikan sebagai:

$$\boldsymbol{Q}_i^* := \begin{pmatrix} \boldsymbol{Q}_i & 0 \\ 0 & \boldsymbol{P}_i \end{pmatrix} \tag{1.9}$$

Selanjutnya, estimator dari *system* GMM dapat diturunkan menjadi:

$$vec(\widehat{\boldsymbol{\Phi}}_{IEE}) = \left(\boldsymbol{S}_{\boldsymbol{Z}^*\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Lambda}_{\boldsymbol{Z}^*}^{-1} \boldsymbol{S}_{\boldsymbol{Z}^*\boldsymbol{X}}\right)^{-1} \boldsymbol{S}_{\boldsymbol{Z}^*\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Lambda}^{-\mathbf{1}} vec(\boldsymbol{S}_{\boldsymbol{Z}^*\boldsymbol{y}}) \tag{1.10}$$

Namun, pada estimasi *two-step* dapat dikembangkan dengan *weighting matrix*:

$$\boldsymbol{\Lambda}_{\boldsymbol{Z}_{\hat{e}}^*} = \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} \boldsymbol{Z}_i^* \boldsymbol{\Gamma}_{\hat{e}} (\boldsymbol{Z}_i^*)^{\mathrm{T}} \tag{1.11}$$

Maka, estimator *two-step* GMM adalah:

$$\begin{aligned} & vec(\widehat{\boldsymbol{\Phi}}_{EFEGMM}) \\ & = \left(\boldsymbol{S}_{\boldsymbol{Z}^*\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Lambda}_{\boldsymbol{Z}^*}^{-1} \boldsymbol{S}_{\boldsymbol{Z}^*\boldsymbol{X}}\right)^{-1} \boldsymbol{S}_{\boldsymbol{Z}^*\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\Lambda}^{-\mathbf{1}} vec(\boldsymbol{S}_{\boldsymbol{Z}^*\boldsymbol{y}}) \end{aligned} \tag{1.12}$$

Setelah kita mengetahui estimator GMM, selanjutnya perlu didefinisikan matrix *asymptotic covariance* pada estimator GMM. (Cornwell & Ruud, 2001) memperlihatkan estimator *one-step* GMM secara *asymptotically* terdistribusi secara normal.

$$\begin{aligned} \hat{V}ar_{\boldsymbol{\Phi}_{\mathrm{IE}}} &:= (\boldsymbol{S}_{ZX}^{\mathrm{T}} \Lambda_Z^{-1} \boldsymbol{S}_{ZX})^{-1} \boldsymbol{S}_{ZX}^{\mathrm{T}} \boldsymbol{\Lambda}_Z^{-1} \boldsymbol{\Lambda}_{\mathrm{Z}_{\hat{e}}} \\ & \times \boldsymbol{\Lambda}_{\mathrm{Z}}^{-1} \boldsymbol{S}_{ZX}^{\mathrm{T}} \left(\boldsymbol{S}_{ZX}^{\mathrm{T}} \boldsymbol{\Lambda}_Z^{-\mathbf{1}} \boldsymbol{S}_{ZX}\right)^{-\mathbf{1}} \end{aligned} \tag{1.13}$$

Dengan $\boldsymbol{\Lambda}_Z = \boldsymbol{\Lambda}_Q \otimes \boldsymbol{I}_m$ dan $\boldsymbol{\Lambda}_{Z_{\hat{e}}}$adalah $\sum_{i=1}^{N} \boldsymbol{Z}_i^{\mathrm{T}} \boldsymbol{\Gamma}_{\hat{e}} \boldsymbol{Z}_i$. Akan tetapi pemilihan *weighting matrix* secara umum tidak optimal pada estimasi *one-step*. Maka, pada *two-step* akan menggunakan *small sample correction* untuk mendapatkan *asymptotic variance matrix* yang memiliki performa baik.

$$
\begin{aligned}
\hat{V}ar^{Wc}_{\boldsymbol{\Phi}_{FEGMM}} = {} & \left(\boldsymbol{S}^{\mathrm{T}}_{ZX}\left(\boldsymbol{\Lambda}_{Z_{\hat{e}}}(\widehat{\boldsymbol{\Phi}}_{IE})\right)^{-1} \boldsymbol{S}_{ZX}\right)^{\boldsymbol{-1}} \\
& + D_{\widehat{\boldsymbol{\Phi}}_{FEGMM},\Lambda_{Z_{\hat{e}}}(\widehat{\boldsymbol{\Phi}}_{IE})}(\boldsymbol{S}_{\boldsymbol{ZX}}(\boldsymbol{\Lambda}_{Z})^{-1}\, \boldsymbol{S}_{ZX})^{\boldsymbol{-1}}\boldsymbol{S}^{\mathrm{T}}_{\boldsymbol{ZX}}(\boldsymbol{\Lambda}_{Z})^{\boldsymbol{-1}} \\
& + \left(\boldsymbol{S}^{\mathrm{T}}_{\boldsymbol{ZX}}\left(\boldsymbol{\Lambda}_{Z_{\hat{e}}}(\widehat{\boldsymbol{\Phi}}_{\boldsymbol{IE}})\right)^{\boldsymbol{-1}} \boldsymbol{S}_{\boldsymbol{ZX}}\right)^{\boldsymbol{-1}} \boldsymbol{D}^{\mathrm{T}}_{\widehat{\boldsymbol{\Phi}}_{FEGMM},\Lambda_{Z_{\hat{e}}}(\widehat{\boldsymbol{\Phi}}_{IE})} \\
& + \boldsymbol{D}_{\widehat{\boldsymbol{\Phi}}_{FEGMM},\Lambda_{Z_{\hat{e}}}(\widehat{\boldsymbol{\Phi}}_{IE})}\hat{V}ar_{\widehat{\boldsymbol{\Phi}}_{IE}}\boldsymbol{D}^{\mathrm{T}}_{\widehat{\boldsymbol{\Phi}}_{\boldsymbol{FEGMM}},\Lambda_{Z_{\hat{e}}}}(\widehat{\boldsymbol{\Phi}}_{\boldsymbol{IE}})
\end{aligned}
\tag{1.14}
$$

Yang mana $\boldsymbol{D}^{*} = \begin{pmatrix} \boldsymbol{D} \\ \boldsymbol{I}_{(T-p)\times(T-p)} \end{pmatrix}$ digunakan untuk transformasi *first difference*.

Terakhir, model GMM-PVAR perlu melalui beberapa pengujian spesifikasi model, diantaranya adalah *Hansen overidentification test* dan *Andrews-Lu model selection procedure. Hansen overidentification test* yang disusun oleh (Hansen, 1982) dapat digunakan untuk menguji validitas dari estimator GMM bahwa variabel instrumen yang digunakan adalah eksogen. Lalu *Andrews-Lu model selection procedure* dapat digunakan untuk memilih model terbaik diantara model-model yang sudah dikembangkan (Andrews & Lu, 2001).

Walaupun model GMM-PVAR ini relatif baru, beberapa penelitian telah menggunakan model tersebut untuk mengantisipasi isu endogenitas yang mungkin muncul pada variabel mereka (lihat (Dogan dkk., 2022; Jahanger dkk., 2023)).

3. Global VAR

Model Global Vectorautoregressive (GVAR) yang dikembangkan oleh (Pesaran dkk., 2004) dengan pendekatan yang serupa dengan PVAR. Perbedaannya, model GVAR memungkinkan untuk memberikan informasi terkait pengaruh unit lain yang disebut *weakly exogenous variables,* y^{*}*,* dikonstruksi dari *cross-sectional weighted average:*

$$
y^{*}_{it} = \sum_{j=1}^{N} w_{ij} y_{jt} \tag{1.15}
$$

w_{ij} merupakan pembobotan yang ditentukan dengan $w_{ii} = 0, w_{ij} \geq 0$ untuk $i \neq j$ dan $\sum_{j=1}^{N} w_{ij} = 1$. Jika memasukkan variabel pembobot w maka $z_{it} = \left(y'^{*}_{it}, y'^{*}_{it-1}, \dots, y'^{*}_{it-p}\right)'$. Sebagai contoh, dengan asumsikan terdapat 1 lag pada model maka model GVAR adalah sebagai berikut:

$$
\boldsymbol{y_{it}} = \boldsymbol{\Phi_{i1} y_{it-1}} + \boldsymbol{\Theta_{i0} y^{*}_{it}} + \boldsymbol{\Theta_{i1} y^{*}_{it-1}} + \boldsymbol{\epsilon_{it}} \tag{1.16}
$$

$\boldsymbol{\Phi}$, sama dengan PVAR, adalah matriks koefisien dengan dimensi $(M \times M(P+1))$ dari unit i terhadap lag, dalam hal ini $P = 1$. Θ_{i} adalah matriks dimensi $(M \times M(P+1))$ yang mana Θ_{i0}

menangkap hubungan langsung antar unit, dan Θ_{i1} berisi koefisien rata-rata dari masing-masing unit berdasarkan lag yang ditentukan. Maka menggunakan vektor z_{it}, persamaan di atas dapat ditulis menjadi:

$$\boldsymbol{A_i z_{it}} = \boldsymbol{B_i z_{it-1}} + \boldsymbol{\epsilon_{it}} \tag{1.17}$$

Dengan $\boldsymbol{A_i} = (\boldsymbol{I_M}, -\boldsymbol{\Theta_{i0}})$ dan $\boldsymbol{B_i} = (\boldsymbol{\Phi_{i1}}, \boldsymbol{\Theta_{i1}})$. $\boldsymbol{I_M}$ mendeskripsikan matriks identitas dari $\boldsymbol{M}$. Lebih lanjut, konstruksi dari variabel pembobot $\boldsymbol{W_i}$ yaitu pada $z_{it} = \boldsymbol{W_i x_t}$ menghasilkan:

$$\boldsymbol{A_i W_i x_t} = \boldsymbol{B_i W_i x_{t-1}} + \boldsymbol{\epsilon_{it}} \tag{1.18}$$

Terakhir, setiap persamaan dari masing-masing unit direpresentasikan menjadi model global

$$\boldsymbol{G x_t} = \boldsymbol{H x_{t-1}} + \boldsymbol{\epsilon_t} \tag{1.19}$$

Hal yang menarik dalam GVAR adalah penentuan skema pembobot sebab terdapat beberapa pendekatan yang digunakan pada literatur selama ini. (Bussière dkk., 2009) menggunakan pembobotan proporsi total ekspor dan impor dari masing-masing negara. Pada artikel milik (Favero, 2013) yang berusaha memproyeksikan *spread* dari surat utang pemerintah, ia menggunakan pendekatan *weighted average* dari *spread* surat utang pemerintah dengan mempertimbangkan kondisi fiskal di masing-masing negara. Di sisi lain, proyeksi siklus bisnis di Indonesia dengan model GVAR oleh (Harahap dkk., 2020) menggunakan proporsi aliran perdagangan masuk dari beberapa negara relatif terhadap total yang diterima Indonesia. Pada beberapa literatur membuktikan model GVAR ini memiliki kemampuan prediksi yang baik, seperti adanya pelemahan perekonomian di beberapa negara (Greenwood-Nimmo dkk., 2012), memprediksi variabel ekonomi makro dan keuangan 1 hingga 4 kuartal ke depan (Pesaran dkk., 2009), hingga (Gunter & Zekan, 2021) yang menggunakan metode tersebut untuk proyeksi jumlah penumpang pesawat.

3.1. Bayesian Global VAR

Bayesian Global VAR (B-GVAR) merupakan pengembangan model GVAR yang memberikan keleluasaan untuk menentukan spesifikasi *shrinkage priors* (Cuaresma dkk., 2016). Menentukan *shrinkage priors* dapat membantu meningkatkan keakuratan model melalui proses *regularization* agar parameter tidak memiliki eror yang besar atau *overfitting* (Bańbura dkk., 2010). Pada implementasinya, dapat dimulai dari memetakan koefisien unit dalam $N_i = k_i K_i$ vektor $\boldsymbol{\Psi_i} = \left(\boldsymbol{\alpha_{i0}}, \boldsymbol{\alpha_{i1}}, \boldsymbol{\Phi_{i1}}, \dots, \boldsymbol{\Phi_{iP}}, \boldsymbol{A_{i0}}, \dots, A_{iQ}^*\right)^T$ dengan $K_i = k_i P + k_i^*(Q+1) + 2)$. Selanjutnya, *prior mean* dari $\boldsymbol{\Phi_{i1}}$ diasumsikan bernilai sama dengan matriks identitas (jika $\boldsymbol{y_{it}}$ merupakan $I(1)$) atau sama dengan matriks (jika $\boldsymbol{y_{it}}$ stasioner di tingkat *level*). Untuk menangkap situasi ketika variabel endogen bersifat stasioner dan persisten, *prior mean* dapat ditentukan berpusat pada proses $AR(1)$. Sedangkan pada koefisien lainnya, *prior mean* selalu ditentukan sama dengan 0.

Berbagai *shrinkage prior* yang digunakan dalam estimasi B-GVAR terdistribusi secara normal (Gaussian) dan memberikan *treatment* pada rata-rata $\underline{\boldsymbol{\Psi}_i}$ dan varians $\underline{\mathbf{V}_i}$. Maka prior $\mathbf{\Psi}$ adalah:

$$\mathbf{\Psi_i} \sim N(\underline{\mathbf{\Psi_i}}, \underline{\mathbf{V_i}}).$$

Terdapat beberapa *shrinkage prior* yang dapat digunakan. Pertama adalah Minnesota prior oleh Litterman (1986). Minnesota prior menentukan $\underline{\mathbf{\Psi_i}}$ sama dengan nol dan $\underline{\mathbf{V_i}}$ kecil. $\underline{\mathbf{V_i}}$ tersebut ditentukan berdasarkan jenis model; jumlah *lag,* adanya model *weakly exogenous*, dan bagian deterministik pada model. Minnesota prior menggunakan *hyperparameter* untuk mengontrol keketatan prior pada variabel endogen dan *weakly exogenous.* Pendekatan ini umum digunaka pada estimasi VAR

Shrinkage prior selanjutnya adalah *Stochastic Search Variable Selection* (SSVS) *prior*. Pendekatan ini digunakan untuk memilih variabel dalam Bayesian (George dkk., 2008). Pemilihan variabel ini menggunakan mekanisme *spike* dan *slab,* yaitu mengatur besaran koefisien *prior* untuk menentukan besar/kecilnya pengaruh variabel dalam model. Lalu, penentuan *spike* dan *slab* tersebut menggunakan distribusi *bernoulli* untuk menentukan signifikansi *spike* dan *slab* dalam estimasi.

Normal-Gamma prior merupakan jenis *shrinkage prior* yang secara umum digunakan pada berbagai estimasi serta mampu bersifat hierarkis. Hierarikis penting dalam estimasi GVAR untuk membedakan posisi *local* dan *global.* Pada implementasi, pendekatan ini memadukan dua jenis distribusi *Normal* dan *Gamma* dalam penentuan *prior,* yang mana $\underline{\boldsymbol{\Psi}_i}$ ditentukan berdasarkan distribusi normal, sedangkan $\underline{\mathbf{V_i}}$ menggunakan distribusi gamma. Hal ini memberikan keunggulan dalam hal pemberian penalti yang adaptif serta penyesuaian *hyperparameter* distribusi *Gamma* untuk menerapkan *regularization.*

Terakhir adalah *Horseshoe* (HS) *prior* oleh Carvalho dkk (2010). *Shrinkage prior* ini memberikan fleksibilitas dalam menentukan *hyperparameter*, namun tetap adaptif pada berbagai situasi. HS *prior* menggunakan parameter *shrinkage* individu untuk setiap koefisien yang mana parameter tersebut menggunakan distribusi *Half-Cauchy* pada parameter *shrinkage local.* Distribusi ini memungkinkan untuk koefisien memiliki tingkat signifikansi tinggi. Lalu, *shrinkage* di level *global* menggunakan sebuah parameter $\boldsymbol{\Psi}$ yang terdistribusi secara normal dengan nilai rata-rata sama dengan 0.

4. Factor Augmented VAR

Factor Augmented Vector Autoregressions (FAVAR) merupakan model yang dibangun oleh (Bernanke dkk., 2004) bertujuan untuk mengurangi dimensi pada data dengan mengekstraksi secara efisien menjadi dimensi lebih kecil pada *unobserved variable*. Hal ini berbeda jika dibandingkan dengan PVAR maupun GVAR yang secara langsung menentukan adanya interdependensi antar negara. Model FAVAR dapat digunakan untuk berbagai konteks, baik data yang bersifat runtun waktu maupun panel.

Pendekatan FAVAR berusaha menghubungan runtun waktu dalam $\boldsymbol{x_t}$ dengan dimensi K menjadi *variable of interest* $\boldsymbol{m_t}$ dalam dimensi M dan *latent factors* $\boldsymbol{f_t}$ dalam dimensi Q sebagai berikut:

$$\boldsymbol{x_t = \Psi_f f_t + \Psi_m m_t + \epsilon_t, \qquad \epsilon \sim N(0, \textstyle\sum_\epsilon)} \tag{1.20}$$

Pada persamaan di atas, $\boldsymbol{\Psi_f}$ dan $\boldsymbol{\Psi_m}$ adalah matriks *factor loadings* dari $K \times Q$ dan $K \times M$. Di sini menjelaskan bahwa $\boldsymbol{x_t}$ dijelaskan oleh dimensi yang lebih kecil $(Q \ll K)$ dari variabel *observed* dan *unobserved,* sehingga sensitivitas dari $\boldsymbol{x_t}$ digambarkan oleh pergerakan $\boldsymbol{f_t}$ dan $\boldsymbol{m_t}$ secara beriringan. Maka, *state equation* dari *joint evolution* pada susunan variabel $(M + Q) - vektor\ z = (\boldsymbol{f'_t, m'_t})'$ dengan asumsi P lag adalah:

$$\boldsymbol{z_t = A_1 z_{t-1} + \cdots + A_p z_{t-P} + \eta_t, \qquad \eta \sim N(0, \textstyle\sum_\eta)} \tag{1.21}$$

$\boldsymbol{A_P}$ adalah matriks koefisien dengan dimensi $(M + Q) \times (M + Q)$ dan eror diasumsikan terdistribusi normal. Pada konteks data panel, dapat mengkategorikan variabel unit i sebagai *observed factors* y_{it}, sedangkan variabel dari unit j sebagai *unobserved factors,* atau $\boldsymbol{x_t} = (\boldsymbol{y'_{1t}}, \dots, \boldsymbol{y'_{Nt}})'$ dan $\boldsymbol{m_t} = \boldsymbol{y_{1t}}$. Model ini digunakan pada beberapa literatur seperti (Dave dkk., 2013) yang melihat adanya transmisi *bank lending* pada kebijakan moneter di tingkat individu bank, (Smith & Zoega, 2005) berusaha mengetahui pengaruh dinamika global terhadap tingkat pengangguran, dan juga memiliki kemampuan proyeksi yang baik dalam proyeksi inflasi di India (Dua & Goel, 2023)

B. Metode Analisis Spasial Perubahan Harga Komoditas

Kami akan menggunakan beberapa metode sebagai berikut untuk melakukan tingkat perubahan harga beberapa komoditas di Kabupaten Nganjuk, serta analisis keterkaitan antar pasar. Penelitian ini menggunakan pendekatan analisis spasial untuk menangkap pengaruh dari wilayah ataupun spatial unit yang berdekatan dengan setiap tempat. Dalam konteks penelitian ini, maka pendekatan spasial digunakan untuk menangkap seberapa besar pengaruh tingkat harga dari pasar-pasar yang berdekatan tersebut dengan masing-masing pasar di Jawa Timur, khususnya di wilayah Kabupaten Nganjuk. Metodologi yang digunakan di sini terbagi dalam dua pendekatan, yaitu: pengujian autokorelasi spasial dan analisis ekonometrika spasial.

1. **Uji Spasial Autokorelasi: Global dan Lokal Moran**

Metode pertama akan terbagi ke dalam 2 sub metodologi yaitu analisis global spasial dan lokal spasial. Untuk analisis global spasial menggunakan metode Global Moran's I statistik. Metode ini digunakan untuk mengukur dan mengidentifikasi apakah ada pola spasial dari data sebuah variabel dalam suatu wilayah. Alat analisis ini digunakan untuk mendeteksi apakah distribusi geografis dari suatu variabel memiliki kecenderungan untuk berkumpul (*clustering*) atau tersebar (*scatteriing*) secara spasial. Metode pengujian tersebut adalah

$$\mathcal{I} = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1.22}$$

atau dalam bentuk matriks, secara lebih sederhana persamaan tersebut dapat ditulis menjadi,

$$I = \frac{n}{\sum_i \sum_j w_{i,j}} \frac{\mathbf{z}'\mathbf{W}\mathbf{z}}{\mathbf{z}'\mathbf{z}} \tag{1.23}$$

dengan $w_{i,j}$ merupakan nilai status *neighborhood* antara wilayah i dan wilayah j. Jika kedua wilayah tersebut bertetangga, maka $w_{i,j} \neq 0$, sebaliknya maka $w_{i,j} = 0$. Untuk keselurahan status bertetangga tersebut, maka dapat ditabulasikan untuk seluruh sub-wilayah ke dalam matriks pembobot spasial ($\mathbf{W}$), sehingga $w_{i,j} \in \mathbf{W}$. Sementara itu, $\mathbf{z}$ merupakan vektor yang mencakup interaksi antara $(y_i - \bar{y})$ untuk seluruh sub-wilayah dalam analisis.

Berdasarkan nilai perhitungan tersebut, maka diketahui bahwa niilai ada dalam rentang $-1 \leq \boldsymbol{I} \leq 1$. Jika $\boldsymbol{I} = -1$, maka pola yang didapat adalah dispersi secara sempurna. Apabila $\boldsymbol{I} = 1$, maka terdapat klusterisasi kewilayahan. Sedangkan jika nilai $\boldsymbol{I} = 0$, maka tidak terdapat hubungan spasial antar wilayah atau tersebar secara acak.

Sedangkan untuk sub metode yang kedua adalah dengan mengeksplore pengujian keterikatan spasial ke masing-masing komponen spatial unit. Untuk itu, maka dibutuhkan pengujian dalam skala lokal. Sub metode ini juga dapat digunakan untuk menghindari dampak outlier pada global Moran's I statistik. Analisis dilakukan menggunakan persamaan berikut:

$$\mathcal{I}_i = \frac{(y_i - \bar{y})\sum_{j=1}^{n} w_{ij}(y_j - \bar{y})}{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}} \tag{1.24}$$

Dari persamaan tersebut akan diperoleh nilai Moran's I untuk setiap wiilayah i. Dari hasil tersebut kemudian kita juga dapat menyajikan nilai-nilai tersebut secara visual menggunakan peta wilayah distribusi local Moran's I.

2. Uji Spasial Autokorelasi: Global dan Lokal Geary

Global Geary's C juga merupakan indikator pengukuran terjadinya keterikatan spasial dan melhat apakah terdapat pola klusterisasi atau dispersi dalam data. Pendekatan ini merupakan salah satu pendekatan untuk menjawab limitasi terkait dengan *handling* outlier data dari global Moran's I. Untuk perhitungan global Geary's C adalah sebagai berikut:

$$C = \frac{(n-1)\sum_i \sum_j w_{ij}(y_i - y_j)^2}{2(\sum_i \sum_j w_{ij})\sum_i (y_i - \bar{y})^2} \tag{1.25}$$

dengan nilai tersebut berada dalam rentang $0 \leq C \leq 2$. Dalam perhitungan tersebut global Geary's C memberikan adjustment terkait dengan covariations yang ada dalam data. Berdasarkan perhitungan di atas, secara umum, global Geary menghitung disparitas dengan menggunakan squared distances, dan ini menyebabkan perhitungannya tidak terlalu terasosiasi dengan hubungan linear antara variabel dan spatial lag nya.

3. Model Ekonometrika Spasial

Setelah melakukan pengujian untuk mendeteksi eksistensi hubungan spasial, baik secara global dan lokal, maka kemudian analisis dilanjutkan dengan menggunakan pendekatan regresi yang mengikutkan informasi hubungan spasial tersebut. Oleh karena itu, dengan berbasis pada LeSage & Pace (2010) dan Elhorst (2014), maka model yang digunakan dalam studi ini adalah sebagai berikut:

$$\mathbf{Y}_t = \alpha \mathbf{i} + \boldsymbol{\mu} + \boldsymbol{\tau} + \delta(\mathbf{WY})_t + \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_t \sim \mathbf{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}) \tag{1.26}$$

Y merupakan variabel independent, yaitu tingkat harga, dan **WY** merupakan spatial lag dari tingkat harga dengan **W** merupakan spatial weights matrix yang berisi informasi interlinkages antar pasar. Karakteristik khusus untuk **W** matriks dalam peneletian ini adalah hubungan kedekatan antar pasar diukur dari nilai inverse jarak antar dua pasar. Jika jarak pasar i dan j dinotasikan dengan d_{ij} maka $w_{ij} = 1/d_{ij}$ dengan $i \neq j$.

α merupakan nilai kontanta. δ adalah parameter spatial autokorelasi dengan nilai $-1 < \delta < 1$. Jika nilai δ bernilai positif, maka terdapat klusterisasi dengan ditandai oleh gradasi nilai yang berada pada kisaran yang sama untuk pasar-pasar yang berdekatan. Jika nilai tersebut negative, maka kecenderungan tingkat harga yang berbeda untuk pasar-pasar yang berdekatan. $\boldsymbol{\mu}$ dan $\boldsymbol{\tau}$ merupakan komponen fixed effect.

Model dalam persamaan (1.26) kemudian diterapkan untuk beberapa komoditas pasar yang tersedia. Hal ini ditujukan untuk melihat karakteristik hubungan antar pasar di masing-masing komoditas dan melihat komoditas manakah yang sensitif terhadap perubahan harga di pasar lain. Dalam proses tersebut juga dilakukan pengujian robustness dengan menggunakan pendekatan lain dalam mendefinisikan hubungan keterikatan antar pasa, di antara lain adalah dengan menggunakan hubungan biner antar pasar dengan $w_{ij} = 1$, jika dan hanya jika pasar j merupakan pasar terdekat dari pasar i. Selain itu, maka nilai $w_{ij} = 0$.

4. Time-series in Space Model

Pendekatan selanjutnya ini digunakan untuk memasukkan unsur spasial kewilayahan dalam proses forecasting tingkat harga per komoditas di pasar di Kabupaten Nganjuk. Model yang digunakan di sini merupakan dekomposisi model dari persamaan (1.25) dengan melepas komponen per unit spatial, sehingga persamaan terbaru adalah sebagai berikut:

$$\begin{aligned} \mathbf{y}_{p,t} = \alpha \boldsymbol{i} + \sum_{k=1}^{T} \boldsymbol{\gamma}_k \mathbf{y}_{p,t-k} + \delta(\mathbf{WY})_{p,t} \\ + \sum_{k=1}^{T} \boldsymbol{\lambda}_k (\mathbf{WY})_{p,t-k} \\ + \boldsymbol{\varepsilon}_{p,t}, \boldsymbol{\varepsilon}_t \sim \mathbf{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}) \end{aligned} \tag{1.27}$$

Dalam persamaan (1.26), besaran tingkat harga $\mathbf{y}_{p,t}$ diestimasi untuk masing-masing komoditas khusus untuk rerata nilai dari pasar-pasar yang ada di Kabupaten Nganjuk. Sehingga, fluktuasi harga tersebut tidak hanya akan melihat data historis tingkat harganya, tetapi juga komponen historis tingkat harga di pasar sekitarnya, $(\mathbf{WY})_{p,t}$. Komponen nilai ini juga tidak hanya dipengaruhi oleh nilai spatial pasar lainnya di Jawa Timur, tetapi juga nilai historisnya. Untuk memudahkan proses estimasi, nilai weighted spasial tetap sama dan bersifat time invariant.

Persamaan (1.27) menjadi basis model ataupun konfigurasi dasar dari model lanjutan yang dijadikan model utama dalam analisis. Dalam studi ini beberapa model digunakan sebagai bahan konsiderasi dan pertimbangan untuk melihat model terbaik untuk setiap komoditas. Beberapa model tersebut adalah 1) standard vector autoregressive (VAR), 2) univariate autoregressive with moving average structure (ARMA) or (ARIMA), and 3) autoregressive with distributed lag (ARDL).

Vector aturogressive model

Proses estimasi pertama didasarkan pada struktur dan konfigurasi berbasis VAR dengan memadukan dengan konfigurasi spasial pada persamaan (1.27). Perhitungan didasarkan pada model sebagai berikut:

$$\boldsymbol{z}_t = c + \sum_{i=1}^{p} A_i z_{t-i} + e_t \tag{1.28}$$

Komponen vector $\boldsymbol{z}$ menggabungkan dua komponen variabel, yaitu variabel runtun waktu untuk rerata harga komoditas di Kabupaten Nganjuk dan juga variabel yang memperhitungkan nilai rerata spasial komoditas dari pasar-pasar lain di Provinsi Jawa Timur. Besaran nilai historis p, ditentukan dari proses pemilihan lag atau nilai historis pada periode sebelumnya. Pemilihan ini menentukan sejauh mana data historis yang memberikan dampak pada nilai harga pada saat ini. Proses ini dilakukan dengan membandingkan nilai Akaike information criterion (AIC), dengan nilai p terpilih yang menghasilkan AIC terkecil.

Proses estimasi tersebut diperoleh dari subset training data, yang kemudian akan dilakukan pengukuran dan komparasi untuk subset test data. Subset test data ini mencakup 12 bulan terakhir dari data yang dimiliki. Dari 12 bulan tersebut kemudian diukur seberapa baik proses forecasting yang diperoleh. Untuk memperoleh persamaan tersebut, terlebih dahulu dilakukan estimasi training data. Estimasi tersebut didasarkan kepada persamaan (1.27) dan diperoleh nilai estimasi setiap parameter, kemudian dilakukan *forecasting* untuk 12-step ahead

$$\hat{\boldsymbol{z}}_{t+12} = \hat{c} + \sum_{i=1}^{p} \widehat{A}_i z_{t-i} + e_t \tag{1.28}$$

Autoregressive with Moving Average Model

Model kedua yang digunakan disini adalah model univariate standar dengan konfigurasi ARMA ataupun ARIMA untuk harga setiap komoditas. Model ini tidak hanya memasukkan unsur data historis variabel terkait, tetapi juga komponen historis nilai residual atau faktor-faktor lain di luar model ini.

$$\boldsymbol{z}_t = c + \sum_{i=1}^{p} \beta_i z_{t-i} + e_t + \sum_{i=1}^{q} \gamma_i e_{t-i} \tag{1.29}$$

Model ini digunakan sebagai bentuk pendekatan yang lebih sederhana dalam proses analisis data runtun waktu. Dalam pendekatan ini, hanya melibatkan satu variabel, yaitu tingkat harga komoditas antara kurun waktu periode. Tentu hal ini juga ditujukan untuk membandingkan dengan metode pertama yang juga melibatkan unsur variabel spasial. Keduanya akan dibandingkan apakah dengan menambahkan unsur variabel tingkat harga komoditas pasar lain dapat menghasilkan estimasi yang lebih baik. Walaupun begitu, model ARMA atau ARIMA ini memiliki kelemahan jika konstruksi ini ingin melibatkan komponen penjelas lain. Oleh karena itu, dalam studi ini juga melibatkan struktur model sederhana autoregressive dengan melibatkan unsur spasial dalam model ARDL.

Autoregressive Distributed Lag (ARDL)

Pemilihan model ketiga, ARDL, merupakan upaya untuk memasukkan peran dan nilai dari variabel tingkat harga dari pasar-pasar lain di luar Kabupaten Nganjuk. Penggunaan ARDL ini memungkan fleksibilitas model untuk mengatasi ataupun mengakomodasi analisis multivariat. Salah satu kelebihan lain dalam model ini adalah dengan memperbolehkan variabel-variabel dalam analisis ini memiliki tingkat keseimbangan runtun waktu, ataupun stasioneritas yang berbeda-beda. Di sisi lain, perhitungan model estimasi yang diperoleh juga mampu mengakomodir persamaan dalam jangka panjang dan pendek. Untuk studi ini, akan difokuskan pada hasil estimasi dan forecasting berbasis analisis jangka pendek. Untuk memperoleh seluruh analisis tersebut, maka perlu dilakukan estimasi menggunakan konfigurasi sebagai berikut:

$$\boldsymbol{z}_t = c + \sum_{i=1}^{p} \beta_i z_{t-i} + \sum_{i=0}^{q} \delta_i x_{t-i} + e_t \qquad (1.30)$$

Komponen $\boldsymbol{z}_t$ merupakan variable nilai rerata per komoditas di Kabupaten Nganjuk dan $\boldsymbol{x}_t$ adalah penjumlahan nilai komoditas dari pasar-pasar lain di Jawa Timur dengan nilai sebesar $\sum_{j=1}^{n} w_{i,j} z_{j,t}$. Sedangkan untuk komponen nilai historis spasial $\boldsymbol{x}_{t-1}, \dots, \boldsymbol{x}_{t-T}$ dihitung dari komponen $\sum_{j=1}^{n} w_{i,j} z_{j,t-1}, \dots, \sum_{j=1}^{n} w_{i,j} z_{j,t-T}$.

Berdasarkan persamaan (1.30) untuk spesifikasi model ARDL(p,q) di atas, kemudian dilakukan pengujian untuk pemilihan lag historis, p dan q terbaik. Setelah proses pemilihan tersebut dilakukan maka besaran ini akan didistribusikan baik untuk *variable of interest* dan juga variabel penjelas lainnya. Dari proses tersebut, kemudian dilanjutkan dengan proses penyusunan konfigurasi jangka pendek sebagai berikut,

$$\Delta \boldsymbol{z}_t = c + \sum_{i=1}^{p} \beta_i \Delta z_{t-i} + \sum_{i=0}^{q} \delta_i \Delta x_{t-i} + ECT_t + e_t \qquad (1.31)$$