

## 3

### Metodologia

#### 3.1

#### Introdução aos Modelos de Equações Múltiplas

Até o momento, ao tratarmos de processos multivariados (isto é, que envolvem múltiplas variáveis), nos limitamos a analisar modelos compostos por uma única variável endógena e, portanto, uma única equação tais como os modelos *ADL-Autoregressive Distribution Lag* e os modelos de componentes não observados com variáveis exógenas.

No contexto desses modelos, a atividade de previsão requer a construção de “cenários” para as variáveis exógenas durante o período de previsão. Por exemplo, a fim de realizar previsões a partir do modelo ADL, precisamos de valores futuros hipotéticos da variável exógena.

Os cenários para as variáveis exógenas podem ser postulados de modo totalmente arbitrário. Isso pode ser útil para realizar “testes de stress” e responder a perguntas do tipo: “O que acontecerá com a demanda industrial se a tarifa aumentar 10% no ano?”. Entretanto, a fim de obter a melhor previsão “incondicional” da variável dependente, necessitamos de boas previsões também para as variáveis exógenas, caso contrário, as previsões finais estarão “contaminadas” pelo uso de cenários equivocados no nosso modelo de previsão.

**Assim, se o foco é a maximização da capacidade de previsão de certa variável**, e não a realização de projeções “condicionais”, é conveniente modelar também as variáveis explicativas da equação de interesse, com o objetivo de gerar cenários razoáveis para estas últimas. Isto pode ser feito aplicando-se os mesmos métodos usados para modelar a variável “final” de interesse, por exemplo, através de modelos uni - equacionais univariados do tipo ARIMA ou de componentes não observados.

Ao “empilhar” o modelo original para a variável de interesse e os modelos para as variáveis explicativas, obtemos então um sistema de previsão de múltiplas variáveis. Se as equações para as variáveis explicativas forem modelos univariados

(ou modelos multivariados que envolvam apenas outras variáveis já modeladas no sistema), será possível usar o sistema para realizar previsões que independam de cenários arbitrários para variáveis “exógenas”.

Por exemplo, poder-se-ia ter um modelo composto por três equações: a primeira determinaria a previsão do consumo residencial de energia elétrica em determinado Estado da Federação (variável endógena) em função da temperatura média e do rendimento médio real da população do Estado (variáveis explicativas), e as outras duas equações explicariam a evolução da temperatura e do rendimento real no estado a partir de modelos univariados. Assim, o sistema (ou “modelo completo”) permitiria prever os valores de cada uma das três variáveis endógenas, período após período: os modelos univariados gerariam previsões da temperatura e do rendimento real, e estas previsões entrariam na primeira equação para determinar o consumo de energia esperado.

Entretanto, o mero “empilhamento” de modelos uniequacionais estimados independentemente – e que poderiam perfeitamente funcionar de modo também independente, cada um gerando previsões de sua própria variável “do lado esquerdo da equação” – não é, em geral, capaz de captar adequadamente as possíveis inter-relações entre as variáveis modeladas.

Para tanto, é necessário um modelo verdadeiramente simultâneo, que permita relações de interdependência e causalidade entre as variáveis analisadas. Por exemplo, pode-se considerar um modelo de oferta e demanda em que, na primeira equação, o nível de demanda de certo produto seja explicado pelo preço do produto; e, na segunda equação, o preço do produto seja explicado pelo nível de demanda. Nesse tipo de modelo, a interdependência entre as variáveis endógenas é um fator fundamental para explicar a dinâmica das variáveis e, conseqüentemente, para gerar melhores previsões.

O processo de especificação e estimação de modelos de equações simultâneas não é, porém, tarefa fácil. De fato, a classificação a priori das variáveis como “endógenas” ou “exógenas” em cada equação é muitas vezes arbitrária e pode estar sujeita a críticas. Por essa razão, é comum o uso de uma abordagem alternativa, **em que todas as variáveis são tratadas como endógenas, o que minimiza a necessidade de impor restrições possivelmente falsas ao modelo.** Nessa linha, os modelos de Auto-Regressão Vetorial (VAR) têm se revelado bastante populares na linha econométrica.

Neste capítulo é mostrado a metodologia das abordagens clássica e bayesiana do VAR para o modelo de previsão de demanda de energia elétrica da classe industrial.

## 3.2

### Revisão de Literatura

Modiano (1984) estimou a elasticidade renda e elasticidade preço para demanda por energia elétrica para o Brasil utilizando dados anuais do período de 1963 a 1981, isto para as classes de clientes: residencial, comercial e industrial. Este autor utilizou para suas estimativas a regressão de uma equação Cobb Douglas, não sendo relatado a utilização de recursos econométricos como teste ADF nem cointegração. O resultado que ele obteve foram as elasticidades renda de longo prazo de 1,130, 1,068 e 1,360 respectivamente. Para as elasticidades preço de longo prazo Modiano (1984) encontrou -0,403, -0,183 e -0,222, isto para as classes residencial, comercial e industrial respectivamente.

Andrade e Lobão (1997) calcularam as elasticidades renda e elasticidade preço para a classe de cliente residencial para demanda por energia elétrica para o Brasil, os valores encontrados foram 0,2132 e -0,05084, respectivamente. Eles utilizaram dados anuais de 1970 a 1995. Além disto, Andrade e Lobão (1997) fizeram previsões para o consumo de energia elétrica para a classe residencial para o período de 1997 a 2005, eles utilizaram para estas estimativas um modelo de correção de erros vetoriais – VEC.

Fiorencio et al. (1998) usaram o BVAR para analisar os impactos da política monetária e cambial sobre o nível de desemprego e inflação pós Plano Real, entre 1994 a 1997. O modelo de Fiorencio et al. (1998) usou como variáveis exógenas o nível de preço (IPCA), a taxa de desemprego, a taxa de câmbio, a taxa de juros, o financiamento do capital de giro e a expansão entre financiamentos de capital e títulos privados (CDBs). Empregando identificação não-recursiva mostraram que a taxa de câmbio tem impactos significativos sobre o aumento do nível de preços e do desemprego, enquanto que choques de política monetária reduzem o nível de preços e aumenta o nível de desemprego.

Giambiagi et al (2001) avaliaram as condições de oferta de energia elétrica do Brasil do período de 2001 a 2009. Os autores verificaram se a restrição de oferta de energia elétrica poderia vir a representar um problema para as perspectivas de crescimento econômico. A metodologia adotada pelos autores foi de se fazer uma análise de alguns indicadores econômicos, como: inflação, investimento estrangeiro, taxa de câmbio, balança comercial, déficit em conta corrente, entre outros, e a partir disto foi feita uma previsão para a taxa de crescimento do PIB brasileiro para o período de análise, chegando à conclusão que o Brasil apresentará um crescimento sustentável deste indicador. Depois os autores verificaram se a capacidade energética seria um limitador para o crescimento econômico, neste tópico foi feita uma referência ao Plano Decenal de Expansão de Energia Elétrica de 2000 a 2009 elaborado pela Eletrobrás, sendo concluído que a energia elétrica seria um entrave para o crescimento econômico, principalmente para os anos de 2001 e 2002, exatamente quando ocorreu o racionamento de energia elétrica no Brasil. Há de se registrar que Giambiagi et al (2001) não apresentaram em seu trabalho nenhum cálculo econométrico para avaliar se a energia elétrica seria um gargalo para a economia, sendo feito somente uma referência ao trabalho da Eletrobrás, eles também não apresentaram nenhuma estimativa para a elasticidade preço elasticidade renda para demanda por energia elétrica.

Schmidt e Lima (2004) tiveram como objetivo estimar a elasticidade preço e a elasticidade renda de longo prazo da demanda por energia elétrica nas três classes de consumo: residencial, comercial e industrial. Eles chegaram aos valores de 0,539, 0,636 e 1,718 para a elasticidade renda de longo prazo respectivamente, e para a elasticidade preço de longo prazo os valores foram -0,085, -0,174 e -0,129. Além disso, são realizadas previsões para o consumo de energia elétrica, vale ressaltar que esses autores não consideraram em suas estimativas a classes de “Outros” clientes, não sendo possível com isto estimar o total de demanda de energia elétrica. Schmidt e Lima (2004) fazem uma análise de cointegração usando a metodologia de Johansen, isto para o cálculo das elasticidades e para a elaboração das previsões de consumo de energia elétrica.

Fernandes-Toro (2005), com amostras entre novembro de 1994 a fevereiro de 2001, estimaram o mecanismo de transmissão monetária para o Brasil após o Plano Real. Identificam vetores de cointegração como relações de equilíbrio de longo prazo. De acordo com Fernandes-Toro (2005) um choque positivo de política

monetária, identificado no modelo de Vetores Auto-Regressivos com Correção de Erros (VECM) como inovações na SELIC, tem temporariamente, um pequeno efeito negativo sobre a inflação. Esse efeito que desaparece em seis meses implica num efeito permanente de política monetária sobre o nível de preço.

Mattos (2005) estimou a elasticidade renda e elasticidade preço para a classe de cliente industrial da energia elétrica para o Brasil utilizando dados anuais de 1974 a 2002, os valores que ele estimou foram 1,588 para a elasticidade renda de longo prazo e de -0,489 para a elasticidade preço de longo prazo. Para verificar se as variáveis do modelo (consumo de energia elétrica, PIB real e tarifa média real) cointegram, ele utilizou a metodologia de Engle - Granger, verificando se o resíduo da equação de cointegração tem raiz unitária, e a conclusão que ele chegou foi que as três variáveis são cointegradas.

Kapetanios-Labhard-Price (2008) geraram previsões para inflação e produto. Comparam modelos lineares e não-lineares. Dentre esses modelos de previsão encontra-se o modelo AR com constante; passeio aleatório (random walk), VAR, BVAR Modelo Markov-Switching, modelo auto-regressivo smooth-transition (STAR) e modelo de média incondicional. Similar ao trabalho que compara performance de modelos de Carvalho-Minella (2009), Kapetanios- Labhard-Price (2008), mostram que, muitas vezes, o VAR/BVAR podem ser superiores a outros métodos de previsão.

Carvalho-Minella (2009), por exemplo, compararam as expectativas de mercado da taxa de inflação, taxas de juros e câmbio com as previsões dos modelos Auto-Regressivos com Média Móvel (ARMA), VAR e BVAR. Entre outras conclusões, Carvalho-Minella (2009), argumentam que algumas expectativas de mercado tem performance melhor ou superior a modelos de previsão ARMA, VAR e BVAR. O oposto para alguns modelos de previsão ocorre, ou seja, existem modelos de previsão que tem performance melhor do que algumas expectativas de mercado.

### 3.3

#### Teste de Causalidade

Consiste na análise de interrelações entre variáveis. O objetivo é responder à pergunta: determinada variável  $X_t$  “causa” outra variável  $Y_t$ ?

Uma variável  $X_t$  “causa” outra variável  $Y_t$  se as defasagens de  $X_t$  ajudam a prever o comportamento de  $Y_t$  isto é se a inclusão das defasagens de  $X_t$  na equação de  $Y_t$  aumentam a capacidade de prever  $Y_t$ .

**Uma das conseqüências desse teste é a utilização (ou não) de modelos de múltiplas equações.**

Logo, pode-se testar se  $X_t$  causa  $Y_t$  através do teste da hipótese de que os coeficientes de todas as defasagens de  $X_t$  na equação de  $Y_t$  são conjuntamente iguais a zero, que é o teste-F usual de restrições lineares. Esse teste é válido no caso de variáveis estacionárias e aproximadamente válido no caso de variáveis não estacionárias, mas cointegradas.

Abaixo a estatística de teste.

$$F = \frac{\frac{SQR_i - SQR_r}{k}}{\frac{SQR_i}{n-2k}} \quad (3.1)$$

Onde:

$SQR_r$  é a soma dos quadrados dos resíduos da equação restrita.

$SQR_i$  é a soma dos quadrados dos resíduos da equação irrestrita.

$n - 2k$  são os graus de liberdade da distribuição.

A bi-causalidade ocorre quando a hipótese nula é rejeitada.

Considere as mesmas séries de tempo  $X_t$  e  $Y_t$ . O teste de causalidade de Granger assume que a informação relevante para a predição das respectivas variáveis  $X_t$  e  $Y_t$  está contida apenas nas séries de tempo sobre essas duas variáveis. Dessa forma, uma série de tempo estacionária  $X_t$  causa, no sentido de Granger, uma outra série estacionária  $Y_t$  se melhores predições estatisticamente significantes de  $Y_t$  podem ser obtidas ao incluirmos valores defasados de  $X_t$  aos valores defasados de  $Y_t$ . Em termos mais formais, o teste envolve estimar as seguintes regressões:

$$X_t = \sum a_i Y_{t-1} + \sum b_i X_{t-1} + u_t \quad (3.2)$$

$$Y_t = \sum c_i Y_{t-1} + \sum d_i X_{t-1} + u_t \quad (3.3)$$

Onde  $u_t$  são os resíduos que assumimos serem não-correlacionados.

A equação (3.2) postula que valores correntes de  $X_t$  estão relacionados a valores passados do próprio  $X_t$  assim como a valores defasados de  $Y_t$ ; a equação (3.3), por outro lado, postula um comportamento similar para a variável  $Y_t$ . Nada impede que as variáveis  $X_t$  e  $Y_t$  sejam representadas na forma de taxas de crescimento, o que aliás tem sido quase que a regra geral na literatura, uma vez que é difícil achar variáveis que sejam estacionárias em seus níveis.

### 3.4

#### Teste de Raiz Unitária – Teste ADF

Um processo estocástico é estacionário quando preenche três requisitos. Primeiro, tem média constante ao longo do tempo; segundo, tem variância constante ao longo do tempo; e terceiro, sua covariância deve indicar que a autocorrelação entre dois valores de uma variável  $y$  qualquer tomados a partir de dois períodos de tempo distintos deve depender apenas do intervalo de tempo entre esses dois valores e não do tempo em si.

Com o objetivo de determinar se as variáveis analisadas são estacionárias ou processos integrados foi realizado o teste de Dickey - Fuller Aumentado.

O teste de raiz unitária ADF utiliza a auto-regressão abaixo.

$$\Delta Y_t = \alpha + \beta_t + \gamma Y_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta Y_{t-1} + \varepsilon_t \quad (3.4)$$

Onde:

$\Delta Y_t$  = operador de diferenças ( $Y_t - Y_{t-1}$ )

$\alpha$  = constante;

$\beta_t$  = componente de tendência do modelo;

$\gamma Y_{t-1}$  é o  $\rho - 1$  que testará a estacionariedade ou não da série ao se regredir a variável  $Y_{t-1}$ ;

$\sum_{j=1}^{\rho=1} \rho_{j+1} \Delta Y_{t-j}$  são as defasagens incluídas no modelo ADF para garantir a não autocorrelação nos resíduos.

A partir da equação (3.4) testam-se as seguintes hipóteses:  $H_0: \delta_i = 0$  e  $H_1: \delta_i < 0$ . A estatística de teste é mostrada abaixo.

$$\tau = \frac{\widehat{\delta}_i}{EP(\widehat{\delta}_i)} \quad (3.5)$$

Onde:  $EP(\widehat{\delta}_i)$  é o erro padrão de  $\widehat{\delta}_i$ .

Por meio da estatística  $\tau$  (tau) pode-se definir se a série original da variável dependente  $Y_t$  é ou não estacionária, utilizando a estatística de teste ADF. Se  $|\tau|_{calculado} > |\tau|_{critico}$ , rejeita-se  $H_0: \delta_i = 0$ , logo, a série é estacionária; se  $|\tau|_{calculado} < |\tau|_{critico}$ , não se rejeita  $H_0: \delta_i = 0$ , e a série é não-estacionária.

Nesse trabalho, foi utilizado o critério de informação de Akaike para selecionar o número de defasagens e incluído no processo auto-regressivo de cada variável tendência e intercepto.

### 3.5

#### Teste de Cointegração

Ao trabalhar com dados de séries temporais existe a possibilidade de, ao estimar uma regressão com séries que apresentam uma mesma tendência, obter resultados que aparentemente sejam satisfatórios, mas que na realidade não tenham nenhum significado econômico. Ou seja, é possível encontrar relações econométricas entre duas ou mais séries, com  $R^2$  elevado e estatística t significante, sem que necessariamente exista alguma relação de causalidade entre elas. Neste tipo de situação, os estimadores obtidos por MQO são ineficientes e o desvio-padrão dos resíduos é inconsistente.

Fatos como esse geralmente ocorrem quando as séries utilizadas não são estacionárias. Uma primeira medida para resolver a questão poderia ser diferenciar as séries e verificar se elas se tornaram estacionárias. Ao adotar esse procedimento, todavia, perde-se uma valiosa relação de longo prazo entre as variáveis, uma vez que não será possível trabalhar com as séries em nível.

Quando as séries analisadas apresentam uma relação de longo prazo entre si, suas tendências se anulam e torna-se possível estimar um modelo de vetores autoregressivos, com a inclusão de um Vetor de Correção de Erros (VEC).

O método utilizado para testar a existência de relação de longo prazo entre as variáveis seguiu o método de estimação de Johansen (1988), que permite a estimação de um VEC simultaneamente aos vetores de cointegração. Além disso, utilizou-se a estatística do traço para determinar se as séries analisadas são cointegradas.

O teste traço, proposto por Johansen e Juselius (1990), foi utilizado para testar a co-integração das séries temporais deste trabalho. Segundo esses autores, as hipóteses do teste traço  $\lambda_{traço}$  são montadas de forma a verificar a existência do número máximo de ( $r$ ) vetores co-integrados, ou seja,  $H_0: r_0 \leq r$  e  $H_1: r_0 > r$ . A estatística do teste traço é denotada por:

$$\lambda_{traço} = -T \sum_{i=r_0+1}^p \ln(1 - \lambda_i) \quad \text{com } r = 0, 1, 2, \dots, p-1 \quad (3.6)$$

em que  $T$  é o número de observações e  $\lambda_i$  são os autovetores estimados.

Valores calculados para a estatística traço maiores que os níveis críticos calculados por Johansen e Juselius (1990) implicam em rejeição de  $H_0$ , concluindo-se que as séries analisadas são co-integradas.

### 3.6

#### O Modelo VAR

A utilização de modelos de equações simultâneas, em virtude das limitações verificadas na crítica de Lucas (1976) e Sims (1980), deixou de ser o centro das atenções no que tange à metodologia utilizada para a resolução de problemas econométricos e cedeu espaço a novos modelos.

Lucas (1976) notou que a ocorrência de alterações nas políticas econômicas influenciava profundamente os resultados dos modelos estruturais. Com isso, ocorriam modificações nos coeficientes analisados, o que implicava na quebra estrutural da série. Outra crítica aos modelos estruturados partiu de Sims (1980). Para ele, na tentativa de tornar os modelos identificáveis, muitas vezes utilizavam-se restrições demasiadamente fortes, o que diminuía a qualidade e a eficácia da previsão.

Através do Modelo Vetorial Auto-Regressivo (VAR), introduzido por Sims (1980), tornou-se possível expressar modelos econômicos completos, bem como estimar seus parâmetros. Nesse modelo as variáveis são conjuntamente determinadas, ou seja, são explicadas por suas defasagens e pelas defasagens das demais variáveis. A vantagem dessa forma de especificação é que ela torna possível analisar o efeito da variação ao longo do tempo de determinada variável sobre as demais. Outros aspectos positivos da metodologia VAR são a não necessidade de impor qualquer restrição inicial de causalidade entre as variáveis - uma vez que isso pode ser verificado pelo teste de Granger - bem como o fato de não ser preciso assumir alguma relação de longo prazo entre as variáveis - o que também pode ser testado pelo método proposto por Johansen.

A metodologia VAR, pelas razões acima mencionadas, mostrou-se superior em muitos testes de previsão aos modelos estruturados.

Suponha que as inter-relações entre duas variáveis,  $y$  e  $z$ , possam ser descritas pelo seguinte modelo:

$$y_t = \alpha_0 + \alpha_1 z_t + \alpha_2 z_{t-1} + e_t^y \quad (3.7)$$

$$z_t = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + e_t^z \quad (3.8)$$

Substituindo  $z_t$  na primeira equação por sua definição dada na segunda equação e remanejando os termos, obtemos a seguinte expressão para  $y_t$ :

$$y_t = \alpha_0 + \alpha_1(\beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + e^z) + e^y \quad (3.9)$$

$$\Rightarrow y_t = \frac{\alpha_0 + \alpha_1 \beta_0}{1 - \alpha_1 \beta_1} + \left( \frac{\alpha_1 \beta_2}{1 - \alpha_1 \beta_1} \right) y_{t-1} + \left( \frac{\alpha_2}{1 - \alpha_1 \beta_1} \right) z_{t-1} + e^y + \frac{\alpha_1 e_t^z + e_t^y}{1 - \alpha_1 \beta_1} \quad (3.10)$$

Usando essa expressão para eliminar  $y_t$  da segunda equação do modelo estrutural, obtemos:

$$z_t = \beta_0 + \beta_1 \left[ \frac{\alpha_0 + \alpha_1 \beta_0}{1 - \alpha_1 \beta_1} + \left( \frac{\alpha_1 \beta_2}{1 - \alpha_1 \beta_1} \right) y_{t-1} + \left( \frac{\alpha_2}{1 - \alpha_1 \beta_1} \right) z_{t-1} + e^y + \frac{\alpha_1 e_t^z + e_t^y}{1 - \alpha_1 \beta_1} \right] + \beta_2 y_{t-1} + e_t^z \quad (3.11)$$

$$\Rightarrow z_t = \beta_0 + \frac{(\alpha_0 + \alpha_1 \beta_0) \beta_1}{1 - \alpha_1 \beta_1} + \left( \frac{\alpha_1 \beta_2 \beta_1}{1 - \alpha_1 \beta_1} + \beta_2 \right) y_{t-1} + \left( \frac{\alpha_2 \beta_1}{1 - \alpha_1 \beta_1} \right) z_{t-1} + \frac{\alpha_1 e_t^z + e_t^y}{1 - \alpha_1 \beta_1} \quad (3.12)$$

Logo, pode-se escrever a “forma reduzida” do modelo (isto é, a “solução” do modelo em função das variáveis predeterminadas e dos choques) como:

$$y_t = \Pi_0^1 + \Pi_1^1 y_{t-1} + \Pi_2^1 z_{t-1} + u_t^1 \quad (3.13)$$

$$z_t = \Pi_0^2 + \Pi_1^2 y_{t-1} + \Pi_2^2 z_{t-1} + u_t^2 \quad (3.14)$$

Note que se trata de um sistema onde  $y$  e  $z$  dependem apenas de suas defasagens, além de choques que correspondem a combinações lineares dos choques do modelo estrutural. Definindo:

$$x_t = \begin{pmatrix} y_t \\ z_t \end{pmatrix}, \quad \Pi_0 = \begin{pmatrix} \Pi_0^1 \\ \Pi_0^2 \end{pmatrix}, \quad \Pi_1 = \begin{pmatrix} \Pi_1^1 & \Pi_1^2 \\ \Pi_2^1 & \Pi_2^2 \end{pmatrix} U_t = \begin{pmatrix} u_t^1 \\ u_t^2 \end{pmatrix} \quad (3.15)$$

Podemos reescrevê-la na forma reduzida como:

$$x_t = \Pi_0 + \Pi_1 x_{t-1} + u_t \quad (3.16)$$

que é um processo auto-regressivo para o vetor  $x$ . Daí o nome “modelo auto-regressivo vetorial” (“Vector-AutoRegressive Model” - VAR).

No exemplo acima, tem-se um modelo VAR de ordem 1, ou VAR(1), pois o vetor  $x$  depende apenas de sua primeira defasagem. Em termos mais gerais, possível expressar um modelo VAR estrutural por:

$$x_t = \Pi_0 + \Pi_1 x_{t-1} + \Pi_2 x_{t-2} + \dots + \Pi_p x_{t-p} + u_t \quad (3.17)$$

$$E(u_t) = 0 \quad E(u_t u_t') = \begin{cases} S & t = T \\ 0 & t \neq T \end{cases}$$

Onde:

$x_t =$  vetor ( $nx1$ ) de variáveis

$\Pi_0 =$  vetor ( $nx1$ ) de interceptos

$\Pi_i =$  matrizes ( $nxn$ ) de coeficientes

$u_t =$  vetor ( $nx1$ ) de distúrbios

$\Sigma =$  matrizes ( $nxn$ ) de covariâncias dos distúrbios

Note que todas as variáveis do VAR têm a mesma ordem de defasagem em todas as equações.

### 3.6.1

#### Estimação

Observa-se que o lado direito do VAR contém apenas variáveis predeterminadas.

Além disso, os distúrbios de cada equação satisfazem as propriedades clássicas.

Logo, cada equação pode ser estimada por MQO.

A matriz de variância-covariância dos resíduos também pode ser estimada da forma usual, a partir dos resíduos da estimação por MQO.

### 3.6.2

#### Especificação

Usualmente, como no caso univariado, utilizam-se critérios de informação para determinar a ordem de defasagens de um modelo VAR. O argumento para a utilização do critério de informação consiste no seguinte fato: cada regressor que é adicionado ao modelo faz com que a soma dos resíduos não aumente e, muitas vezes, diminua. Isso implica que a redução da soma dos resíduos ocorre devido ao aumento do número de regressores. Para balancear a redução dos erros e o aumento do número de regressores, é associada uma penalidade a esse aumento. Quando a penalidade decorrente da incorporação de um regressor adicional for menor que a diminuição da soma de resíduos, deve-se adicioná-lo ao modelo. Caso contrário, o regressor não deve ser utilizado.

Dado um modelo VAR(p), onde  $m = 0,1,2,\dots,p$ , deve-se escolher a ordem p que minimiza a seguinte fórmula geral do critério de informação:

Os critérios de informação mais utilizados na literatura são os de Akaike (AIC), Hannan-Quinn (HQ) e Schwartz (SIC ou BIC). Sendo n o tamanho da amostra, p a ordem do defasamento, e  $\Sigma$  da matriz de covariância dos resíduos. Tais critérios são generalizações diretas dos modelos uniequacionais.

$$AIC = \ln[\det(\hat{\Sigma})] + \frac{2}{T}pn^2 \quad (3.18)$$

$$HQ = \ln[\det(\hat{\Sigma})] + \frac{\ln(\ln(T))}{T}pn^2 \quad (3.19)$$

$$BIC = \ln[\det(\hat{\Sigma})] + \frac{\ln(T)}{T}pn^2 \quad (3.20)$$

Quanto mais regressores são incorporados no mesmo período da amostra, menor será o erro estimado. Por outro lado, esse maior número de regressores sofrerá uma penalidade na segunda parte da estatística, o que confirma que a escolha de um modelo mais parcimonioso – no sentido da utilização de menos parâmetros – produz menos imprecisão das estimativas.

Em relação ao comportamento entre si dos critérios mencionados tem-se que para grandes amostras o critério BIC é o mais consistente, tendendo a escolher um modelo mais parcimonioso que o AIC. Por outro lado, para pequenas amostras, o AIC funciona melhor que o BIC e o HQ.

### 3.6.3

#### Modelos Vetoriais de Correção de Erros (MVEC)

Para a especificação correta do modelo, o primeiro passo é testar se as variáveis são estacionárias ou processos integrados. O teste utilizado para analisar a presença ou não de raiz unitária foi o de Dickey-Fuller Aumentado (ADF). Através desse teste é possível verificar o comportamento da série, identificando a presença de não-estacionariedade pela incidência de tendências estocástica, determinísticas ou pela junção de ambas.

Caso as variáveis analisadas sejam não-estacionárias,  $I(1)$ <sup>1</sup> por exemplo, o VAR deve ser estimado em primeiras diferenças. Contudo, o resultado obtido por meio de primeiras diferenças nem sempre é razoável (HARVEY, 1989) uma vez que existem informações nos níveis que são desconsideradas.

Ao optar por trabalhar com variáveis em nível, duas dificuldades podem aparecer: o aparecimento de regressões espúrias em séries não estacionárias e o fato de que as estimações e testes para regressões não estacionárias não serem padronizados.

Um caso particular surge quando as séries em nível são co-integradas. Nessa situação é possível mostrar que os estimadores são superconsistentes, isto é, convergem mais rapidamente que o usual, e as informações sobre os níveis continuam mantidas (HOLLAUER, BAHIA, ISSLER, 2006). Na presença de co-integração Engle e Granger (1987) recomendam utilizar uma representação MCE, que explicita a utilização do modelo de correção de erro, o qual representa o comportamento de longo prazo.

O Modelo de Correção de Erros (MCE) pode ser escrito da maneira a seguir.

---

<sup>1</sup> Ou seja, as séries são estacionárias na primeira diferença.

$$\Delta x_t = \Delta \Pi_0 + \Delta x_{t-1} + \Phi_1 \Delta x_{t-1} + \Phi_2 \Delta x_{t-2} \dots + \Phi_{p-1} \Delta x_{t-(p-1)} + u_t \quad (3.21)$$

Ou seja, trata-se do VAR original reescrito em função das primeiras diferenças do vetor  $x$ , defasadas até a ordem  $p-1$ , e de  $x$  defasado de um período. O vetor  $\Delta x_{t-1}$  contém justamente a relação de longo prazo existente entre as variáveis do vetor  $x - p$  que deixa claro por quê um modelo especificado somente em primeiras diferenças, que ignora esse termo, “joga fora” as informações referentes ao longo prazo.

O uso da abordagem VECM requer testar a existência de cointegração entre as variáveis do vetor  $x$  e estimar as relações eventualmente existentes, o que não é trivial em muitos casos. **Por outro lado, o VECM apresenta, em relação ao VAR “irrestrito”, a vantagem de que a imposição das relações de cointegração conhecidas pode propiciar significativos ganhos de eficiência na estimação e, conseqüentemente, na previsão.**

#### 3.6.4

##### Inclusão de Termos Determinísticos

A inclusão de termos determinísticos no modelo (constante, tendência, dummies sazonais, dummies de intercepto etc.) é frequentemente necessária para retratar adequadamente as características observadas nas séries de interesse.

Entretanto, a inclusão de todos os termos determinísticos necessários pode diminuir significativamente os graus de liberdade na estimação do VAR (principalmente quando se trata de dummies sazonais). Tendo em vista que um VAR sem termos determinísticos já inclui número relativamente grande de coeficientes, esse pode ser um problema bastante severo. É importante, portanto, investir algum tempo numa análise prévia dos dados, a fim de só incluir termos determinísticos que sejam realmente necessários.

Para este estudo de caso, as intervenções dar-se-ão no mês de dezembro de 2008 e Janeiro de 2009.

### 3.6.5

#### Teste para Quebra Estrutural

Estacionariedade no processo de geração de dados é uma importante condição para derivar as propriedades dos estimadores e elaborar previsões pontuais e intervalares. Também é uma propriedade que garante que as médias, variâncias e covariâncias são constantes ao longo do tempo. Muitas séries econômicas não são estacionárias. Podem conter sazonalidades e tendências que fazem com que a variância seja heterocedástica ao longo do tempo.

Estes componentes podem ser eliminados através de algumas transformações simples, por exemplo, diferenciando as séries e incluindo dummies sazonais. No entanto outra grande origem da não estacionariedade das séries econômicas são eventos que causam turbulência no sistema econômico em um país ou no conjunto de países. São eventos que não exibem tendências de longo prazo, não são sazonais, porém podem causar quebra econômica estrutural. Estes eventos ocorrem em alguns períodos específicos como guerras, implementação de pacotes econômicos fiscais e de política monetária, mudanças de governo, choques de preços do petróleo, mudanças de cambiais e crises financeiras.

O teste para de Chow para quebra estrutural examina se uma variação nos parâmetros ocorreu em algum ponto do tempo pela comparação de parâmetros estimados antes e depois da possível data da quebra estrutural.

Supõe-se que existe uma suspeita de uma variação nos parâmetros do processo VAR(p), após o período  $T_1 > T$ .

No teste de Chow são requeridas dois conjuntos de amostras  $Y_1$  e  $Y_2$ . O primeiro conjunto são amostras  $y_1, \dots, y_T$ , usualmente inseridas no VAR e o conjunto  $Y_{(2)}$  são as amostras da previsão pelo VAR. Conseqüentemente, tem-se as matrizes dos coeficientes estimados,  $B_1$  e  $B_2$ . O teste de Chow é montado com a seguinte estimação:

$$[Y_1 : Y_2] = [B_1 : B_2]Z + [U_1 : U_2] = BZ + U \quad (3.22)$$

Onde:

$$Z = \begin{bmatrix} Z_{(1)} & 0 \\ 0 & Z_{(2)} \end{bmatrix}$$

$$Z_{(1)} = [Z_0, \dots, Z_{T-1}]$$

$$Z_{(2)} = [Z_{T1}, \dots, Z_{T-1}]$$

$$Z'_t = [1, y_t, \dots, Y_{t-p+1}]$$

A hipótese nula e alternativa baseia-se na verificação das matrizes dos parâmetros dos dois conjuntos de amostras,  $B_1$  e  $B_2$ . As hipóteses para teste são:

$$H_0: B_1 = B_2$$

$$H_1: B_1 \neq B_2$$

Pelo método de estimação MQ ou MV,  $B_1$  e  $B_2$  podem ser facilmente obtidos.

O Teste de Chow de quebra no ponto ou *break point* testa se existe uma quebra estrutural em um período  $T_B$  qualquer. O modelo em consideração é calculado usando todas as amostras, do primeiro a última observação,  $T_1$  e  $T_2$ , onde  $T_1 < T_B$  e  $T_2 \leq T - T_B$ . Denotando os resíduos resultantes do VAR,  $\hat{u}_t$ ,  $u_t^{(1)}$  e  $u_t^{(2)}$  e usando as seguintes notações para matriz de covariância  $\Sigma_u$ .

$$\hat{\Sigma}_u = \frac{\sum_{t=1}^T \hat{\mu}_t \hat{\mu}_t'}{T} \quad (3.23)$$

$$\hat{\Sigma}_{1,2} = \frac{\sum_{t=1}^{T_1} \hat{\mu}_t \hat{\mu}_t' + \sum_{t=T-T_2+1}^{T_1} \hat{\mu}_t \hat{\mu}_t'}{T_1 + T_2} \quad (3.24)$$

$$\hat{\Sigma}_{(1)} = \frac{\sum_{t=1}^{T_1} \hat{\mu}_t^{(1)} \hat{\mu}_t^{(1)'}}{T_1} \quad (3.25)$$

$$\hat{\Sigma}_{(2)} = \frac{\sum_{t=T-T_2+1}^T \hat{\mu}_t^{(2)} \hat{\mu}_t^{(2)'}}{T_2} \quad (3.26)$$

A estatística BP de Chow segue a distribuição  $\chi^2$ , sendo é dada por:

$$\lambda_{BP} = (T_1 + T_2) \log \det \hat{\Sigma}_{1,2} - T_1 \log \det \hat{\Sigma}_{(1)} - T_2 \log \det \hat{\Sigma}_{(2)} \approx \chi^2(K) \quad (3.27)$$

Onde  $k$  é a diferença entre a soma do número de parâmetros estimados no primeiro e último sub-período e o número de parâmetros no modelo completo, incluindo potencialmente os diferentes parâmetros da matriz de covariância dos resíduos ruído branco.

No teste BP  $\lambda_{BP}$ , a hipótese nula de parâmetros constantes ou de que não há quebra estrutural é rejeitada se BP  $\lambda_{BP} > \chi^2(k)$ .

### 3.6.6

#### Análise de Resíduos

Na seção 4.6.2, mostramos procedimentos para escolher a ordem do VAR. Os critérios para escolha da ordem do modelo podem ser considerados como critérios de decisão de resíduos. **Na teoria os resíduos devem satisfazer a condição de ruído branco. Na prática, segundo Lutkepohl (2005, p. 157), se o objetivo é previsão e o modelo prevê bem, a condição de ruído branco pode não ser de importância principal.**

Porém, há situações que existe interesse para conferir a hipótese do ruído branco para os resíduos de um modelo particular. Por exemplo, se a ordem do modelo é escolhida através da teoria econômica e não estatisticamente, pode ser útil usar ferramentas estatísticas disponíveis para investigar as propriedades dos resíduos.

Nesta seção, apresentam-se ferramentas estatísticas para conferir as propriedades de auto-correlação dos resíduos do VAR. Especificamente, baseado na hipótese que os resíduos são realmente ruído branco. Em seguida é mostrada uma estatística para conferir o significado global da autocorrelação residual: o teste LM. Por fim abordaremos o teste de normalidade Jarque - Bera.

Analogamente, os testes de autocorrelação dos resíduos no âmbito de um VAR são generalizações dos testes usados em modelos uniequacionais.

### 3.6.6.1

#### Teste Multiplicadores de Lagrange

O teste LM Breusch-Godfrey para auto-correlação residual foi proposto por Breusch (1978) e Godfrey (1978), com modificação de Godfrey (1988). O teste parte do modelo VAR. Assumindo um vetor de erros,  $U_t = D_1 U_{t-1} + D_2 U_{t-2} + \dots + D_h U_{t-h} + v_t$  é um ruído branco, o teste tem como hipótese nula a não correlação residual. A Hipótese nula e alternativa do teste são:

$$H_0: D_1 = D_2 = \dots = D_h = 0$$

$$H_1: D_j \neq 0 \text{ para pelo menos um } j \in \{0, 1, 2, \dots, h\}$$

O teste estatístico é determinado com ajuda da regressão auxiliar, dado por:

$$u_t = A_1 Y_{t-1} + \dots + A_p Y_{t-p} + D_1 \hat{u}_{t-1} + \dots + D_h \hat{u}_{t-h} + \varepsilon_t \quad (3.28)$$

Para  $t=1, 2, \dots, T$

Em particular os resíduos  $\hat{u}_t$  são obtidos do VAR estimado por MQ. Caso  $\hat{u}_t = 0, t \leq 0$ . Matricialmente, o modelo de regressão auxiliar é:

$$\hat{U} = BZ + D\hat{u} + \varepsilon \quad (3.29)$$

Onde  $D_{(k \times k_h)} = [D_1, \dots, D_h]$ ;  $\hat{u} = (I_h \otimes \hat{U})F$ ;  $\varepsilon_{k \times t} = [\varepsilon_1 \dots \dots \varepsilon_t]$ .

O Estimador MQ de  $[B:D]$ , do modelo auxiliar é:

$$[B:D] = \hat{U}[Z': \hat{u}] \left( \begin{pmatrix} Z \\ \hat{u} \end{pmatrix} (Z \quad \hat{u})^{-1} \right) \quad (3.30)$$

Aplicando regras de partição de matrizes<sup>2</sup>, chega-se a:

$$\hat{D} = \hat{U}\hat{u} \left( (\hat{u}\hat{u}' - \hat{u}Z'(ZZ')^{-1}Z\hat{u}) \right)^{-1} \quad (3.31)$$

<sup>2</sup> Ver Lutkepohl(2005,p659,regra2).

Sob a hipótese nula, a estatística do teste LM Breusch-Godfrey segue uma distribuição  $\chi^2$ . Formalmente, temos:

$$LM_h = \lambda_{LM} = T \hat{c}_h' \hat{\Sigma}_c(h)^{-1} \hat{c}_h \xrightarrow{d} \chi^2(hK^2) \quad (3.32)$$

Onde:

$$\hat{c}_h = T(\hat{U}\hat{u}) e \sum_c^{\wedge}(h) = T^{-1}[\hat{u}\hat{u}' - \hat{u}Z'(ZZ')^{-1}Z\hat{u}]^{-1} \otimes \sum_u \sum$$

### 3.6.6.2

#### Teste para Normalidade dos Resíduos do VAR

Para montar intervalos de previsão, a normalidade dos dados  $y_t$  gerados por processos é uma condição clássica desejada. Resíduos que não seguem a normalidade gaussiana indicam, geralmente, que o modelo VAR não é uma boa representação do processo de geração de dados  $y_t$ . A normalidade de  $y_t$  pode ser observada pelos resíduos  $u_t$ . Na prática os resíduos verdadeiros são substituídos por resíduos do modelo VAR. Para fazer a verificação, testa-se a hipótese de normalidade sobre a distribuição de probabilidade dos resíduos do VAR.

Apresenta-se nesta seção testes para normalidade dos resíduos. Especificamente será mostrado o teste Jarque-Bera para normalidade multivariada.

O teste Jarque-Bera (1987) para normalidade é baseado no terceiro e quarto momentos centrais da distribuição normal padronizada, respectivamente assimetria e curtose. O teste verifica se os momentos da série univariada estimada,  $x$ , são iguais aos da distribuição normal padronizada. Sob esta hipótese, a assimetria é igual a 0,  $E(x^3)=0$ , e a curtose é igual a 3,  $E(x^4)=3$ . Tecnicamente, testam-se as duas hipóteses conjuntamente.

Dado  $y_t$  estacionário,  $K$ -dimensional, com processo VAR gaussiano estável, onde  $u_t$  é um processo ruído branco gaussiano de média zero, com matriz de covariância não singular  $\Sigma_u$  e  $\hat{A}_1, \dots, \hat{A}_p$  os estimadores dos coeficientes, consistentes e assintoticamente normalmente distribuídos, calculados por MV ou MQO, baseados nas amostras  $y_1, y_2, \dots, y_T$ , os resíduos do VAR e a matriz de covariância são:

$$u_t = (Y_t - \bar{Y}) - \hat{A}_1(Y_{t-1} - \bar{Y}) - \dots + \hat{A}_p(Y_{t-p} - \bar{Y}) \quad (3.33)$$

Para  $t=1,2,\dots,T$

$$\sum_u^{\wedge} = \frac{\sum_{t=1}^T \hat{\mu}_t \hat{\mu}_t'}{T - KP - 1} \quad (3.34)$$

Assume-se que os resíduos são não correlacionados e, portanto, são ortogonais. E também a matriz  $\hat{P}$  de fatorização, satisfaz:

$$\hat{P}\hat{P}' = \sum_u^{\wedge} \quad (3.35)$$

Sendo:

$$\text{Plim}(\hat{P} - P) = 0 \quad (3.36)$$

Ademais:

$$\hat{P}^{-1}\hat{\mu}_t = \hat{w}_t \sim N(0, I_k) \text{ para } t=1,\dots,T \quad (3.37)$$

$$\hat{w}_t = (\hat{w}_{1t} \dots \dots \hat{w}_{kt})' \quad (3.38)$$

Onde  $w_t$  são variáveis randômicas independentes e normalmente distribuídas, iid. As esperanças de assimetria e curtose, respectivamente, são:

$$E \begin{pmatrix} w_{1t}^3 \\ \vdots \\ w_{kt}^3 \end{pmatrix} = 0 \text{ e } E \begin{pmatrix} w_{1t}^4 \\ \vdots \\ w_{kt}^4 \end{pmatrix} = 3k \quad (3.39)$$

Para construir o teste utilizam-se os resíduos estimados pelo VAR,  $u_t$ 's. Os resultados são usados para testar a hipótese nula de normalidade do processo ruído branco  $u_t$  do VAR, contra a hipótese alternativa de não normalidade, ou seja:

$$H_0: E \begin{pmatrix} \widehat{W}_{1t}^3 \\ \vdots \\ \widehat{W}_{kt}^3 \end{pmatrix} = 3k \text{ contra } H_1: E \begin{pmatrix} \widehat{W}_{1t}^4 \\ \vdots \\ \widehat{W}_{kt}^4 \end{pmatrix} \neq 0$$

$$H_0: E \begin{pmatrix} \widehat{W}_{1t}^3 \\ \vdots \\ \widehat{W}_{kt}^3 \end{pmatrix} = 3k \text{ contra } H_1: E \begin{pmatrix} \widehat{W}_{1t}^4 \\ \vdots \\ \widehat{W}_{kt}^4 \end{pmatrix} \neq 3k$$

Os valores da assimetria e curtose estimados, dados respectivamente por  $\widehat{b}_1$  e  $\widehat{b}_2$  são:

$$\widehat{b}_1 = (\widehat{b}_{11} \dots \widehat{b}_{k1})' \text{ com } \widehat{b}_{k1} = \frac{\sum_{t=1}^T \widehat{W}_{kt}^3}{T} \text{ para } K = 1 \dots k \quad (3.40)$$

$$\widehat{b}_2 = (\widehat{b}_{12} \dots \widehat{b}_{k2})' \text{ com } \widehat{b}_{k2} = \frac{\sum_{t=1}^T \widehat{W}_{kt}^4}{T} \text{ para } K = 1 \dots k \quad (3.41)$$

Com a distribuição assintótica de assimetria e curtose,  $\widehat{u}_t \sim N(\mu_1, \Sigma_u)$ . Então  $b_1$  e  $b_2$  são assintoticamente iid:

$$\sqrt{T} \begin{pmatrix} \widehat{b}_1 & 0 \\ \widehat{b}_2 & -3K \end{pmatrix} \xrightarrow{d} N \left( 0, \begin{pmatrix} 6IK & 0 \\ 0 & 24Ik \end{pmatrix} \right) \quad (3.42)$$

A proposição da distribuição assintótica de assimetria e curtose implica que:

$$\widehat{\lambda}_s = T \frac{\widehat{b}_1' \widehat{b}_1}{6} \xrightarrow{d} \chi^2(K) \quad (3.43)$$

$$\widehat{\lambda}_k = T \frac{(\widehat{b}_2 - 3K)' (\widehat{b}_2 - 3K)}{24} \xrightarrow{d} \chi^2(K) \quad (3.44)$$

### 3.7

#### O Modelo BVAR

Os modelos de vetores auto-regressivos (VAR) surgiram na década de 80 como resposta às críticas ao grande número de restrições impostas às estimações pelos modelos estruturais. A idéia era desenvolver modelos dinâmicos com o mínimo de restrições, nos quais todas as variáveis econômicas fossem tratadas como endógenas. Sendo assim, os modelos VAR examinam relações lineares entre cada variável e os valores defasados dela própria e de todas as demais variáveis, impondo como restrições à estrutura da economia somente: a escolha do conjunto relevante de variáveis e do número máximo de defasagens envolvidas nas relações entre elas. Nos modelos VAR, o número de defasagens é normalmente escolhido com base em critérios estatísticos, como os de Akaike ou Schwarz.

Os modelos VAR, evidentemente, têm as suas limitações, que foram objeto de um grande volume de pesquisa nas décadas seguintes. Duas limitações são lembradas com frequência. A primeira refere-se ao elevado número de parâmetros dos modelos VAR, com reflexo no tamanho de amostra requerido para que se obtenha uma estimação confiável. De fato, o número de parâmetros a ser estimado,  $K(Kp+K)$ , cresce geometricamente com o número de variáveis,  $K$ , e proporcionalmente com o número de defasagens,  $p$ . Quando o número de variáveis do VAR é relativamente alto e o período da amostra é relativamente baixo, é provável que as estimativas sejam influenciadas pelo ruído ao invés do sinal das variáveis. Quando isto acontece, é recomendável calcular os parâmetros do VAR impondo restrições para reduzir a dimensão do espaço dos parâmetros. Então, o problema é encontrar as restrições que são coerentes com o modelo<sup>3</sup>. A segunda diz respeito ao fato de que cada modelo VAR é simplesmente uma “forma reduzida”, ou seja, as mesmas relações entre as variáveis e suas defasagens são simultaneamente compatíveis com vários diferentes modelos que descrevem também as relações contemporâneas entre as variáveis (chamados de “formas estruturais”).

---

<sup>3</sup> Ver Lutkepohl (2005, Capítulo 5)

Uma maneira simples de amenizar a sobre-parametrização dos modelos VAR é impor que os coeficientes de algumas variáveis sejam iguais a zero. Entretanto, segundo Litterman (1986), o investigador não pode estar seguro que alguns coeficientes são zero e, portanto, não devem ignorar uma possível variação. A perspectiva bayesiana pode ajudar a resolver tal situação. Em outras palavras, existe incerteza sobre o valor exato dos parâmetros do modelo como uma distribuição de probabilidade para o vetor de parâmetros sem inserir “pesos” em certos valores. Os modelos BVAR (Bayesian Vector Autoregression) surgiram como uma resposta mais satisfatória a esse problema. Nos modelos BVAR, **em vez de se impor a exclusão de determinadas variáveis, opta-se por estipular uma distribuição de probabilidade *a priori* (informativa) para cada um dos coeficientes. Essa distribuição *a priori* é combinada com a informação amostral para gerar as estimações.** Esse processo, portanto, difere da estimação clássica utilizada nos modelos VAR.

### 3.7.1

#### Estatística Bayesiana Econométrica

Na abordagem econométrica bayesiana assume-se uma informação *prior* obtida por uma função densidade de probabilidade. Denotando os parâmetros de interesse por  $\alpha$ , a informação *prior* é resumida pela função densidade de probabilidade prior,  $g(\alpha)$ . A informação da amostra é resumida na função densidade de probabilidade (FDP), condicional ao parâmetro  $\alpha$ , dito  $f(y|\alpha)$ . A função  $f(y|\alpha)$  é algebricamente idêntico a função de verossimilhança  $l(\alpha|y)$ . Os dois tipos de informação,  $g(\alpha)$  e  $f(y|\alpha)$  são combinadas pelo teorema de Bayes:

$$g(\alpha|y) = \frac{f(\alpha|y)*g(\alpha)}{f(y)} \quad (3.45)$$

Onde  $f(y)$  é função densidade da amostra incondicional que, para uma determinada amostra, é uma constante normalizada. Em outras palavras, a distribuição de  $\alpha$ , dada a informação amostral contida em  $y$ , é resumida através da função densidade de probabilidade posterior,  $g(\alpha|y)$ . Essa função é proporcional a

função de verossimilhança,  $l(y|\alpha)$ , e a função de densidade de probabilidade prior,  $g(\alpha)$ . Portanto:

$$g(\alpha|y) = f(y|\alpha) * g(\alpha) \propto l(\alpha/y) g(\alpha) \quad (3.46)$$

A densidade condicional,  $g(\alpha|y)$ , contém toda a informação disponível no vetor de parâmetros,  $\alpha$ . Estimadores pontuais de  $\alpha$  podem ser derivados da distribuição posterior. Por exemplo, a média da distribuição posterior, chamada de média posterior, é frequentemente usada como um estimador de ponto para  $\alpha$ . Para incorporar a estimação bayesiana no processo VAR, necessita-se, portanto, saber a distribuição prior a ser utilizada no modelo.

Kadyiala-Karlsson (1997) e Ciccarelli-Rebucci (2003) revisam algumas distribuições prior como: *Normal, de Minnesota ou Litterman, Diffuse e Distribuições Conjungadas (Normal-Wishart, Normal-Diffuse e Natural Estendida)*.

### 3.7.1.1

#### Litterman ou Minnesota Priors

Doan-Litterman-Sims (1984) e Litterman (1980, 1986) sugeriram uma distribuição prior específica para os parâmetros do VAR, hoje conhecida na literatura bayesiana como distribuição prior de Minnesota ou Litterman<sup>4</sup>.

A Minnesota prior consiste em usar um pequeno conjunto de hiperparâmetros para caracterizar a distribuição adequadamente. No caso de processos estacionários, se a dependência intertemporal das variáveis é fraca, um modo para descrever isso é fixar a média das prior's dos coeficientes do VAR em zero, com variâncias prior diferentes de zero, ou seja,  $\alpha^*=0$  e  $V_{\alpha^*} \neq 0$ . Com essa escolha de  $\alpha^*$  a média *posterior*<sup>5</sup> em é:

---

<sup>4</sup> Inicialmente foi recomendada pelos autores para ser utilizada em certos processos não estacionários. No entanto, hoje, existem adaptações da Minnesota prior para processos estacionários.

<sup>5</sup> Este estimador para  $\alpha$  parece com o estimador multivariado LS com exceção da matriz de covariância inversa  $V_{\alpha}^{-1}$ . Na prática  $\Sigma_u$  é substituído por  $\hat{\Sigma}_u$ .

$$\bar{\alpha} = [V_{\alpha}^{-1} + (X'X \otimes \Sigma_u^{-1})]^{-1}[(X' \otimes \Sigma_u^{-1})]y \quad (3.47)$$

Litterman (1986) especifica a matriz de covariância prior,  $V_{\alpha}$ , como uma matriz diagonal. Os elementos de  $V_{\alpha}$  são dados por:

$$v_{ij,1} = \begin{cases} (\lambda/1)^2 & \text{se } i = j \\ (\lambda\theta\sigma_1)^2/1\sigma_j & \text{se } i \neq j \end{cases} \quad (3.48)$$

Onde  $v_{ij,1}$  é a variância prior de  $\alpha_{ij,1}$ ;  $\lambda$  é o desvio-padrão prior dos coeficientes  $\alpha_{kk,1}$  com  $k= 1, \dots, K$ ;  $\sigma_i^2$  é o  $i$ -th elemento da diagonal de  $\Sigma_u$  e  $\theta$  especifica a fração do desvio-padrão prior,  $\lambda$ , ligado aos coeficientes defasados das variáveis explicativas, em cada equação. Como restrição,  $0 < \theta < 1$ .

Para cada equação,  $\lambda$  controla como o coeficiente do primeiro lag da variável dependente diminui de maneira harmônica e concentrada ao redor de zero. Na prática, diferentes valores de  $\lambda$  podem ser experimentados. Inclusive diferentes valores em diferentes equações, podem ser considerados.

A variância prior diminui com o aumento da defasagem,  $\mathbf{1}$ , pela crença provável que os coeficientes com ordem alta de defasagem são próximos de zero. Além disso, existe a crença que a maioria da variação em cada uma das variáveis é devida à própria defasagem. Então, coeficientes de variáveis diferentes da variável dependente têm uma variância relativa menor, escolhendo  $\theta$  entre 0 e 1.

A relação  $(\sigma_1/\sigma_j)^2$  é incluída pela preocupação das diferenças na variabilidade das diferentes variáveis. Assume-se que a resposta de uma variável em relação à outra é determinada por movimentos inesperados refletidas, portanto, na variância residual. Portanto, as variâncias residuais são preferidas em relação às variâncias de  $y_k$ .

### 3.8

#### Previsões não Condicionadas e o Teste CP

Conforme dito anteriormente, a realização de previsões a partir de um modelo uniequacional requer hipóteses sobre a evolução futura das variáveis explicativas, para que seja possível projetar valores esperados para a variável endógena.

Por exemplo, suponha que estejamos trabalhando com a primeira equação de um VAR bivariado de ordem 2:

$$y_t = \Pi_0^1 + \Pi_1^1 y_{t-1} + \Pi_1^1 y_{t-2} + \Pi_2^1 z_{t-1} + \Pi_2^1 z_{t-2} + u_t^1 \quad (3.49)$$

Claramente, previsões de  $y$  mais do que um passo à frente (isto é, para  $t+2$ ,  $t+3$ , etc.) exigem também previsões para  $z$  (em  $t+1$ ,  $t+2$ , etc.).

Ou seja, previsões de  $y$  a partir desse modelo uniequacional são condicionadas às previsões da variável explicativa  $z$ .

Mas se trabalharmos com o **VAR completo**:

$$y_t = \Pi_0^1 + \Pi_1^1 y_{t-1} + \Pi_1^1 y_{t-2} + \Pi_2^1 z_{t-1} + \Pi_2^1 z_{t-2} + u_t^1 \quad (3.50)$$

$$z_t = \Pi_0^2 + \Pi_1^2 y_{t-1} + \Pi_1^2 y_{t-2} + \Pi_2^2 z_{t-1} + \Pi_2^2 z_{t-2} + u_t^2 \quad (3.51)$$

Podemos gerar previsões para as variáveis endógenas  $y$  e  $z$  sem que seja necessário condicioná-las à evolução de alguma variável exógena. Ou seja, trata-se de **previsões não condicionadas**.

O teste de capacidade preditiva (CP) consiste em apresentar qual dos modelos terá melhor desempenho através da capacidade preditiva que pode ser gráfica ou métrica.

A análise métrica consiste na avaliação do desempenho dos candidatos através do indicador MAPE.

O erro médio absoluto percentual ou MAPE é o valor médio do erro das previsões sobre todo o conjunto de teste. O MAPE é definido da seguinte forma:

$$MAPE = \frac{\sum_1^N \left| \frac{(Ajustado - Observado)}{Observado} \right|}{\# \text{ Previsões}} \quad (3.52)$$