

### 3 Referencial teórico

Neste capítulo, será apresentado de forma resumida o arcabouço teórico que suporta a análise dos dados e conclusões deste estudo.

#### 3.1. Análise de séries temporais

Segundo Morettin (2006, p.1), “*uma série temporal é qualquer conjunto de observações ordenadas no tempo*”. Downing et al. (2005, p.299) apresenta uma definição um pouco mais restrita: “*séries temporais (ou séries históricas) são conjuntos de medidas de uma mesma grandeza, relativas a vários períodos consecutivos*”. Todos os dados tratados nesta pesquisa enquadram-se perfeitamente nestas definições e disso resulta a importância da análise de séries temporais para o presente trabalho.

Para este estudo, será necessário tratar das componentes de séries temporais, especialmente da sazonal. Também deverá ser apresentada a operação de diferenciação, assim como os conceitos de estacionariedade e co-integração e seus respectivos testes.

##### 3.1.1. Componentes das séries temporais

As séries temporais podem ser caracterizadas por componentes de tendência, cíclica, residual (ou irregular) e sazonal. A componente de tendência está relacionada com movimentos ascendentes ou descendentes continuados. A cíclica é caracterizada por movimentos periódicos. A componente residual é constituída por movimentos ascendentes ou descendentes aleatórios não explicados por variações cíclicas ou pela tendência (DOWNING et al., 2005).

Segundo McClave et al. (2005), a componente sazonal descreve flutuações recorrentes nas séries durante períodos específicos do ano. Downing et al. (2005) comentam que a sazonalidade tem importância para séries mensais, trimestrais e semanais, pois, quando se deseja analisar tendências, é necessário remover as variações sazonais. Como as séries temporais deste estudo são mensais e algumas delas incluem décimo-terceiro salário, a identificação e tratamento de sazonalidade são importantes.

Morettin et al. (2006, p.64) afirmam que *“é difícil definir, tanto do ponto de vista conceitual como estatístico, o que seja sazonalidade”* e, empiricamente, consideram como sazonais os fenômenos que ocorrem regularmente de ano para ano. Afirmam também que as séries sazonais são caracterizadas por apresentarem correlação alta em “lags sazonais”, isto é, intervalos múltiplos de doze meses no caso das séries de valores mensais. Esta foi a orientação seguida neste estudo. Portanto, a sazonalidade será identificada por meio de gráficos de autocorrelação parcial das séries onde os picos significantes em intervalos regulares durante o ano confirmam a presença da componente sazonal. Quando detectada esta componente, o método das médias móveis, descrito em Downing et al. (2005, p.309) e implementado no software SPSS versão 13.0, será usado para ajustamento sazonal da série.

### **3.1.2. Estacionariedade**

Segundo Seiler (2004), estacionariedade significa estabilidade das séries temporais ao longo do tempo. Formalmente, estacionariedade significa que médias e autocovariâncias não dependem do tempo. Gujarati (2000, p. 719) também apresenta formalização desse conceito afirmando que a série é estacionária se *“média e variância forem constantes ao longo do tempo e o valor da covariância entre dois períodos depender apenas da distância ou defasagem entre os dois períodos”*.

Morettin (2006, p. 4) complementa:

*“Uma das suposições mais frequentes que se faz a respeito de uma série temporal é a de que ela é estacionária, ou seja, ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de equilíbrio”*.

Existem testes para estacionariedade cujo detalhamento teórico foge ao escopo deste trabalho. O mais simples deles, baseado na função de autocorrelação, requer o cálculo da estatística de Ljung-Box. Este método está disponível no software SPSS versão 13.0 por meio da função Graph/Time Series/Autocorrelations.

Outro procedimento muito adotado é o teste aumentado de Dickey-Fuller (ADF). Gujarati (2000) explicita que ele pode ser aplicado a processos com termo de erro autocorrelacionado. O conceito de autocorrelação será apresentado adiante, mas pode-se adiantar que os dados desta pesquisa apresentam esta característica e, por isso, o teste ADF revelou-se mais adequado a este estudo. Detalhes sobre o teste ADF e sua implementação no software EViews 5.0 são apresentados por Seiler (2004, p. 279).

### 3.1.3. Diferenciação

Segundo Morettin (2006), a maioria dos procedimentos de análise estatística de séries temporais supõe que estas sejam estacionárias. Quando elas não apresentam esta característica, às vezes é preciso transformá-las para obter a estacionariedade.

A transformação mais comum é a operação de diferenciação que consiste em tomar diferenças de valores sucessivos da série original. A série constituída pelas primeiras diferenças de uma série  $Z_t$ , indicada por  $\Delta Z_t$ , é definida como:

$$\Delta Z_t = Z_t - Z_{t-1}$$

Conclui-se, portanto, que a série diferenciada ( $\Delta Z_t$ ) possui um elemento a menos que a série original ( $Z_t$ ). Esta é a razão pela qual, em alguns *outputs* de softwares estatísticos, o número de observações é reduzido quando se usam explicita ou implicitamente processos de diferenciação.

Se uma série temporal for diferenciada uma vez e a série resultante for estacionária, diz-se que a série original é integrada de ordem 1. Analogamente, se forem necessárias duas diferenciações seguidas para atingir a estacionariedade, isto é, se for preciso tomar a primeira diferença da série diferenciada, diz-se que a série original é integrada de ordem 2, e assim sucessivamente (GUJARATI, 2000).

### **3.2. Análise de regressão**

A análise de regressão linear é uma técnica estatística que pode ser usada para analisar a relação entre uma variável, dita dependente, e uma ou mais variáveis, ditas independentes ou preditoras. O objetivo da análise de regressão é prever as mudanças na variável dependente como resposta a mudanças nas variáveis independentes por meio da regra estatística dos mínimos quadrados. A regressão múltipla também pode ser usada para comparar dois ou mais conjuntos de variáveis para avaliar seu poder preditivo, comparando assim dois ou mais modelos alternativos ou concorrentes. Esta técnica pode ser aplicada em duas classes de problema de pesquisa: previsão e explicação (HAIR et al., 2005).

Conclui-se, portanto, que a análise de regressão linear é adequada para este estudo uma vez que ele tem por objetivo explicar como cada variável previdenciária (receita, pagamento de benefícios e saldo) responde a mudanças em variáveis independentes (econômicas, demográficas etc).

Não faz parte do escopo deste trabalho a apresentação detalhada do método de análise de regressão. Para detalhes sobre os fundamentos teóricos e componentes do modelo (variável dependente, variáveis independentes, variável estatística), indicadores (coeficientes de determinação, significância do modelo e dos coeficientes), pressupostos estatísticos envolvidos no método, verificação de premissas, análise de resíduos etc, as referências sugeridas para consulta são: McClave et al. (2005), Hair et al.(2005) e Downing et al. (2005).

#### **3.2.1. O processo de análise de regressão**

Neste trabalho, foi usado o processo recomendado por Hair et al. (2005) para construir, estimar, interpretar e validar as análises de regressão.

O processo inicia com a determinação do objetivo e seleção das variáveis. A segunda etapa trata da verificação do tamanho da amostra para garantir poder estatístico e capacidade de generalização das conclusões. O requisito mais crítico

é encontrado para procedimentos *stepwise* em que são recomendadas pelo menos 50 observações por variável independente.

A terceira etapa do processo consiste em verificar se as variáveis individuais atendem às suposições de normalidade, linearidade, e variância constante. Caso alguma dessas premissas seja violada, sugere-se a criação de novas variáveis por meio de transformação, isto é, de aplicação de funções matemáticas aos valores da série original para criar outra série com características mais adequadas. Assim, a aplicação de funções logarítmicas ou inversas, entre outras, podem ser usadas com o objetivo de atender às suposições do método.

A quarta etapa recomendada consiste na seleção de uma técnica de estimação. O pesquisador pode especificar um modelo ou utilizar um procedimento de regressão que selecione as variáveis independentes como os métodos de busca seqüencial (*forward/backward* e *stepwise*) e abordagem combinatória, onde são montados todos os subconjuntos possíveis das variáveis. Hair et al. (2005, p. 159) observam que:

“Se um método confirmatório, de busca seqüencial ou combinatório, for escolhido, o critério mais importante será o bom conhecimento do pesquisador sobre o contexto da pesquisa, que permita uma perspectiva objetiva e fundamentada quanto às variáveis a serem incluídas e aos sinais e magnitudes esperados de seus coeficientes. Sem esse conhecimento, os resultados da regressão podem ter elevada precisão preditiva sem qualquer relevância gerencial ou teórica. O pesquisador jamais deve ser totalmente orientado por qualquer uma dessas abordagens, mas, em vez disso, deve usá-las após considerar todas cuidadosamente e então aceitar os resultados somente após cuidadosa investigação.

Ainda nesta quarta etapa, para cada modelo investigado, o processo recomenda testar se a variável estatística, isto é, a combinação linear das variáveis independentes, satisfaz as suposições de regressão (linearidade, variância constante, independência e normalidade). Esta investigação é conduzida pela análise dos resíduos. Também, para cada modelo, devem ser analisadas as significâncias dos coeficientes (estatística *t*) e do modelo como um todo (estatística *F*). Esta etapa é concluída com o exame do coeficiente de determinação ajustado apresentado nos principais softwares estatísticos como “*Adjusted R Square*”.

A quinta etapa do processo recomendado consiste em interpretar a variável estatística analisando os coeficientes estimados pela regressão em termos da sua

explicação da variável dependente. Também deve ser analisado o impacto potencial das variáveis omitidas em razão de multicolinearidade “*para garantir que a significância gerencial seja avaliada juntamente com a significância estatística*” (HAIR et al., 2005, p.164). A última etapa consiste na validação dos resultados.

Duas características relacionadas com o método de análise de regressão são especialmente importantes para este trabalho: a multicolinearidade e a autocorrelação. Os próximos itens tratam desses conceitos.

### **3.2.2. Multicolinearidade**

Segundo Hair et al.(2005, p.142):

“A colinearidade é a associação, medida como a correlação, entre duas variáveis independentes. A multicolinearidade refere-se à correlação entre três ou mais variáveis independentes (evidenciada quando uma é regressada em relação às outras). Apesar de haver uma distinção precisa em termos estatísticos, é prática comum usar os termos alternadamente”.

Quando o método de análise de regressão é usado para explicação, como no presente trabalho, é necessário um cuidado especial com os efeitos da multicolinearidade porque os pesos relativos dos coeficientes acabam não representando bem a intensidade com que uma variável independente é capaz de explicar a variável dependente. Sobre tais efeitos, Hair et al. (2005, p.165) alertam:

“Os efeitos na explicação se referem principalmente à habilidade do procedimento de regressão e do pesquisador em representar e compreender os efeitos de cada variável independente na variável estatística de regressão. Quando ocorre multicolinearidade (mesmo em níveis relativamente baixos de 0,30 ou próximo disso), o processo de separação de efeitos individualmente se torna mais difícil. Primeiro, ela limita o tamanho do coeficiente de determinação e fica cada vez mais difícil acrescentar uma única previsão explanatória a partir de variáveis adicionais. Segundo, mas igualmente importante, ela torna a contribuição de cada variável independente mais complicada porque os efeitos das variáveis independentes estão ‘misturados’ ou confusos. A multicolinearidade resulta em partes maiores de variância compartilhada e níveis mais baixos de variância única da qual os efeitos das variáveis independentes individuais podem ser determinados”.

Hair et al. (2005) sugerem alguns cuidados e ações corretivas no tratamento de multicolinearidade. Podem ser omitidas uma ou mais variáveis independentes altamente correlacionadas e procuradas outras para ajudar na previsão. Usar as correlações simples entre cada variável independente e a variável dependente para compreender a relação entre elas também pode ser uma alternativa.

Existem testes para detectar a multicolinearidade e um processo de análise de regressão em dois estágios para contornar o problema. Tais procedimentos serão apresentados a seguir.

### 3.2.2.1.

#### **Detectando a multicolinearidade**

A multicolinearidade pode ser identificada pela matriz de correlação entre as variáveis independentes. Mas este não é um teste definitivo porque identifica apenas relações entre os pares de variáveis. Há multicolinearidade quando uma variável independente é explicada por outras duas ou mais variáveis independentes juntas e isso pode não ser captado pela análise da matriz de correlação (SEILER, 2004).

Segundo Hair et al. (2005), uma medida frequentemente usada para a multicolinearidade é a tolerância, definida como  $1 - R_i^2$ , onde  $R_i^2$  é o coeficiente de determinação para previsão de uma variável independente “i” pelas outras variáveis predictoras. Os valores de tolerância próximos de zero indicam que a variável é altamente prevista pelas outras variáveis independentes e, por isso, há multicolinearidade. Outra medida comumente usada é o VIF (Variance Inflation Factor) que é o inverso da tolerância. Valores altos de VIF indicam alto grau de multicolinearidade.

Outro método para detecção de multicolinearidade recomendado por Hair et al.(2005) e por Seiler (2004) envolve dois componentes fornecidos pelos softwares estatísticos. O primeiro é o índice de condição que representa a colinearidade de combinações das variâncias no conjunto de dados. O segundo é a matriz de decomposição da variância dos coeficientes de regressão. Em primeiro lugar, devem-se identificar todos os índices de condição com valor acima de 30. Se, para algum dos índices identificados, houver dois ou mais coeficientes com proporção de variância acima de 90%, há multicolinearidade séria nos dados.

### 3.2.2.2. Contornando a multicolinearidade

Para contornar o problema da multicolinearidade, Seiler (2004) recomenda a regressão em dois estágios (*two-stage least squares regression*), uma variante do método convencional, exemplificada a seguir.

Suponha que se pretenda construir uma regressão de Y com as variáveis independentes A, B, C e D. Suponha também que D seja explicada pelas variáveis A,B,C, F e G. Nota-se que ocorre multicolinearidade porque uma variável independente D é explicada por outras variáveis independentes (A,B,C). Neste caso, a variável D é dita endógena e as demais, exógenas. A solução recomendada quando ocorre este tipo de problema consiste em executar a regressão de D com A,B,C,F,G e obter valores previstos para D, que indicaremos por “pD”. Em seguida, executa-se a regressão de Y com A,B,C,”pD”. Assim, ao substituir a variável D por “pD”, os coeficientes estimados nesta última regressão são confiáveis e podem ser interpretados da mesma forma que na regressão convencional. Na terminologia deste tipo regressão, as variáveis independentes utilizadas no primeiro estágio são chamadas de “instrumentos”.

Seiler (2004) também observa que o coeficiente de determinação e os erros não são corretamente reportados quando esses passos são executados separadamente. Mas informa que os softwares SPSS e EViews contemplam a regressão em dois estágios com todos os resultados corretos. Neste estudo, a regressão em dois estágios, quando necessária, será executada utilizando a funcionalidade “*2-Stage Least Squares*”, disponível no menu “*Analyze / Regression*” no SPSS versão 13.0.

### 3.3. Análise de regressão aplicada a séries temporais

Quando se aplica o método de análise de regressão linear usando dados constituídos por séries temporais, há premissas importantes que devem ser observadas. A primeira delas é relacionada com a autocorrelação dos resíduos que, embora não seja exclusiva das séries temporais, é frequentemente encontrada

nesse tipo de dados. Em segundo lugar, devem ser observadas as hipóteses de estacionariedade e de co-integração para que o resultado da regressão seja válido.

### **3.3.1. Autocorrelação dos resíduos**

Um problema comum na aplicação do método de análise de regressão linear em dados constituídos por séries temporais é a violação da premissa de que os resíduos, isto é, as diferenças entre valores previstos na regressão e valores observados, são independentes. A correlação entre resíduos de séries temporais em diferentes pontos no tempo é chamada autocorrelação. A correlação entre dois pontos sucessivos “vizinhos” é chamada de autocorrelação de primeira ordem.

O teste de Durbin-Watson para identificar autocorrelação de primeira ordem é amplamente utilizado e tratado na literatura. De modo muito resumido, pode se dizer que é um teste de hipótese da existência de autocorrelação baseado no número de observações e de variáveis usadas no modelo de regressão. A aplicação do teste é muito facilitada porque o valor da estatística de Durbin-Watson é apresentado com os resultados das análises de regressão pelos principais softwares estatísticos. Mas algumas condições devem ser observadas para que o teste seja válido. O modelo deve incluir um termo de intercepto, ser compatível com esquema auto-regressivo de primeira ordem, não incluir valores defasados da variável dependente como variáveis explicativas e não apresentar observações ausentes (GUJARATI, 2000). Para maiores detalhes sobre este teste e a interpretação dos seus resultados, consultar McClave et al. (2005, p.1044).

Quando a etapa de verificação de premissas do processo de análise de regressão indica a presença de autocorrelação dos resíduos, devem ser usados processos auto-regressivos como contorno do problema. Tais processos são apresentados a seguir.

### 3.3.1.1. Processos auto-regressivos

Os processos auto-regressivos têm o objetivo de solucionar o problema da violação de premissa de independência dos resíduos. Nesta pesquisa, o uso de processos auto-regressivos de primeira ordem revelou-se suficiente.

A equação básica usada para representar um esquema auto-regressivo de primeira ordem é  $Y_t = \beta_0 + \beta_1 X_t + u_t$  onde  $u_t = \rho u_{t-1} + \varepsilon_t$ . Se  $\rho$  é zero, então não há autocorrelação dos resíduos. Caso contrário,  $\rho$  representa a correlação dos termos de erro ao longo do tempo. Os processos auto-regressivos têm por objetivo obter uma estimativa de  $\rho$ .

Neste estudo, foi usado o processo iterativo de Cochrane-Orcutt associado à transformação de Prais-Winsten apresentado em Gujarati (2000) e resumido a seguir. Suponha uma regressão linear com variável dependente  $Y_t$  e variável independente  $X_t$  que resulte em erros autocorrelacionados. Deve-se, inicialmente, obter a série de resíduos  $\varepsilon_t$ . Depois disso, deve-se executar uma regressão com variável dependente  $\varepsilon_t$  e variável independente  $\varepsilon_{t-1}$  para se obter uma primeira aproximação do valor de  $\rho$ . Depois disso, transformam-se as séries originais fazendo  $Y'_t = Y_t - \rho Y_{t-1}$  e  $X'_t = X_t - \rho X_{t-1}$ . Para os termos iniciais das séries, usa-se  $Y'_1 = Y_1 * \text{sqrt}(1 - \rho^2)$  e  $X'_1 = X_1 * \text{sqrt}(1 - \rho^2)$ , onde “sqrt” é a função raiz quadrada. A partir das séries transformadas, repete-se todo o processo e calcula-se novo  $\rho$ . Estas iterações são executadas até se obter uma convergência satisfatória de  $\rho$ .

O software SPSS implementa este processo na função *Autoregression* de análise de séries temporais por meio da opção Prais-Winsten. A adequação desta função às necessidades desta pesquisa é revelada pela ajuda on-line do SPSS:

“The Autoregression procedure is an extension of ordinary least-squares regression analysis specifically designed for time series. [...] The Autoregression procedure accounts for first-order autocorrelated residuals and provides reliable estimates of both goodness-of-fit measures and significance levels of chosen predictor variables”.

### 3.3.2. Séries temporais co-integradas

Outro problema comum na análise de regressão aplicada a séries temporais é apresentado a seguir, extraído de Gujarati (2000, p. 715):

“Ao regredir uma variável de série temporal sobre outra variável de série temporal, muitas vezes obtemos um valor de  $R^2$  bem alto, embora não haja uma relação significativa entre as duas. [...] Este problema ocorre porque, se ambas as séries temporais envolvidas exibirem fortes tendências (movimentos ascendentes ou descendentes continuados), o alto  $R^2$  observado se deve à presença da tendência, e não à verdadeira relação entre as duas”.

Estas regressões que exibem resultados duvidosos são conhecidas como regressões espúrias e ocorrem quando se regredem algumas séries temporais não-estacionárias sobre outras também não estacionárias. Nestes casos, os testes “t” e “F” usuais nem sempre são válidos. Uma alternativa seria diferenciar as séries até obter-se a estacionariedade e usar as séries estacionárias nos processos de regressão. Mas estas diferenciações geralmente levam à perda de valiosas relações de longo prazo existentes entre as séries.

Como, na prática, a maioria das séries econômicas é não-estacionária e grande parte da teoria econômica envolve relações de longo prazo entre estas variáveis, muitos estudos no campo da econometria das séries temporais foram realizados. Estes estudos concluíram que uma combinação linear de variáveis não estacionárias pode ser estacionária e que, neste caso, a metodologia tradicional de regressão é válida. As séries temporais que apresentam esta característica de estacionariedade quando combinadas são ditas co-integradas. Dois requisitos devem ser satisfeitos para que séries sejam co-integradas: os resíduos da regressão devem ser estacionários e as séries devem ser integradas de mesma ordem, isto é, devem requerer quantidade idêntica de diferenciações para atingir a estacionariedade (GUJARATI, 2000).

Assim, pelo que foi exposto, além de verificar as premissas tradicionais do método de análise de regressão, é necessário verificar se as séries envolvidas nos modelos deste estudo são estacionárias. Se não forem, deve-se verificar se são co-integradas, isto é, se o número de diferenciações para obter estacionariedade é igual para todas as séries de um determinado modelo e se os resíduos da regressão são estacionários. Este trabalho é muito facilitado pelo software EViews 5.0 que

apresenta duas formas de avaliação da premissa de co-integração das séries envolvidas em regressões.

A primeira delas é o teste de co-integração (*Johansen Cointegration Test*) para determinar se duas ou mais séries não estacionárias podem ser combinadas para criar séries estacionárias. Este teste, que inclui verificações de estacionariedade e ordem de integração para cada série, é descrito em detalhes por Seiler (2004, p. 296-310).

A segunda forma, muito mais simples, mas restrita a processos auto-regressivos, contempla o indicador denominado “*Inverted AR roots*”. Quando este indicador apresenta valor menor que um em módulo, o processo é estacionário e a regressão é válida. No caso de regressão espúria, além de apresentar o indicador, o software emite a mensagem “*Estimated AR process is nonstationary*”. O texto a seguir foi extraído da ajuda on-line do programa EViews 5.0 e demonstra sua utilidade para as verificações necessárias a este estudo.

“The stationarity condition for general AR(p) processes is that the inverted roots of the lag polynomial lie inside the unit circle. EViews reports these roots as Inverted AR Roots at the bottom of the regression output. There is no particular problem if the roots are imaginary, but a stationary AR model should have all roots with modulus less than one.”

Como todas as regressões montadas nesta pesquisa incluem um termo auto-regressivo, esta é a opção preferencial para verificar a validade dos modelos.