

4

Base de dados, métricas estatísticas e metodologia

4.1.

Base de dados

Foram coletados dados do boletim estatístico do Ministério da Previdência Social de forma temporal para os meses de dezembro de 2002 (ano do início da regulamentação pela Secretária de Previdência Complementar) até dezembro de 2008 trimestralmente. A análise temporal ocorrerá, basicamente, através dos resultados disponíveis nos quadros 7, 8 e 9 do boletim estatístico e de uma regressão dinâmica.

Com base nestes resultados será possível observar o comportamento dos investimentos dos 10 maiores fundos de pensão no Brasil, além de verificar os seus comportamentos condicionados a variáveis macroeconômicas. Para a análise de regressão, serão utilizadas também informações obtidas no Instituto de Pesquisas Econômicas Aplicadas (fonte secundária), Banco Central do Brasil e Instituto Brasileiro de Geografia e Estatística¹ referentes à:

- 1) **Dólar comercial:** Parâmetro para operações oficiais de compra e de venda de moeda no comércio exterior, geradas pelos seguintes tipos de negócio: exportação, importação, emissão de passagens aéreas e marítimas, bônus, comercial paper, pagamento do serviço da dívida externa e remessa de dividendos das empresas com sede no exterior do País. A cotação é expressa em real (moeda brasileira) por dólar americano (moeda americana). (ADFN, 2010)
- 2) **Retorno Fundo de Ações:** Fundo de investimento cuja carteira de títulos e valores mobiliários mantém aplicações de, no mínimo, 80% em ativos relacionados diretamente, ou sintetizados via derivativos, no fator de risco que dá nome à classe, ou seja, variação de preços de ações admitidas à negociação no mercado a vista de bolsa de valores ou

entidade do mercado de balcão organizado. O fundo deve possuir, no mínimo, 67% da carteira em ações admitidas à negociação no mercado a vista de bolsa de valores ou entidade do mercado de balcão organizado. (ADFN, 2010)

- 3) **Taxa de Juros:** O juro pode ser compreendido como uma espécie de "aluguel sobre o dinheiro". A taxa seria uma compensação paga pelo tomador do empréstimo para ter o direito de usar o dinheiro até o dia do pagamento. O credor, por outro lado, recebe uma compensação por não poder usar esse dinheiro até o dia do pagamento e por correr o risco de não receber o dinheiro de volta (risco de inadimplência). (ADFN, 2010)
- 4) **Índice de Ações (Ibovespa):** Ação é o título representativo da menor parcela do capital social de uma sociedade por ações (sociedade anônima) e podem ou não ser negociadas em bolsa de valores. As ações são classificadas de acordo com a natureza dos direitos e das vantagens que dão aos seus detentores. (ADFN, 2010)
- 5) **PIB (preços de mercado):** O PIB constitui um indicador da atividade econômica de um determinado país na medida em que representa o valor total da produção de bens e serviços. (ADFN, 2010)

Vale observar que em todas as séries, tal como a variável independente compreendem o período de dezembro de 2002 à dezembro 2008 sendo observadas trimestralmente. A fonte primária do dólar comercial, fundo de ações, poupança e taxa de juros foram são oriundas do boletim do mercado financeiro do Banco Central do Brasil (BCB Boletim/M. Finan). Já o índice de ações Ibovespa é fornecido pela ANDIMA enquanto que o PIB – preços de mercado é calculado pelo IBGE (IBGE/SCN 2000 Trimestral).

É importante observar que foram testadas outras variáveis explicativas ao modelo, tais como, rendimento do dólar, rendimento de fundo de ação, taxa de juros CDB, no entanto, todas não se apresentaram significativas para explicar os modelos, dessa forma, não serão retratadas na análise a seguir. Em outras palavras, levar-se-á em consideração as variáveis destacadas de 1 a 5 com o objetivo de mostrar alguma relação causal entre estes indicadores

¹ Ambas citadas na parte 6 deste trabalho (Referências).

macroeconômicos com o comportamento dos investimentos dos 10 maiores fundos de pensão no Brasil.

As variáveis dependentes serão os investimentos realizados tanto pelos fundos públicos quanto pelos fundos privados.

4.2. Métricas estatísticas²

Dando seqüência, são descritas estatísticas para a avaliação do desempenho do modelo de previsão. Isto é, a partir da comparação dos valores reais e dos valores “ajustados” pelo modelo, podem ser calculadas várias métricas para medir o desempenho. Estas medidas servem, então, para avaliar o desempenho do modelo estimado dentro (*in sample*) e fora (*out of sample*) da amostra de dados utilizados na modelagem.

4.2.1. Coeficiente de explicação (R^2)

O coeficiente de explicação R^2 indica o quanto da variação total dos dados (série dependente) é explicada pelo modelo. Este coeficiente é calculado através da comparação do erro do modelo e a variação dos dados da série dependente (série a ser prevista) em torno de sua média. Importante salientar que o coeficiente varia entre 0 e 100% e quanto maior seu valor significa que o modelo está sendo melhor explicado. Matematicamente, tem-se:

$$R^2 = \left(1 - \frac{\sum_{t=1}^N (Y(t) - \hat{Y}(t))^2}{\sum_{t=1}^N (Y(t) - \bar{Y})^2} \right) \times 100$$

(1)

onde $Y(t)$ é o valor da série temporal no período (t); $\hat{Y}(t)$ é o valor ajustado da série temporal para o período (t); \bar{Y} é a média de Y e N é o total de observações.

² Embora nem todas as métricas são citadas no texto, ambas são amplamente utilizadas na análise para a escolha do melhor modelo.

4.2.2. Mean Absolute Percent Error (MAPE)

O MAPE (erro médio absoluto percentual) é calculado através da diferença entre valores estimados e reais e equivale às previsões um passo-à-frente (por exemplo, para o ano seguinte). Matematicamente, tem-se:

$$MAPE = \frac{\sum_{t=1}^N \frac{|Y(t) - \hat{Y}(t)|}{Y(t)} \times 100}{N}$$

(2)

Vale lembrar que o MAPE dentro da amostra (*in-sample*) é uma medida de ajuste, já, fora da amostra (*out-of-sample*) é uma medida real de ajustamento da amostra. No entanto, em ambos os casos, quanto menor for o MAPE, melhor será a adequação do modelo, e conseqüentemente, a previsão.

4.3. Metodologia

A análise de regressão está entre as técnicas estatísticas mais utilizadas em aplicações empíricas, pois permite que o analista observe a relação existente entre a variável de interesse (dependente) e as variáveis denominadas explicativas (independentes). Esta técnica permite que se observe a elasticidade de uma variável sobre a outra, isto é, é possível observar o efeito direto que uma variação marginal na variável independente tem sobre a variável dependente.

Variável dependente é aquela que se quer explicar enquanto que variáveis independentes são as variáveis utilizadas para explicar a variável dependente. Outra possível nomenclatura possível seria chamar a variável dependente de variável explicada e as variáveis independentes de variáveis explicativas. Para ilustrar esta explicação, neste trabalho, as variáveis dependentes são os investimentos e as variáveis independentes ou explicativas são o dólar comercial, a poupança, a taxa de juros, o Ibovespa e o PIB³.

³É importante destacar que ainda foram testados modelos *log-log* e modelos com percentuais (isto é, regrediu-se a variação percentual entre os períodos), no entanto, ambas as métricas estatísticas destes modelos apresentaram resultados não satisfatórios.

Diversos são os modelos matemáticos que podem ser especificados para caracterizar tal relação. A partir de cada tipo de modelo empregado define-se um específico método de regressão tal como a regressão linear, regressão logística, regressão dinâmica entre outros. A escolha do método de regressão deve levar em conta peculiaridades dos dados tais como o tipo da variável dependente (contínua, binária, etc.), a existência de relação ou não entre as observações etc.

Como as informações sobre os fundos de investimento utilizados neste trabalho são amostras dependentes serialmente, ou seja, são séries temporais (observações equidistantes no tempo e que apresentam correlação), a técnica de regressão mais apropriada é a regressão dinâmica.

4.3.1. Modelo de regressão dinâmica

Os modelos de regressão dinâmica surgiram com o intuito de suprir algumas deficiências encontradas ao se aplicar um modelo de regressão clássico a observações com dependência serial. Esta dependência serial faz com que os pressupostos para a utilização de uma regressão linear não sejam válidos. Um modelo de regressão clássico tem por hipótese que os resíduos do modelo têm: média zero, variância constante, distribuição normal e independência. Basicamente, a hipótese de independência dos erros é muito forte e praticamente impossível de ser alcançada quanto se trabalha com séries temporais.

Segundo Barros e Souza (1995), a falta de independência nos resíduos acarreta que, (a) os estimadores de mínimos quadrados são tendenciosos e com isso não possuem mais a variância mínima; (b) os estimadores da variância e dos erros padrões dos coeficientes da regressão são subestimados, o que levaria à conclusão de que os estimadores são mais precisos do que na realidade. Isto implica que os intervalos de confiança e os testes de hipótese perdem a sua validade

Dessa forma, surgem os modelos de regressão dinâmica que tem como principal característica unir o poder das variáveis explicativas ao dinamismo das séries temporais. De acordo com Zanini (2000), os modelos de regressão dinâmica combinam a dinâmica de séries temporais e o efeito de variáveis explicativas. Atenta-se que o termo “regressão dinâmica” não indica que os parâmetros do modelo evoluem no tempo. Ao contrário, a palavra “dinâmica” significa aqui um

modelo de regressão no qual incluímos a estrutura de dependência de uma série temporal.

Souza (2005) também salienta a origem do nome dinâmico. Para o autor, os modelos dinâmicos não consideram apenas os valores correntes das variáveis, mas também consideram os seus valores defasados. Esta peculiaridade, onde é possível descrever as relações entre as variáveis consideradas ao longo do tempo, é que lhes dão o adjetivo de dinâmicos. Modelos de regressão dinâmica devem ser usados quando existe uma estrutura de dependência entre a variável de interesse e variáveis causais e, ao mesmo tempo, quando a estrutura de correlação da série dependente indicar que não podemos supor a independência dos erros.

Barbetta (2004) observa que em modelos dinâmicos, procura-se construir um modelo estatístico-matemático para analisar a relação entre as variáveis independentes e a variável dependente e a partir do conhecimento adquirido sobre estas, pode-se prever o valor futuro da variável dependente com base nas variáveis independentes.

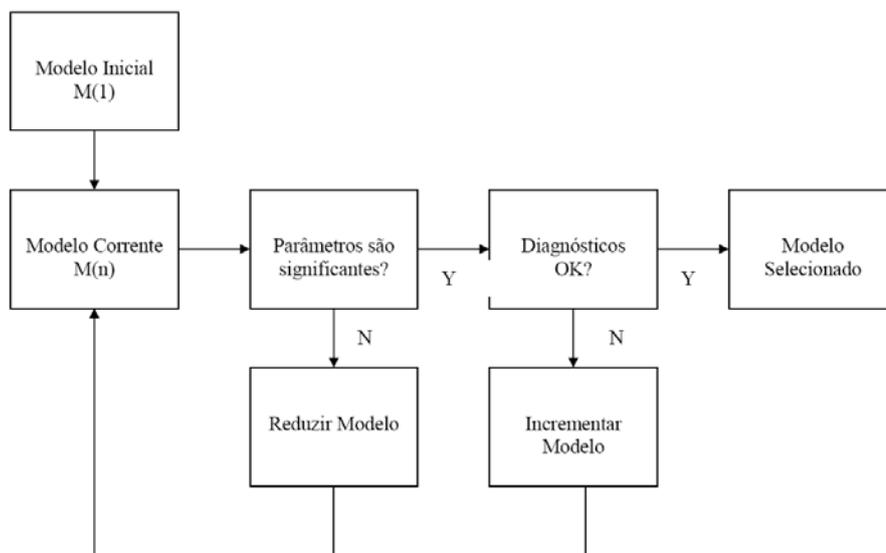
É importante observar que a diferença básica entre os modelos de Regressão Dinâmica e os modelos ARIMA consiste no fato dos modelos de regressão dinâmica incluírem efeitos de variáveis causais através do termo βx_t . Os modelos ARIMA, por sua vez, não incluem tais efeitos, e apenas o passado da série Y_t e os valores defasados da série de erros são usados na modelagem e previsão da série Y_t . (ZANINI, 2000). Procura-se, portanto, dentro de metodologia definida, evoluir na modelagem, buscando em outras causalidades, que não somente a própria série de demanda, um melhor modelo para prever esta demanda.

Quanto ao modelo de regressão dinâmica é importante observar sua diferença para os modelos econométricos. Isto porque, no modelo de regressão dinâmica a estratégia é construir modelos a partir dos dados, isto é, utiliza-se o procedimento *bottom-up*, ou seja, a partir de um modelo simples, refina-se as variáveis explicativas, incluindo e retirando variáveis até encontrar um modelo apropriado, sempre adotando o princípio da parcimônia. Já em um modelo econométrico, é previamente conhecido a estrutura do modelo sendo o único problema a estimação dos parâmetros.

Segundo Souza (2005), outra abordagem seria possível. Esta abordagem consiste em se partir de um modelo mais geral, com diversos regressores, e com o decorrer da análise ir se retirando as variáveis que não são significativas, ou seja, através de análises de significância vai se reduzindo o número de variáveis do modelo. Esta abordagem é academicamente conhecida como abordagem do geral para o específico ou então como metodologia de Hendry e foi desenvolvida na *London School of Economics* na década de 60.

O fluxograma a seguir indica, de maneira genérica, os passos usados na construção de um modelo de regressão dinâmica com o procedimento *bottom-up* que será empregado neste trabalho

Figura 1 – Construção de um modelo de regressão dinâmica



Fonte: ZANINI (2000)

Cabe salientar que neste tipo de modelagem, a inclusão de uma variável ou não em um modelo está sujeito a critérios estatísticos, mais precisamente a significância dos parâmetros analisados através de testes de hipótese, como também a um resultado lógico e plausível com a realidade, isto é, deve-se levar em consideração o sinal de causalidade entre a variável dependente e a independente.

Segundo Zanini (2000) o modelo de regressão pode ser representado por:

$$\varphi(B)Y_t = \beta x_t + \varepsilon_t$$

onde:

Y_t = variável dependente no instante t

β = vetor de coeficientes das variáveis causais, que é estimado por

mínimos quadrados

B = é o operador de defasagem

x_t = vetor de variáveis causais no instante t

ε_t = ruído aleatório associado ao modelo

$\varphi(B)$ = polinômio autoregressivo de ordem p ,

O polinômio $\varphi(B)$ é o que permite incluir variáveis defasadas ao modelo.

Quando $\varphi(B) = 1$, o modelo não apresenta estrutura de defasagem, enquanto que quando $\varphi(B) \neq 1$, o modelo pode ser utilizado para expressar as mais complexas relações possíveis. Por exemplo, se o modelo tiver uma estrutura de defasagem na variável explicativa de ordem 1, a equação do modelo ficará como,

$$Y_t = \beta x_t + \phi_1 Y_{t-1} + \varepsilon_t$$

Alguns testes são utilizados durante a construção do modelo dentre os quais pode-se citar o teste para verificar a dinâmica do modelo, o teste para verificar a causalidade do modelo e o teste para verificar o ajuste do modelo. O teste para verificar a dinâmica do modelo consiste em um teste de hipótese que tem como objetivo identificar se a inclusão de “lags” tanto da variável dependente quanto da independente é necessária para o modelo.

Segundo Zanini (2000), o teste de causalidade consiste em verificar se a inclusão de uma ou mais variáveis ainda não contempladas no modelo resulta numa melhora do ajuste. Atenta-se para o fato de que os testes para a especificação de variáveis causais não se referem à parte dinâmica do modelo, e não tratam da inclusão de “lags” da variável dependente e de erros estruturados

Para testar a adequação, é necessário realizar uma análise dos erros obtidos pelo modelo. Deve-se atentar para que não exista nenhuma estrutura de correlação entre os erros, porém se essa correlação serial for encontrada, esta deve ser atribuída segundo Barros e Souza (1995) ao fato de existir alguma estrutura da variável dependente que não ainda não foi capturada pelo atual modelo. Segundo os autores isto pode ser solucionado incluindo mais “lags” da variável dependente ou independentes contidas no modelo ou então incluir novas variáveis causais. Neste momento também deve-se observar se a correlação existente é factível ou não. O que se quer dizer é que, a primeira vista, não faz sentido algum uma observação hoje ter uma correlação com uma observação muito distante. Dentre os testes de correlação pode-se citar o de Durbin-Watson e o Ljung-Box. O teste de Durbin-Watson testa só a correlação dos erros de lag 1 enquanto que o de Ljung-Box testa todas as correlações até o lag 18.

Para se avaliar a adequação do modelo, várias métricas sobre o ajuste do modelo e a capacidade de previsão podem ser utilizadas. Neste trabalho se utilizará o R^2 ajustado e o MAPE. R^2 ajustado é uma medida de como as observações são explicadas pelo modelo, ou seja, mostra quanto da variação da variável dependente é explicada pela variável independente. Este valor varia entre 0% e 100% e quanto maior, melhor é. Já o MAPE significa o erro percentual média absoluto, isto é, o quanto o modelo está errando quando se faz a previsão um passo a frente.

Como neste trabalho o objetivo maior não é montar um modelo preditivo para os investimentos dos fundos de pensão, e sim dar maior ênfase aos testes de causalidade, então, buscar-se-á apenas encontrar uma relação entre a variável dependente e as variáveis independentes (macroeconômicas) deixando-se dessa forma de lado a busca por estruturas defasadas. A aplicação desta ferramenta estatística tem como principal objetivo encontrar diferença entre as características dos fundos privados e os fundos públicos, ou seja, responder a perguntas como: o portfólio dos dois tende a ficar similar? Os dois reagem de forma parecida às variáveis macroeconômicas? Existe algum fato que influencia de forma diferente esses fundos?

No próximo capítulo, serão apresentados os modelos estimados e os resultados encontrados. O software utilizado para implementar tal modelo foi o FPW (Forecast Pro for Windows) e as saídas deste programa estão em anexo.