

6

Tratamento e análise dos dados

Este capítulo apresenta as séries temporais das variáveis previdenciárias e dos indicadores apresentados na revisão de literatura. Relata as operações nos dados conduzidas em conformidade com o processo descrito no referencial teórico para torná-los compatíveis com os objetivos do estudo e com as suposições do método usado. Apresenta também as análises estatísticas e os melhores modelos encontrados.

6.1.

Identificação das variáveis

As variáveis tratadas neste trabalho estão identificadas na tabela 4 a seguir. As três primeiras representam as séries previdenciárias e as demais representam os indicadores selecionados na revisão de literatura sobre previdência.

Sigla	Descrição
RP	Receitas Previdenciárias
BP	Benefícios Previdenciários
SP	Saldos Previdenciários
INPC	Índice Nacional de Preços ao Consumidor
SM	Salário Mínimo
SELIC	Taxa de Juros SELIC
RDDI	Razão de Dependência Demográfica dos Idosos
RS	Razão de Sexo
TD	Taxa de Desemprego
RMNH	Rendimentos Médios Nominais Habituais
TI	Trabalho Informal

Tabela 4 – Identificação das variáveis

6.2.

As séries previdenciárias

As séries previdenciárias, resultantes das demonstrações contábeis do INSS, são receitas previdenciárias (RP), benefícios previdenciários (BP) e saldo previdenciário (SP), este último, resultante da diferença entre as duas primeiras. Elas são constituídas por valores mensais correntes em milhões de reais para o período de janeiro de 1995 a agosto de 2006, totalizando 140 observações.

Todos os valores das séries previdenciárias foram divididos pelos valores correspondentes do Produto Interno Bruto (PIB), cuja série mensal é calculada e divulgada pelo Banco Central do Brasil em milhões de reais.

Assim, constituiu-se uma nova série adimensional de receitas previdenciárias divididas pelo PIB, identificada por “RP/PIB”. Analogamente, foram construídas as séries “BP/PIB” e “SP/PIB”, respectivamente dos benefícios e saldos previdenciários divididos pelo PIB. Esta representou a primeira etapa na preparação dos dados.

6.2.1.

Ajustamento sazonal das séries previdenciárias

As séries resultantes foram investigadas para detectar a existência de componentes sazonais. Nota-se que o décimo-terceiro salário resulta em picos a cada 12 meses, como revelam os gráficos de autocorrelação parcial da figura 6.

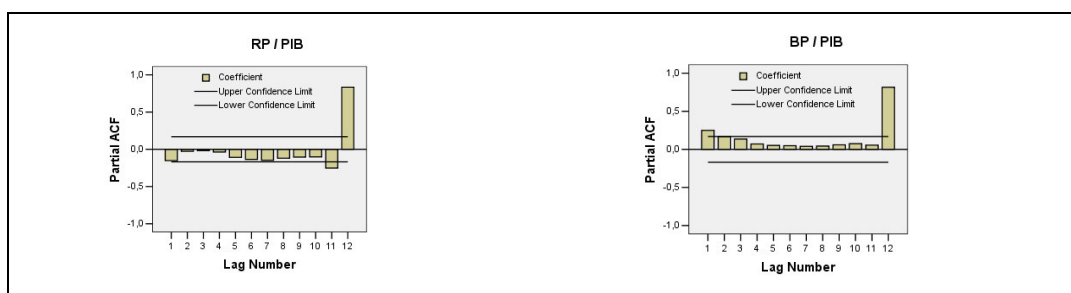


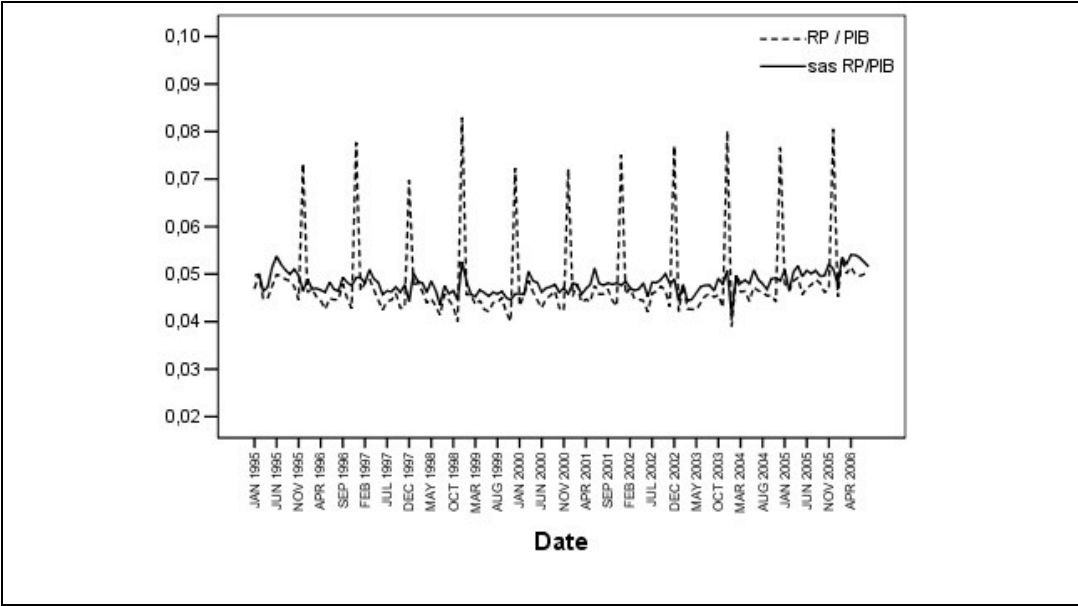
Figura 6 – Autocorrelação parcial das séries previdenciárias

Foi, então, aplicado o ajustamento sazonal com auxílio do SPSS versão 13.0 por meio da função *Analyze / Time Series / Seasonal Decomposition*. A tabela 5 apresenta os fatores sazonais resultantes.

Receitas Previdenciárias			Benefícios Previdenciários		
	Period	Seasonal Factor (%)		Period	Seasonal Factor (%)
	1	94,3		1	94,8
	2	99,9		2	97,3
	3	96,2		3	92,8
	4	95,2		4	92,1
	5	92,3		5	89,2
	6	92,9		6	92,0
	7	95,3		7	95,2
	8	96,2		8	94,7
	9	97,1		9	98,0
	10	92,9		10	92,0
	11	89,8		11	96,3
	12	157,7		12	165,7

Tabela 5 – Fatores de ajustamento sazonal das séries previdenciárias

As séries ajustadas foram identificadas pelo prefixo “sas”. Assim, “sasRP/PIB” é a sigla que representa a série ajustada de receitas previdenciárias em relação ao PIB. Analogamente, a série “sasBP/PIB” representa os benefícios previdenciários após o ajustamento sazonal. As séries originais e ajustadas são apresentadas na figura 7.



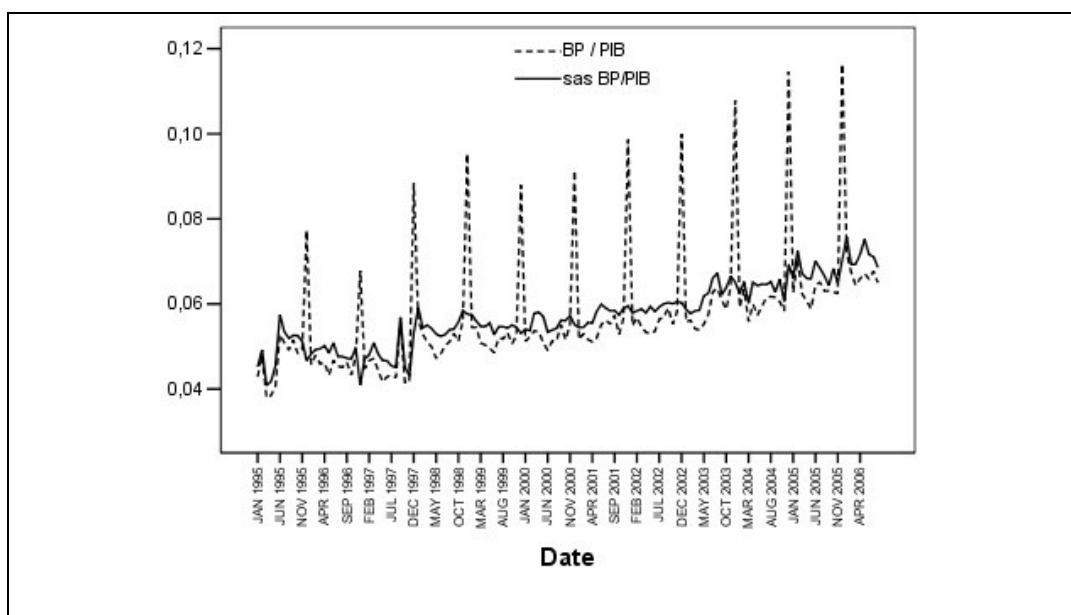


Figura 7 – Séries previdenciárias antes e após o ajuste sazonal

Novos gráficos de autocorrelação parcial foram gerados para as séries ajustadas e são apresentados na figura 8.

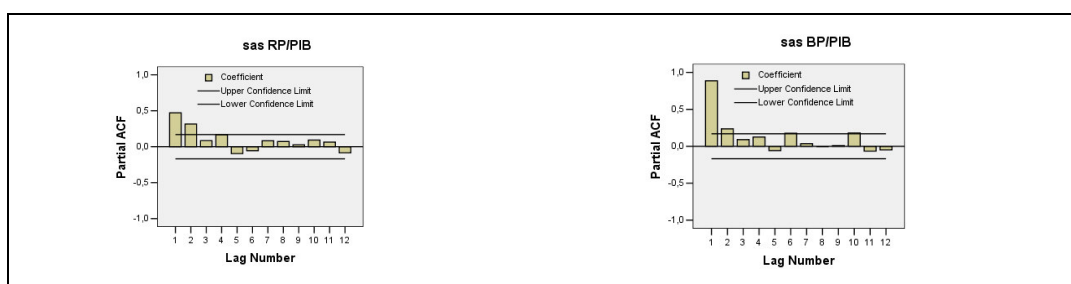


Figura 8 – Autocorrelação parcial após ajustamento sazonal

Esses gráficos, além de comprovar remoção das componentes sazonais, mostram que as séries ajustadas apresentam autocorrelação de primeira ordem e, por isso, indicam o uso de processos auto-regressivos nas análises desse estudo.

6.3. Verificação de premissas

O tratamento dos dados segue o processo de análise de regressão descrito no referencial teórico. A primeira etapa consistiu na determinação do objetivo (explicar variações nas séries previdenciárias) e na seleção das variáveis a partir de referências da literatura.

Para a etapa de verificação do tamanho da amostra, foi usada a regra geral comentada por Hair et al. (2005, p. 148) em que os autores afirmam que “*o nível desejado está entre 15 e 20 observações para cada variável independente. [...] No entanto, se um procedimento stepwise é empregado, o nível recomendado aumenta para 50 por 1*”. O conjunto de dados desta pesquisa é constituído por séries temporais com, no mínimo, 54 observações cada. Atende, portanto, à recomendação quanto à quantidade de itens.

A terceira etapa consistiu em verificar se as variáveis individuais atendiam às suposições do modelo de análise de regressão múltipla. As séries de INPC, SELIC e salário mínimo não atenderam ao requisito desejável de normalidade e, para tentar obtê-la, as transformações sugeridas em Hair et al. (2005) foram aplicadas às séries. Mas a única que produziu efeitos positivos foi a função inversa aplicada à SELIC. Esta série transformada será identificada pela sigla “inv_SELIC”. Para as outras séries, optou-se por usar os valores originais uma vez que o método é robusto para desvios nesta premissa e há um número grande de observações nas séries temporais que contribui para diminuir efeitos nocivos de não-normalidade.

Ainda nesta etapa, verificou-se, por meio de gráficos de dispersão, que a suposição de linearidade da relação entre as variáveis previdenciárias e os fatores demográficos era melhor atendida com a transformação raiz quadrada aplicada às séries RDDI e RS que passaram a ser identificadas por “sqrtRDDI” e “sqrtRS” respectivamente.

Verificou-se também que os dados e modelos usados neste estudo são compatíveis com as premissas listadas no referencial teórico para aplicação do teste de Durbin-Watson.

6.3.1. Exame da matriz de dispersão

O anexo 1, apresenta a matriz de dispersão dos dados. Nela, se podem observar as distribuições univariadas (histogramas), as correlações bivariadas e os diagramas de dispersão entre as variáveis. Trata-se de uma forma compacta e abrangente de apresentar os dados pesquisados e é recomendada por Hair et al. (2005, p.54) como “*um formato particularmente adequado a técnicas multivariadas*”.

A tabela 6, a seguir, reproduz as três primeiras colunas da matriz de dispersão dos dados que contêm as correlações entre as variáveis previdenciárias e as variáveis independentes.

	sas RP/PIB	sas BP/PIB	sas SP/PIB
INPC	,009	-,229(**)	,251(**)
SM	,392(**)	,914(**)	-,859(**)
inv SELIC	,073	,550(**)	-,571(**)
sqrt RDDI	,250(**)	,912(**)	-,903(**)
sqrt RS	-,241(**)	-,910(**)	,904(**)
TD	-,319(**)	,365(**)	-,499(**)
RMNH	,724(**)	,799(**)	-,508(**)
TI	-,226	,031	-,197

** Correlation is significant at the 0.01 level (2-tailed).

Tabela 6 – Correlação das variáveis independentes com as variáveis previdenciárias

Segundo Hair et al. (2005, p.165): “*A situação ideal para um pesquisador seria ter diversas variáveis independentes altamente correlacionadas com a variável dependente, mas com pouca correlação entre elas próprias*”. Nota-se que a maioria das variáveis apresenta alta correlação com as variáveis previdenciárias, fato que caracteriza uma situação próxima à ideal e indício de que a literatura pesquisada indicou corretamente elementos relevantes para estudos sobre Previdência Social. Mas resta verificar a correlação entre as variáveis independentes.

O próximo passo consistiu na análise da matriz de correlações e indicadores de multicolinearidade. A tabela 7 reproduz esta parte da matriz de dispersão. Por ela, nota-se que existem altas correlações entre variáveis independentes.

	INPC	SM	inv SELIC	sqrt RDDI	sqrt RS	TD	RMNH
SM	-,257(**)						
inv SELIC	-,401(**)	,611(**)					
sqrt RDDI	-,246(**)	,960(**)	,701(**)				
sqrt RS	,248(**)	-,956(**)	-,705(**)	-,999(**)			
TD	-,209(*)	,329(**)	,350(**)	,453(**)	-,462(**)		
RMNH	-,410(**)	,932(**)	,398(**)	,953(**)	-,952(**)	-,724(**)	
TI	-0,24	-0,001	0,149	0,081	-0,085	0,112	-0,162

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Tabela 7 – Correlação entre as variáveis independentes

Foram obtidos também os valores de tolerância e índices de condição, apresentados nas tabelas 8 e 9, para uma regressão envolvendo todas as variáveis.

Collinearity Statistics		
	Tolerance	VIF
INPC	0,5047	1,9814
SM	0,0432	23,1465
invSELIC	0,7450	1,3422
sqrtRDDI	0,0002	4119,1212
sqrtRS	0,0002	4088,8582
TD	0,3031	3,2998
RMNH	0,0298	33,5133
TI	0,3370	2,9677

Tabela 8 – Tolerâncias e fatores de inflação de variância

Condition Index	Variance Proportions								
	(Constant)	INPC	SM	invSELIC	sqrtRDDI	sqrtRS	TD	RMNH	TI
1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
3,98	0,00	0,46	0,00	0,00	0,00	0,00	0,00	0,00	0,00
18,04	0,00	0,20	0,02	0,03	0,00	0,00	0,04	0,00	0,00
23,56	0,00	0,01	0,01	0,91	0,00	0,00	0,00	0,00	0,00
51,66	0,00	0,18	0,08	0,04	0,00	0,00	0,39	0,00	0,00
127,09	0,00	0,13	0,18	0,00	0,00	0,00	0,11	0,14	0,15
253,64	0,00	0,01	0,38	0,01	0,00	0,00	0,32	0,46	0,32
2006,11	0,00	0,00	0,32	0,01	0,01	0,00	0,15	0,40	0,48
118274,81	1,00	0,00	0,01	0,00	0,99	1,00	0,00	0,01	0,05

Tabela 9 – Índices de condição e proporções de variância

O exame das tabelas 7, 8 e 9 permite concluir que as séries escolhidas para representar os fatores demográficos (sqrtRDDI e sqrsRS) apresentam altíssima colinearidade, indicada pelo coeficiente de correlação de $-0,999$, pela tolerância de $0,0002$ e pelas proporções de variância de $0,99$ e $1,00$ existentes na última linha da tabela 9. Além disso, não se identificaram variáveis exógenas que pudessem ser usadas como instrumentos numa regressão em dois estágios para substituir uma delas. Por esta razão, apenas a série raiz quadrada da razão de dependência demográfica dos idosos (sqrtRDDI), escolhida porque apresenta maior correlação com as variáveis previdenciárias, foi usada para representar os fatores demográficos. Esta decisão foi tomada com base nas “ações corretivas para a multicolinearidade” recomendadas por Hair et al. (2005, p.169), entre as quais se inclui “omitir uma ou mais variáveis independentes altamente correlacionadas”.

As mesmas tabelas indicaram baixa tolerância das séries de rendimentos (RMNH) e salário-mínimo (SM), mas os índices de condição e proporções de variância não apresentaram os critérios que são considerados críticos no referencial teórico. Tentativas de identificar e comprovar a existência de variáveis endógenas entre os dados não foram bem sucedidas. Assim as regressões em dois estágios não puderam ser usadas para otimizar os modelos multivariados.

6.3.2. Testes de estacionariedade

Finalmente, aplicaram-se os testes ADF para verificação da hipótese de estacionariedade das séries originais e de suas primeiras diferenças. Os resultados encontram-se na tabela 10.

P-values dos testes de estacionariedade das séries (ADF) Hipótese nula: a série é não-estacionária		
	Série original	Primeira diferença
sas RP/PIB	0.0017	0.0000
sas BP/PIB	0.6631	0.0000
sas SP/PIB	0.4500	0.0000
INPC	0.0006	0.0000
SM	0.9927	0.0000
inv SELIC	0.2006	0.0000
sqrt RDDI	0.1525	0.0000
TD	0.1361	0.0348

RMNH	0.9294	0.0000
TI	0.5459	0.0000

Tabela 10 – Testes ADF de estacionariedade das séries (p-values)

Verificou-se que, ao nível de significância de 5%, as séries “sas RP/PIB” e INPC são estacionárias e que as demais são integradas de ordem 1, isto é, suas primeiras diferenças são estacionárias.

6.4.

Abordagem combinatória dos processos auto-regressivos

Terminada a fase de preparação dos dados e avaliação das suposições do método, passou-se à fase de execução das regressões em que as séries previdenciárias assumem o papel de variável dependente e os indicadores tornam-se as variáveis preditoras. Como o conjunto de variáveis independentes ficou constituído por sete séries (“sqrtRDDI”, TD, SM, INPC, TI, RMNH e “invSELIC”), é possível construir 127 subconjuntos resultantes das diferentes possibilidades de combinação. Assim, para cada variável previdenciária, foram testados 127 modelos construídos a partir desta abordagem combinatória que, segundo Hair et al. (2005, p. 159) representa “*um processo de busca generalizado em todas as possíveis combinações de variáveis independentes*”.

Os testes foram executados com auxílio do software SPSS versão 13.0 por meio da função *Autoregression*, disponível no menu *Analyze / Time Series*. Esta função é referida no manual do SPSS como uma extensão da análise de regressão linear comum voltada especificamente para séries temporais. Ela leva em conta a presença de resíduos autocorrelacionados e produz estimativas confiáveis dos coeficientes de determinação e dos níveis de significância das variáveis preditoras.

Entre as opções disponíveis na função *Autoregression*, foi escolhido o método de Prais-Winsten porque não há dados perdidos nas séries e os *outputs* gerados são muito semelhantes aos da regressão múltipla convencional. Trata-se de um processo iterativo que termina após uma convergência satisfatória do coeficiente de autocorrelação. As opções do software foram, então, configuradas para apresentar a primeira e a última iteração.

Verifica-se que a primeira iteração apresenta os valores obtidos numa regressão linear convencional. Todos os modelos revelaram autocorrelação de resíduos na primeira iteração confirmando o fato de que o método convencional não produziria resultados eficientes. Entretanto, após a última iteração, os modelos apresentaram resultados do teste de Durbin-Watson na faixa de rejeição da hipótese de autocorrelação dos resíduos revelando, portanto, sucesso dos processos auto-regressivos usados.

A execução das regressões iniciou-se mantendo como variável dependente a série com ajustamento sazonal de benefícios previdenciários divididos pelo PIB (sasBP/PIB). Foram construídos modelos contendo todas as combinações possíveis das variáveis independentes. Este processo foi repetido para as receitas previdenciárias divididas pelo PIB (sasRP/PIB) e para os saldos previdenciários divididos pelo PIB (sasSP/PIB). Ao todo, 381 modelos foram obtidos nesta abordagem combinatória.

Além dos modelos acima relacionados, foram testados modelos com razões de dependência alternativas, calculadas usando limites de idade inferiores de 15 e 20 anos e limites superiores de 55, 60 e 65 anos. Desta forma, verificou-se a influência daquelas aposentadorias concedidas a segurados com menos de 65 anos nos resultados deste estudo. Os modelos obtidos com as RDDI alternativas revelaram-se muito semelhantes e, por isso, foi mantida a definição original encontrada na literatura.

Desse conjunto, foram selecionados para avaliação da variável estatística com relação às suposições do método, aqueles que apresentaram todos os coeficientes significantes no nível de 5%.

Além dos modelos selecionados pelo critério de significância dos coeficientes, todos aqueles em que havia apenas uma variável independente também foram selecionados para apresentação. Esse segundo subconjunto permitiu separar as variáveis que confirmaram as referências da literatura e identificar aquelas não revelaram significância estatística na relação com as variáveis previdenciárias.

Para cada modelo selecionado, foram verificadas as premissas do método de análise de regressão convencional por meio da análise de resíduos, incluindo testes de estacionariedade. Além disso, quando submetidos ao software Eviews versão 5.0 para confirmação, os modelos selecionados apresentaram-se válidos

com valores absolutos de “*Inverted AR Roots*” menores que 1 (um). Desta forma, garantiu-se que nenhuma regressão espúria foi usada para fundamentar as conclusões deste estudo.

6.5.

Modelos selecionados para benefícios

Os modelos que envolvem a variável previdenciária de benefícios relativos ao PIB e apenas uma variável independente têm seus principais parâmetros apresentados na tabela 11, a seguir.

Modelos com variável dependente sasBP/PIB e uma variável preditora					
Variável independente	Adjusted R Square	Durbin-Watson	Unstandardized Coefficients	t	Sig
SM	0,692	2,051	0,000094	17,710	0,000
RMNH	0,477	2,094	0,000050	7,095	0,000
sqrt RDDI	0,671	2,039	0,837079	16,707	0,000
inv SELIC	0,005	2,441	0,005813	1,623	0,107
TD	-0,014	2,436	0,000000	0,291	0,771
INPC	-0,011	2,452	0,000000	-0,721	0,472
TI	-0,039	2,486	0,000092	0,081	0,936

Tabela 11 – Modelos para benefícios com uma variável independente

Nota-se que o salário mínimo (SM), a renda mensal nominal habitual (RMNH) e a raiz quadrada da razão de dependência demográfica dos idosos (sqrtRDDI) explicaram variações em benefícios previdenciários relativos ao PIB no período analisado. Para estas variáveis, rejeita-se a hipótese de que seus coeficientes individuais sejam zero com nível de significância de 5%. Para as demais variáveis, falha-se em rejeitar esta mesma hipótese e, portanto, conclui-se que elas, individualmente, não contribuem para explicar variações em benefícios relativos ao PIB.

Entre os modelos selecionados, aquele que apresentou a variável estatística com maior poder preditivo para benefícios é constituído pelas variáveis sqrtRDDI e SM. Os parâmetros do modelo são apresentados na tabela 12.

Model Fit Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
,858	,737	,731	,003	2,017

Regression Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig
	B	Std. Error	Beta		
sqrtRDDI	,447	,147	,453	3,039	,003
SM	4,62E-005	,000	,415	2,778	,006
(Constant)	-,082	,040		-2,033	,044

The Prais-Winsten estimation method is used.

ANOVA

	Sum of Squares	df	Mean Square
Regression	0,002884	2	0,001442049
Residual	0,00103	136	7,57468E-06

F = 190,4**Autocorrelation Coefficient**

Rho (AR1)	Std. Error
,349	,080

Tabela 12 – Modelo multivariado para benefícios

6.6.**Modelos selecionados para receitas**

Os procedimentos anteriores foram repetidos, tomando-se como variável dependente as receitas previdenciárias relativas ao PIB. Os modelos com uma variável independente são apresentados na tabela 13, a seguir.

Modelos com variável dependente sasRP/PIB e uma variável preditora					
Variável independente	Adjusted R Square	Durbin-Watson	Unstandardized Coefficients	t	Sig
SM	0,065	2,203	0,00001	3,416	0,001
RMNH	0,562	1,956	0,00003	8,366	0,000
sqrt RDDI	0,012	2,253	0,06962	1,920	0,057
inv SELIC	-0,008	2,287	0,00149	0,912	0,363
TD	0,027	2,228	-0,00052	-2,421	0,017
INPC	-0,012	2,289	-0,00021	-0,555	0,580
TI	-0,016	2,345	-0,00059	-1,086	0,282

Tabela 13 – Modelos para receitas com uma variável independente

Quando tomadas individualmente, as variáveis de rendimentos médios nominais habituais (RMNH), salário mínimo (SM) e taxa de desemprego (TD) explicaram variações em receitas previdenciárias relativas ao PIB uma vez que apresentaram coeficientes significantes a 5%. Por esses resultados, falha-se em

rejeitar a hipótese de que os coeficientes das demais variáveis sejam zero e conclui-se que elas não são boas preditoras de receitas quando isoladas.

Entre os modelos multivariados para receitas previdenciárias destaca-se aquele constituído pelas variáveis trabalho informal (TI) e salário mínimo (SM), cujos parâmetros são apresentados na tabela 14.

Model Fit Summary					
R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson	
,781	,611	,587	,002	1,950	

Regression Coefficients					
	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
TI	-,001	,000	-,242	-2,744	,008
SM	3,87E-005	,000	,742	8,413	,000
(Constant)	,065	,010		6,747	,000

The Prais-Winsten estimation method is used.

ANOVA			
	Sum of Squares	df	Mean Square
Regression	0,000248	2	0,000123961
Residual	0,000158	50	3,16226E-06

F = 39,2

Autocorrelation Coefficient	
Rho (AR1)	Std. Error
-,168	,139

Tabela 14 – Modelo multivariado para receitas

6.7.

Modelos selecionados para o saldo previdenciário

Novamente os procedimentos foram repetidos usando o saldo previdenciário relativo ao PIB (sasSP/PIB) como variável dependente. Os modelos com uma variável independente são apresentados na tabela 15, a seguir.

Modelos com variável dependente sasSP/PIB e uma variável preditora					
Variável independente	Adjusted R Square	Durbin-Watson	Unstandardized Coefficients	t	Sig
SM	0,538	2,174	-8,2E-05	-12,810	0,000
RMNH	0,207	2,003	-2,6E-05	-3,984	0,000
sqrt RDDI	0,748	2,023	-0,76750	-20,179	0,000
inv SELIC	-0,001	2,613	-0,00533	-1,340	0,182
TD	-0,007	2,586	-0,00069	-1,043	0,299
INPC	-0,014	2,632	0,000132	0,172	0,864
TI	-0,015	2,132	-0,00077	-1,096	0,278

Tabela 15 – Modelos para saldo previdenciário com uma variável independente

As variáveis sqrtRDDI, SM e RMNH apresentaram coeficientes diferentes de zero com o nível adotado nesta pesquisa (5%). Mesmo ao nível de significância mais rígido de 1%, estas variáveis explicam variações no saldo previdenciário relativo ao PIB. As demais variáveis não revelaram significância quando tomadas individualmente.

Entre os modelos selecionados, aquele que apresentou a variável estatística com maior poder preditivo para o saldo previdenciário contém as variáveis: sqrtRDDI e TD, conforme a tabela 16.

Model Fit Summary

R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
,882	,779	,774	,003	2,004

Regression Coefficients

	Unstandardized Coefficients		Standardized Coefficients	t	Sig
	B	Std. Error	Beta		
sqrtRDDI	-,725	,039	-,829	-18,364	,000
TD	-,001	,000	-,108	-2,396	,018
(Constant)	,210	,011		19,819	,000

The Prais-Winsten estimation method is used.

ANOVA

	Sum of Squares	df	Mean Square
Regression	,004	2	,002018
Residual	,001	136	8,43441E-06

Autocorrelation Coefficient

Rho (AR1)	Std. Error
,150	,085

F = 239.26

Tabela 16 – Modelo multivariado para o saldo previdenciário