

4. PROCEDIMENTOS METODOLÓGICOS

Este capítulo se inicia com a apresentação do tipo de pesquisa utilizado. Em seguida, são descritas a população e a amostra do estudo, o método de coleta dos dados, as técnicas estatísticas aplicadas e as formas de tratamento dos dados. Por fim, o capítulo se encerra com a discussão das limitações do método.

4.1. Tipo de pesquisa

Para a classificação da pesquisa, utiliza-se a taxonomia proposta por Aaker, Kumar & Day (1997), que a qualifica em relação a dois aspectos: quanto à finalidade (exploratória, descritiva ou causal) e quanto ao método de coleta dos dados (a partir de fontes primárias ou de fontes secundárias). Adicionalmente, uma pesquisa também pode ser classificada quanto à abordagem temporal: corte transversal (*cross-sectional*) ou longitudinal (Malhotra, 2012).

Quanto à finalidade, esta pesquisa pode ser classificada como descritiva, uma vez que tem como objetivo primordial a descrição das características de determinado fenômeno (no caso deste estudo, a relação entre modo de entrada e desempenho de subsidiárias das empresas multinacionais brasileiras), bem como o estabelecimento de relações entre variáveis, sem interferência ou manipulação por parte do pesquisador.

Quanto ao método de coleta dos dados, esta pesquisa faz uso tanto de fontes secundárias quanto de fontes primárias de dados. Preferencialmente, os dados foram coletados a partir de fontes secundárias (neste caso, bancos de dados de empresas públicas e privadas, artigos, livros e revistas especializadas em negócios internacionais e arquivos de empresas e instituições de pesquisa). Dados oriundos de fontes primárias (neste caso, dados obtidos por meio de contatos pessoais ou telefônicos com os gestores das empresas) foram utilizados apenas em última instância, ou seja, quando não houve possibilidade de coletá-los a partir de fontes secundárias.

A opção pela coleta dos dados de forma preferencial a partir de fontes secundárias pode ser justificada por três razões: (1) economia de tempo e de recursos financeiros em comparação com fontes primárias (Kinneer & Taylor, 1979); (2) como todas as empresas analisadas nesta pesquisa são oriundas de um

único país (Brasil), e, assim, encontram-se sujeitas à mesma legislação tributária, não há diferença significativa nas práticas contábeis utilizadas para a confecção de suas demonstrações financeiras (balanços e demonstrações de resultados), o que possibilita a comparação entre os dados provenientes de diferentes empresas; e (3) como o número total de empresas industriais brasileiras com investimentos diretos no exterior antes da realização desta pesquisa era desconhecido e poderia ser pequeno, o levantamento dos dados a partir de uma *survey* (método comumente empregado em pesquisas de negócios internacionais) poderia resultar em uma amostra final insuficiente para a realização dos testes estatísticos, tendo em vista a grande possibilidade de se obter baixa taxa de respostas⁴.

Finalmente, no que se refere à abordagem temporal, o presente estudo se caracteriza como pesquisa de corte transversal, uma vez que os dados coletados cobrem determinado momento no tempo, e não tratam da evolução do fenômeno estudado ao longo do tempo.

4.2. População e amostra

A população-alvo do estudo foi constituída pelas subsidiárias internacionais das empresas industriais brasileiras (conforme CNAE Seção “C” do IBGE – Indústrias de Transformação – códigos CNAE 10 a CNAE 33) com controle de capital nacional e investimentos diretos no exterior. Optou-se por não envolver outros segmentos de mercado nesta pesquisa devido à grande dificuldade para o estabelecimento de antecedentes comuns dos modos de entrada, bem como de critérios para a comparação do desempenho de empresas, de segmentos muitas vezes díspares (p.ex., segmento industrial *versus* financeiro).

Sendo o objetivo desta tese a avaliação dos modos de entrada adotados pelas MNEs industriais brasileiras, bem como dos impactos decorrentes destas escolhas no desempenho de suas subsidiárias, a unidade amostral considerada foi a subsidiária operacional de produção ou de distribuição, tendo sido descartadas as

⁴ Harzing (1997) afirma que as taxas de resposta para *surveys* internacionais enviadas por correio geralmente variam entre 6% e 16%. No caso específico de estudos empíricos de modos de entrada no mercado externo, várias *surveys* apresentaram taxas de resposta iguais ou inferiores a 20%, (chegando, em alguns casos, a números inferiores a 10%), como, por exemplo: Carneiro, 2007 (15,4%); Dikova & Witteloostuijn, 2007 (7,5%); Ekeledo & Sivakumar, 2004 (20%); Harzing, 2002 (20%); Kim & Gray, 2008 (17%); Rajan & Pangarkar, 2000 (17%); Slangen & Tulder, 2009 (19,2%); Taylor, Zou & Osland, 1998 (18%).

subsidiárias que apresentavam função de *holding* pura ou que apenas exerciam atividades administrativas, financeiras ou de vendas.

O levantamento desta população, entretanto, mostrou-se desde o início do trabalho de difícil execução. Infelizmente, ainda hoje, não existem dados oficiais sobre a atuação das empresas brasileiras no exterior, algo comum em países desenvolvidos como Japão e Estados Unidos, onde as empresas fornecem informações detalhadas sobre suas operações internacionais para fins estatísticos. A única indicação do tamanho da população das MNEs brasileiras foi a informação obtida junto ao BCB de que 960 pessoas jurídicas em 2007, e, ainda, 997 pessoas jurídicas em 2008, afirmaram possuir IDE na declaração “Capitais Brasileiros no Exterior”, documento de cunho obrigatório para residentes no Brasil com ativos no exterior. Cabe ressaltar, entretanto, que nas estatísticas divulgadas anualmente pelo BCB, as empresas declarantes não são identificadas, não há indicação de quais investimentos são de natureza operacional, não há discriminação do número de empresas por CNAE e tampouco indicação sobre a origem do capital.

Mesmo que a população-alvo fosse plenamente conhecida, certamente não haveria possibilidade da coleta dos dados de todas as MNEs brasileiras para a operacionalização das variáveis envolvidas neste estudo. Em casos como este, torna-se necessária a utilização de uma amostra da população-alvo, ou seja, um subconjunto desta população. Segundo Aaker, Kumar & Day (1997), as amostras podem ser divididas em probabilísticas e não-probabilísticas. As amostras do primeiro tipo são formadas por subconjuntos de uma população que asseguram determinada representatividade, ao dar a todo elemento da população uma chance maior do que zero de ser escolhido. Já as do segundo tipo não asseguram esta representatividade a todos os membros da população.

No caso específico deste estudo, a amostra utilizada pode ser considerada como não-probabilística, já que foi composta pelas subsidiárias onde foi possível a coleta dos dados necessários para a realização dos testes estatísticos, ou seja, a escolha das subsidiárias não se deu de forma aleatória. Entretanto, a fim de que a amostra final fosse a mais representativa possível da população, tomou-se o cuidado de coletar dados suficientes para que todos os segmentos industriais estivessem nela representados.

Tendo em vista a carência de dados oficiais consolidados sobre a atuação internacional das empresas brasileiras, optou-se pela construção de um banco de dados, que foi especialmente desenhado para os propósitos desta pesquisa. Na próxima seção, são apresentados, em detalhes, todos os procedimentos realizados para executar o mapeamento da população-alvo do estudo, bem como a amostra efetivamente utilizada para a realização dos testes estatísticos.

4.3. Coleta dos dados

Os dados foram coletados em três etapas distintas: as duas primeiras etapas executadas por meio de pesquisa documental e a terceira por meio de contatos pessoais e telefônicos com gestores das empresas.

A primeira etapa da pesquisa teve por objetivo identificar a população de empresas industriais brasileiras com investimentos diretos no exterior, a origem do capital destas empresas (para posterior eliminação daquelas com controle de capital estrangeiro) e os países onde estavam sediadas. Para a execução desta etapa, foram realizadas consultas às seguintes fontes públicas e privadas:

- Órgãos da administração pública federal: BNDES e BCB;
- Instituições especializadas em negócios internacionais: SOBEET e UNCTAD;
- Publicações especializadas em negócios: Valor Econômico, Exame e Época Negócios;
- Arquivos do NUPIN;
- *Websites* corporativos das empresas industriais identificadas.

Após minucioso trabalho de pesquisa, foram identificadas 224 empresas industriais com 738 subsidiárias internacionais distribuídas por 55 países. Neste número total de subsidiárias, além daquelas ligadas às atividades de produção ou distribuição, também estavam incluídas as subsidiárias exclusivamente dedicadas à atividade de vendas, que foram eliminadas apenas na etapa seguinte. Entretanto, apesar de todo o esforço empreendido para a realização deste levantamento, cabe ressaltar que não há garantias de que todos os elementos da população de MNEs industriais brasileiras estejam incluídos na listagem final obtida, assim como

não se pode assegurar que todas as suas subsidiárias internacionais operacionais tenham sido devidamente identificadas.

O Quadro 23 a seguir apresenta a relação de todos os países onde as MNEs industriais brasileiras possuem subsidiárias com atividades operacionais e a quantidade de empresas em cada um deles devidamente agrupadas pelos códigos CNAE 2 dígitos.

A análise deste quadro releva que os países com maior quantidade de MNEs industriais brasileiras são, em ordem decrescente, Estados Unidos (108), Argentina (82), México (43), Chile (33) e China (31). Cabe ressaltar que apenas quatro segmentos industriais (CNAE 11, 12, 18 e 21) não apresentam subsidiárias nos Estados Unidos, fato que confirma a relevância deste país no processo de inserção internacional das empresas brasileiras.

Este quadro também mostra que os segmentos com maior quantidade de MNEs industriais brasileiras são, em ordem decrescente, produtos alimentícios (CNAE 10; 26), veículos automotores e carrocerias (CNAE 29; 21), produtos eletrônicos (CNAE 26; 18), produtos de borracha e material plástico (CNAE 22; 16) e máquinas e equipamentos (CNAE 28; 16). Vale destacar que o segmento de produtos alimentícios (CNAE 10) é aquele que está presente em um maior número de países (31), o que confirma sua importância no processo de internacionalização das empresas brasileiras.

Quadro 23 – Países com subsidiárias operacionais das MNEs industriais brasileiras

Países hospedeiros dos investimentos	Códigos CNAE 2 dígitos																				Total				
	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29		30	31	32	33
África do Sul	1												1							2					4
Alemanha	5		1										1	1	1	2	3	1	3	4					22
Arábia Saudita											1														1
Argentina	7	1	8		9	1	4		1	7	1	3	5	5	1	7	5	8	7				2	82	
Austrália	1											1		1				1		1	1			6	
Áustria																			1					1	
Bahrein						1																		1	
Bélgica	2				1				1		1	2					1							8	
Bolívia		1							1		1	1	1		1									6	
Canadá		1	1									2		2		1								5	
Chile	4	1	1	1	4		1		1	5	1	2	4	1	1	1	1	1	1	2		1	2	33	
China	4		2		2		2		1			1	1	1	2	4	2	1	6	2	2	1		31	
Cingapura	2								1					1			1				1			6	
Colômbia	1				1				1	4	2	1	1	3	1	1	1	1	2				1	21	
Croácia																	1							1	
Dinamarca																	1							1	
Egito	1																			1				2	
El Salvador					1																			1	

Fonte: Elaboração do autor.

Países hospedeiros dos investimentos	Códigos CNAE2 dígitos																																						
	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	Total														
Emirados Árabes Unidos	1															1		1		2															1			6	
Equador		1	1										1																									3	
Eslôvãquia													2			2																						4	
Eslôvênia													1																									1	
Espanha	1		1		2	1				3		1	1		2	1	1	1	2																			16	
Estados Unidos	11		10	2	5	5	5	5	1	3	3	8	5	9	3	11	4	5	12	3	3	1	3	2													108		
França	2		1						2	2	1	1	1	1	1	1	1	1	1	1	1	1	1														14		
Guatemala		1													1																							2	
Holanda	6	1			1				1	1	1	1	1	1	1			1																				13	
Honduras																1																						1	
Hungria	1																			1																		2	
Índia				1											1			1		1																		4	
Israel																	1																					1	
Itália	2		1	1	1					2				1				1	1	1	1	1																12	
Japão	4									1								1		1																		7	
Líbano	1																																						1
Luxemburgo															2																								2
Malásia	1													1																									2
México	3		2	2	2			1		8	2	3	1	1	1	1	4	3	3	6																			43

Fonte: Elaboração do autor.

Países hospedeiros dos investimentos	Códigos CNAE 2 dígitos																										Total
	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33			
Noruega	1																									1	
Panamá	1												1							1		1				4	
Paraguai	1	1	2	1		1			1	1	2	1	1					1							13		
Peru	1	1				2			1	2	1	2	2	1	1				1					1	16		
Porto Rico											1							1							2		
Portugal	2								1	3		1	2		1	1	1	1	1	1	1	1			14		
Reino Unido	3			1		1	3	1		1		1						2	2	2			1		18		
Republica Dominicana		1										1													2		
Republica Tcheca															1										1		
Romênia							1																		1		
Rússia	2																2			1					5		
Suécia													1					1							2		
Suíça				1			2										2								5		
Taiwan													1			1									3		
Turquia	1															1									2		
Ucrânia																		1							1		
Uruguai	9	2				2			1	4		3	2	1					1						25		
Venezuela	2	1							1	3	1			1					1						11		
Total de Empresas	26	3	2	13	4	14	6	6	0	1	14	8	16	8	13	8	18	10	16	21	4	4	6	3	224		

Fonte: Elaboração do autor.

Nota: A linha "Total de Empresas" refere-se ao número de empresas representadas em cada CNAE e não ao somatório dos valores lançados em cada coluna.

Na segunda etapa, foram mapeados os modos de entrada adotados (nível de propriedade, forma de estabelecimento e tipo de atividade) pelas MNEs industriais brasileiras quando da criação de suas subsidiárias internacionais, assim como foram levantadas informações contábeis e financeiras das empresas. Nesta fase, foram eliminadas todas as subsidiárias exclusivamente dedicadas à atividade de vendas. Para a execução desta etapa, foram necessárias consultas às seguintes fontes públicas e privadas:

- Órgãos da administração pública federal: BNDES, BCB, CVM e IBGE;
- Publicações e consultorias especializadas: Valor Econômico, Exame, Época Negócios, Economática e Broadcast;
- Literatura de negócios internacionais: SOBEET (2007) e Fleury & Fleury (2011);
- Ferramentas de busca na internet: Google e Bing;
- Arquivos do NUPIN;
- *Websites* corporativos e demonstrações financeiras das empresas industriais identificadas.

Finalmente, na terceira e última etapa, foram levantados os demais dados necessários para a realização dos testes estatísticos, cuja coleta não foi possível nas duas fases anteriores. Para a realização desta etapa, foram efetuados contatos pessoais e telefônicos com gestores das empresas identificadas.

Ao final das três etapas da coleta de dados, várias subsidiárias tiveram que ser descartadas devido à falta de informações imprescindíveis para a realização dos testes estatísticos. A maior dificuldade encontrada durante este levantamento foi a obtenção de dados relativos ao desempenho (lucratividade), informações estas normalmente consideradas de caráter estratégico pelas empresas, apesar da promessa do pesquisador de não apresentar dados individualizados no relatório final (confidencialidade).

Apesar de todas estas dificuldades, a amostra final apresentou um número de empresas e subsidiárias em percentual superior ao normalmente obtido com a realização de *surveys*: 133 empresas com 280 subsidiárias sediadas em 39 países. Assim, 60% das empresas e 38% das subsidiárias internacionais operacionais da suposta população de MNEs industriais brasileiras, bem como 71% dos países

hospedeiros de seus investimentos, fizeram parte da amostra final. Além desta quantidade relativamente expressiva, cabe ressaltar que empresas de todos os segmentos industriais (com exceção apenas do CNAE 18, que já não apresentava nenhuma empresa listada no Quadro 23) tiveram subsidiárias incluídas na amostra final, o que confirma sua representatividade em relação à população.

O Quadro 24 a seguir apresenta a caracterização das empresas industriais e suas subsidiárias internacionais incluídas na amostra final.

Quadro 24 – Caracterização da amostra final

Empresas			Subsidiárias		
Ativos (em R\$ milhões)	Total	Partic.	Nível de propriedade	Total	Partic.
Inferiores a 100	31	23,3%	WOS	213	76,1%
Entre 100 e 250	29	21,8%	JV	67	23,9%
Entre 250 e 500	23	17,3%			
Entre 500 e 1.000	14	10,5%	Forma de estabelecimento	Total	Partic.
Acima de 1.000	36	27,1%	Aquisição	108	38,6%
			<i>Greenfield</i>	172	61,4%
Receitas (em R\$ milhões)	Total	Partic.	Tipo de atividade	Total	Partic.
Inferiores a 100	33	24,8%	Produção	139	49,6%
Entre 100 e 250	29	21,8%	Distribuição	141	50,4%
Entre 250 e 500	28	21,0%			
Entre 500 e 1.000	13	9,8%	Idade	Total	Partic.
Acima de 1.000	30	22,6%	Menor ou igual que 2 anos	77	27,5%
			Entre 3 e 5 anos	72	25,7%
Número de funcionários	Total	Partic.	Entre 6 e 10 anos	65	23,2%
Menor ou igual que 250	19	14,3%	Entre 11 e 20 anos	49	17,5%
Entre 250 e 500	24	18,1%	Maior que 20 anos	17	6,1%
Entre 500 e 2.000	40	30,1%			
Entre 2.000 e 5.000	24	18,0%	Tipo de indústria	Total	Partic.
Acima de 5.000	26	19,5%	Alta tecnologia	125	44,6%
			Baixa tecnologia	155	55,4%
Experiência internacional	Total	Partic.	Localização geográfica	Total	Partic.
Menor ou igual que 2 anos	18	13,5%	América do Sul	137	48,9%
Entre 3 e 5 anos	28	21,1%	América do Norte	55	19,7%
Entre 6 e 10 anos	26	19,5%	América Central e Caribe	25	8,9%
Entre 11 e 20 anos	33	24,8%	Europa	44	15,7%
Maior que 20 anos	28	21,1%	Ásia e Oceania	12	4,3%
Abertura do capital	Total	Partic.	Oriente Médio	5	1,8%
Capital aberto	43	32,3%	África	2	0,7%
Capital fechado	90	67,7%			

Fonte: Elaboração do autor.

4.4. Técnicas estatísticas e tratamento dos dados

Esta seção dedica-se à apresentação do tratamento dos dados e das técnicas estatísticas utilizadas para a avaliação da escolha dos modos de entrada e do desempenho das subsidiárias. Cabe destacar que foi utilizado o pacote estatístico SPSS PASW Statistics 18TM, rodando sobre o sistema operacional Microsoft Windows 7TM, para a operacionalização de todos os procedimentos aqui descritos.

4.4.1. Técnicas estatísticas

Escolha do modo de entrada

No caso da escolha do modo de entrada (nível de propriedade, forma de estabelecimento e tipo de atividade), a técnica estatística selecionada para a realização dos testes dos relacionamentos entre as variáveis independentes e dependentes foi a regressão logística, tendo em vista a natureza discreta das variáveis dependentes. Técnicas de regressão múltipla (p.ex., mínimos quadrados ordinários – OLS) não podem ser utilizadas nestes casos, pois as variáveis em questão não são contínuas. Como todas as variáveis dependentes relacionadas à escolha do modo de entrada são dicotômicas (WOS *versus* JV, aquisição *versus* investimento *greenfield* e produção *versus* distribuição), a técnica estatística mais apropriada é um modelo logit binomial (Hair et al., 2009).

Modelos de regressão logística são uma combinação de regressão múltipla e análise discriminante múltipla. Esta técnica é semelhante à análise de regressão múltipla, onde uma ou mais variáveis independentes são utilizadas para prever uma única variável dependente. O que distingue um modelo de regressão logística de uma regressão múltipla é o fato de a variável dependente ser não-métrica, tal como na análise discriminante. A escala não-métrica da variável dependente exige diferenças no método de estimação e nas premissas sobre o tipo de distribuição subjacente, porém, nas demais características, a regressão logística é bastante similar à regressão múltipla. Desta forma, uma vez que a variável dependente esteja corretamente especificada e a técnica de estimação apropriada seja empregada, os fatores básicos considerados na regressão múltipla também podem ser utilizados na regressão logística.

As principais vantagens da regressão logística quando comparadas à análise discriminante, e até mesmo à regressão múltipla, se devem, em grande parte, à falta de premissas requeridas em uma análise de regressão logística (Hair et al., 2009). Esta regressão não requer qualquer forma específica de distribuição das variáveis independentes e questões como heterocedasticidade não tem tanta importância como na análise discriminante. Além disso, a regressão logística não exige relacionamentos lineares entre as variáveis independentes e a variável dependente como no caso da regressão múltipla e, ainda, pode lidar com efeitos não-lineares, mesmo quando termos exponenciais e polinomiais não estejam explicitamente adicionados como variáveis independentes por causa do relacionamento logístico.

Segundo Hair et al. (2009), o único grande problema da regressão logística em relação às premissas refere-se ao fato de que ela também é afetada pela multicolinearidade entre as variáveis independentes, de forma similar à análise discriminante. A multicolinearidade cria uma variância “compartilhada” entre as variáveis, diminuindo assim a capacidade de prever a medida dependente, bem como determinar os papéis relativos de cada variável independente.

Uma forma de identificar a multicolinearidade é produzir uma matriz de correlações de todas as variáveis preditoras e verificar se algumas delas possuem correlação muito alta, ou seja, valores acima de 0,8 ou 0,9 (Field, 2009). Outra forma para se averiguar a multicolinearidade é por meio do fator de inflação da variância (VIF), que indica se um preditor tem uma forte relação linear com outros preditores. Embora não existam regras rígidas sobre qual valor de VIF deve causar preocupação, Field (2009) e Hair et al. (2009) sugerem o valor limite de 10, acima do qual há indicação de alta colinearidade. Relacionada com o VIF está o valor de tolerância, que é seu recíproco ($1/VIF$). Como tal, valores abaixo de 0,1 indicam sérios problemas de multicolinearidade.

Outra vantagem da regressão logística é que precisamos saber apenas se determinado evento ocorreu para então utilizar um valor dicotômico como variável dependente. A partir desse valor dicotômico, o procedimento prevê sua estimativa da probabilidade de que o evento ocorrerá ou não. Se a probabilidade for maior que 0,5, então a previsão será sim, caso contrário será não. Devido a todas estas características, os modelos logit binomiais têm sido utilizados com

bastante frequência em estudos empíricos de modos de entrada (p.ex., Benito, 1997; Brouthers & Dikova, 2010; Cho & Padmanabhan, 2005; Cui, Jiang & Stening, 2011; Demirbag, McGuinness & Altay, 2010; Dow & Ferencikova, 2010; Drogendijk & Slangen, 2006; Elango, 2005; López-Duarte & Vidal-Suárez, 2010; Tseng & Lee, 2010; Wang & Schann, 2008).

Apesar do uso generalizado de modelos logit nas pesquisas de modos de entrada, alguns autores (p.ex., Arregle, Hébert & Beamish, 2006; Hough, 2006; Makino, Isobe & Chan, 2004; Mani, Anita & Rindfleisch, 2007) argumentam que as decisões sobre modos de entrada aparentam ser um fenômeno de múltiplos níveis: as explicações vêm de variáveis ao nível da subsidiária, ao nível da empresa e de interações entre elas. No entanto, segundo estes autores, muitas pesquisas misturam ou confundem estes níveis de análise, pois seus resultados são derivados de análises que não são multiníveis, tais como regressões logísticas. Ou seja, estes estudos implicitamente apresentam modelos conceituais multiníveis com variáveis sendo medidas em dois níveis diferentes (empresa e subsidiária), mas ignoram essa dimensão multinível em seu método de análise de dados.

Esta omissão pode ter importantes consequências para a compreensão dos determinantes das decisões dos modos de entrada, bem como para a validade e robustez das evidências empíricas apresentadas na literatura. Primeiro, ela limita o desenvolvimento conceitual de modelos mais abrangentes que considerem as variáveis ao nível da subsidiária, ao nível da empresa e suas interações. Depois, ela envolve potenciais problemas estatísticos de desagregação, de correlação intraclasse e de precisão incorreta (Hox, 2010). Como uma mesma empresa pode ter várias subsidiárias internacionais, estes IDEs não podem ser considerados como independentes, sendo a estrutura de dados hierárquica. Métodos de múltiplos níveis foram projetados especificamente para superar tais problemas, contudo, eles têm sido praticamente ignorados pela literatura acadêmica de negócios internacionais (Arregle, Hébert & Beamish, 2006).

Para lidar com tais dados hierárquicos em estudos de modos de entrada, os pesquisadores frequentemente desagregam dados ao nível da empresa para utilizar um método não-multinível (p.ex., Brouthers, 2002; Lu, 2002; Pak & Park, 2004; Yiu & Makino, 2002). Assim sendo, cada IDE é tratado como uma observação independente e variáveis ao nível da empresa (p.ex., tamanho da empresa) são

duplicadas para cada observação ao nível da subsidiária. Por exemplo, 12 IDEs de uma mesma empresa são considerados como 12 IDEs independentes e possuem 12 (idênticas) observações para uma mesma variável. Esta não-independência das observações pode levar a resultados estatísticos enviesados (Hox, 2010).

Segundo Arregle, Hébert & Beamish (2006), esta questão de desagregação não tem tanta importância quando as empresas estão envolvidas em um ou poucos IDEs. Nesses casos, mesmo se os modelos conceituais forem multiníveis, as hipóteses podem ser testadas com métodos que não sejam multiníveis sem sérios problemas. Como esta observação se encaixa perfeitamente na situação deste estudo, onde o modelo conceitual pode ser considerado como multinível, porém o número médio de IDEs é de apenas dois por empresa, optou-se pela não utilização de uma análise multinível no lugar da regressão logística inicialmente planejada.

Avaliação do desempenho

Já para a análise do impacto no desempenho da subsidiária em função do modo de entrada adotado, a técnica estatística selecionada para os testes dos relacionamentos entre as variáveis independentes e a variável dependente foi a regressão múltipla de mínimos quadrados ordinários (OLS), tendo em vista que a variável dependente (retorno sobre ativos), neste caso, é contínua. Entretanto, uma regressão múltipla OLS somente pode estimar corretamente o impacto do modo de entrada no desempenho se a estratégia for escolhida aleatoriamente ou se todos os fatores que influenciam tanto o modo de entrada quanto o desempenho estiverem incluídos no modelo a ser analisado. Caso contrário, esta regressão produzirá estimativas tendenciosas e inconsistentes (Tan, 2009). Tendo em vista que as MNEs não selecionam aleatoriamente seus modos de entrada, mas o fazem com base em seus atributos e nas condições do mercado, a escolha estratégica torna-se endógena e autosselecionada (Brouthers, Brouthers & Werner, 2008b). Desta forma, seguindo sugestão de Shaver (1998), foi incluída uma correção para autosseleção na equação de regressão do desempenho utilizando o procedimento proposto por Heckman (1979).

A variável de correção para autosseleção (também conhecida como “razão inversa de Mills”) foi adicionada ao modelo porque outros estudos empíricos forneceram evidências de que características não-observáveis da empresa podem

afetar seu desempenho (p.ex., Brouthers, Brouthers & Werner, 2003, 2008b; Shaver, 1998; Slangen & Hennart, 2008b; Tan, 2009). Esta variável é calculada a partir dos parâmetros estimados de cada equação de seleção. Assim sendo, todas as variáveis do modelo são utilizadas para estimar a razão inversa de Mills. A eventual significância da variável de correção para autoseleção indica que características não-observáveis da empresa estão relacionadas com a escolha do modo de entrada e o desempenho. Tendo em vista que o pacote estatístico SPSS 18 não possui uma função própria que contemple o procedimento proposto por Heckman (1979), sua operacionalização no SPSS foi realizada a partir do algoritmo desenvolvido por Smits (2003).

Tal como a regressão logística, a regressão múltipla também é afetada pela multicolinearidade entre as variáveis independentes. Segundo Hair et al. (2009), entretanto, algumas premissas adicionais também precisam ser atendidas para que o modelo gerado a partir de uma regressão múltipla possa ser generalizado para uma população mais ampla: (1) linearidade do fenômeno medido; (2) normalidade da distribuição dos termos de erro; (3) variância constante (homocedasticidade) dos termos de erro; e (4) independência dos termos de erro. Para a avaliação do atendimento às três primeiras premissas, foi utilizada a análise gráfica dos resíduos, seguindo recomendação de Field (2009) e Hair et al. (2009). No caso da independência dos erros, uma eventual violação a esta premissa foi verificada a partir do teste de Durbin-Watson. Segundo Field (2009), a estatística deste teste pode variar entre 0 e 4, com um valor de 2 (ou bem próximo a ele) significando que os resíduos não são correlacionados. Como regra conservadora, valores menores do que 1 ou maiores do que 3 devem ser motivo de preocupação.

4.4.2. Tratamento dos dados

Antes da aplicação das técnicas estatísticas em si, os dados coletados foram submetidos aos seguintes tratamentos:

- Dados ausentes (*missing data*);
- Observações atípicas (*outliers*) e casos influentes (*influential cases*);
- Adequabilidade do tamanho da amostra.

Dados Ausentes (missing data)

Um processo de dados ausentes é qualquer evento sistemático externo ao respondente (como erros na entrada de dados ou problema na coleta de dados) ou ação por parte do respondente (como recusa a responder) que conduz a valores ausentes (Hair et al., 2009). Qualquer resultado estatístico baseado em dados com um processo não-aleatório de dados ausentes pode estar enviesado. Este viés ocorre quando o processo de dados ausentes faz com que certos dados fiquem faltando na tabulação, levando a resultados incorretos. Segundo Hair et al. (2009), os tratamentos ou ações corretivas para lidar com dados ausentes podem ser classificados em quatro categorias, com base na aleatoriedade do processo de dados ausentes e no método empregado para estimar os mesmos:

- (1) O tratamento mais simples e direto para lidar com dados ausentes é incluir na análise apenas as observações com dados completos, método também conhecido como abordagem de caso completo;
- (2) Outro remédio simples para dados ausentes é eliminar os casos e/ou variáveis problemáticas. Nesse tratamento, o pesquisador determina a extensão dos dados ausentes em cada caso e em cada variável e, então, elimina aqueles com níveis excessivos;
- (3) Uma terceira categoria de ações corretivas para dados ausentes é a utilização de toda a informação disponível de um subconjunto de casos que é, então, generalizado para a amostra como um todo. Conhecido como abordagem de disponibilidade total, esse método é utilizado principalmente para estimar correlações e maximizar a informação de pares disponíveis na amostra;
- (4) A última categoria para o tratamento de dados ausentes envolve a substituição destes dados por valores estimados com base em outras informações disponíveis na amostra. Alguns exemplos deste tipo de tratamento de dados ausentes são: substituição pela média, substituição por um caso, atribuição por regressão e atribuição múltipla.

No caso deste estudo, optou-se pela inclusão apenas das observações completas na amostra final, ou seja, foi utilizada a abordagem de caso completo. A principal razão para esta escolha foi devido ao fato de a grande maioria dos

dados ausentes referirem-se à lucratividade das subsidiárias, uma variável fundamental para a execução das análises estatísticas.

Observações atípicas (outliers) e casos influentes (influential cases)

Observações atípicas (*outliers*) são aquelas com uma combinação única de características identificáveis como sendo notadamente diferentes das demais observações. Estas observações não podem ser categoricamente caracterizadas como benéficas ou problemáticas, mas devem ser vistas no contexto da análise e avaliadas pelos tipos de informações que possam fornecer. Segundo Hair et al. (2009), as ocorrências de *outliers* podem ser divididas em quatro classes:

- (1) A primeira classe surge de um erro de procedimento, tal como erro na entrada de dados ou falha na codificação. Essas observações atípicas devem ser identificadas durante o estágio de limpeza dos dados, mas se algumas permanecerem, elas devem ser eliminadas ou registradas como valores ausentes;
- (2) A segunda classe é aquela que ocorre como o resultado de um evento extraordinário, o que então explica a peculiaridade da observação. O pesquisador deve decidir, então, se o evento extraordinário deve ser mantido ou eliminado da análise;
- (3) A terceira classe refere-se a observações extraordinárias para as quais o pesquisador não tem explicação. Apesar de estes *outliers* serem aqueles com maior probabilidade de eliminação, eles poderão ser mantidos se o pesquisador considerar que os mesmos representam um segmento válido da população;
- (4) A quarta e última classe contém as observações que estão no intervalo usual de valores para cada variável, mas são únicas em sua combinação de valores entre as variáveis. Em tais situações, o pesquisador deve reter a observação, a não ser que exista uma evidência que desconsidere o *outlier* como um membro válido da população em estudo.

Para a identificação de *outliers*, foi utilizada a distância D^2 de Mahalanobis, uma avaliação multivariada de cada observação em um conjunto de variáveis (Hair et al., 2009). Este método mede cada distância da observação no espaço multidimensional do centro médio (centróide) de todas as observações,

fornecendo um único valor para cada observação, independentemente de quantas variáveis forem consideradas. Valores mais elevados de D^2 representam observações mais distantes da distribuição geral de observações neste espaço multidimensional. Para fins de interpretação, a medida D^2 de Mahalanobis possui propriedades estatísticas que permitem realizar testes de significância. O valor de D^2/df , a medida D^2 de Mahalanobis dividida pelos graus de liberdade (número de variáveis envolvidas), é aproximadamente distribuída como uma *t-value*. Dada a natureza dos testes estatísticos, Hair et al. (2009) sugerem que níveis de significância conservadores (p.ex., 0,005 ou 0,001) sejam utilizados como valor limite para a designação como um *outlier*. Assim sendo, observações com valores de D^2/df superiores a 2,5 em amostras pequenas e 3 ou 4 em grandes amostras podem ser designados como possíveis *outliers*.

Já os casos influentes (*influential cases*) são observações que têm uma influência desproporcional sobre um ou mais aspectos das estimações de regressão. Essa influência pode estar baseada em valores extremos das variáveis independentes ou dependentes, ou de ambas. Segundo Hair et al. (2009), observações influentes podem ser “boas”, quando reforçam o padrão dos demais dados restantes, ou “más”, quando um único ou pequeno conjunto de casos afeta indevidamente as estimações de regressão. Casos influentes, contudo, são de difícil detecção, pois nem sempre um caso influente é também um *outlier* e nem sempre um *outlier* pode ser considerado um caso influente. Além disso, o pacote estatístico SPSS 18 não apresenta nenhum mecanismo próprio para a identificação de casos influentes. Desta forma, optou-se pela utilização da distância de Cook, que é uma estatística que leva em consideração o efeito de um único caso sobre o modelo como um todo. Field (2009) sugere que os valores maiores do que 1 devem ser investigados como possíveis casos influentes.

Adequabilidade do tamanho da amostra

A regressão logística, como qualquer outra técnica estatística, é afetada pelo tamanho da amostra analisada. Amostras muito pequenas possuem tantos erros de amostragem que a identificação das maiores diferenças é improvável. Já amostras muito grandes aumentam tanto o poder estatístico, que qualquer diferença, se relevante ou não, será considerada estatisticamente significativa. Segundo Hair et

al. (2009), como a maioria das situações de pesquisa se situa entre estes dois extremos, há necessidade de se considerar o impacto do tamanho da amostra nos resultados, tanto ao nível geral quanto em uma base de grupo-por-grupo.

O primeiro aspecto do tamanho da amostra é o tamanho total necessário para suportar adequadamente a estimação do modelo logístico. Um fator que distingue a regressão logística de outras técnicas multivariadas é o uso da máxima verossimilhança como técnica de estimação. Esta técnica exige amostras maiores de tal forma que, *ceteris paribus*, a regressão logística necessita de uma amostra maior do que uma regressão múltipla. Hosmer & Lemeshow (2000), por exemplo, recomendam tamanhos de amostra maiores que 400 observações para o uso da regressão logística. Além disso, alguns autores (p.ex., Field, 2009 e Hair et al., 2009) recomendam a divisão da amostra em amostra de análise (*analysis sample*) e amostra de teste (*holdout sample*) como um meio de validar o modelo logístico. Ao fazer esta divisão, os requisitos de tamanho da amostra ainda permanecem para ambas as amostras separadamente, duplicando, assim, o tamanho total da amostra necessário para validação do modelo.

Além do tamanho total da amostra, também é importante o tamanho da amostra por grupo da variável dependente. Hair et al. (2009) recomendam que o tamanho da amostra para cada grupo seja de pelo menos dez observações por variável independente em uma regressão logística, valor bem superior ao mínimo sugerido para uma regressão múltipla, que é de cinco observações por parâmetro.

Para a realização dos testes empíricos, apesar da amostra final atender ao critério mínimo de observações por variável preditora em ambas as regressões, optou-se pela não divisão da amostra para validação do modelo logístico. Caso esta fosse dividida pela metade, por exemplo, cada uma das amostras resultantes (amostras de análise e de teste) ficaria com apenas 140 observações, quantidade bem inferior ao recomendado por Hosmer & Lemeshow (2000). Cabe ressaltar, entretanto, que esse procedimento resulta em um viés ascendente (*upward bias*) na precisão preditiva do modelo, mas certamente é melhor do que não testar o modelo de alguma forma (Hair et al., 2009). Neste caso, a capacidade de previsão do modelo foi verificada a partir da análise do percentual de casos com classificação correta (*hit ratio*) e do coeficiente de determinação (R^2) ajustado.

4.5. Limitações do método

Toda pesquisa científica está sujeita a algumas limitações conceituais e metodológicas e esta tese não constitui exceção. A primeira delas diz respeito ao uso preferencial de dados secundários. Por definição, dados secundários são aqueles que foram coletados no passado para propósitos distintos da pesquisa corrente. Desta forma, estes dados dificilmente atenderão às necessidades de informação do estudo atual (Kinnear & Taylor, 1979). Outra séria limitação de dados secundários é a dificuldade para avaliar sua precisão, tendo em vista que os únicos parâmetros disponíveis para este fim são o exame dos métodos utilizados para a coleta dos dados e a evidência de ter sido um trabalho consciencioso (Aaker, Kumar & Day, 1997).

Outra limitação deve-se ao fato de que muitas empresas, em especial as de pequeno e médio porte, não são obrigadas a publicar informações contábeis e financeiras, por serem de capital fechado. Além disso, a legislação tributária brasileira também não obriga a divulgação de resultados individualizados por subsidiária, independente do porte da empresa e da abertura de seu capital. Devido a estas características, em alguns casos, a única forma encontrada para a coleta dos dados foi por meio de contatos pessoais ou telefônicos com os gestores das empresas. Contudo, muitos deles relutaram em divulgar informações consideradas estratégicas, notadamente aquelas relacionadas à lucratividade. Esta dificuldade causou uma redução significativa na amostra final utilizada nos testes estatísticos.

Adicionalmente, tal como a maioria dos estudos empíricos que abordam o desempenho de empresas, esta pesquisa também pode ter sofrido do chamado “viés do sobrevivente” (*survivor bias*) (Makino & Neupert, 2000), ou seja, aquelas subsidiárias que deixaram de existir (muitas das quais provavelmente devido ao baixo desempenho da operação) não estão representadas na amostra final. Este fato limita o alcance das conclusões sobre o impacto dos modos de entrada no desempenho das subsidiárias.

Outro problema que também pode limitar o alcance das conclusões desta pesquisa é a própria representatividade da amostra final, que foi prejudicada pela inexistência de dados oficiais sobre as empresas brasileiras com investimentos diretos no exterior. Este fato obrigou o pesquisador a consultar diversas fontes

primárias e secundárias e a desenvolver um banco de dados para a consolidação das informações. Entretanto, apesar de todos os cuidados para que os dados fossem os mais precisos possíveis, não há garantias de que o quadro amostral utilizado na pesquisa seja representativo da real população existente, que ainda hoje permanece desconhecida.

Algumas considerações de cunho prático, como disponibilidade de dados e conveniência, além das limitações do pacote estatístico utilizado e da própria capacidade cognitiva do pesquisador, também obrigaram a uma simplificação do modelo testado. Desta forma, nem todas as variáveis explicativas levantadas ao longo da revisão bibliográfica foram incluídas no modelo conceitual, bem como não foram modelados todos os efeitos relevantes, o que exige cuidados especiais quando da interpretação dos resultados.