

3

Modelos de Equações Estruturais Aplicados na Estimação da Satisfação do Consumidor

3.1

Introdução

Este capítulo é dedicado a uma revisão bibliográfica sobre os Modelos de Equações Estruturais aplicados na estimação da Satisfação do Consumidor. A ênfase será dada aos modelos baseados em estruturas de covariâncias, ou abreviadamente, CSM (*Covariance Structure Model*), e aos modelos baseados em Mínimos Quadrados Parciais, ou abreviadamente, PLS (*Partial Least Square*). A fundamentação teórica desses modelos será tratada no próximo capítulo.

Os modelos baseados em estruturas de covariâncias descrevem as variâncias e as covariâncias das variáveis observadas. Estes modelos assumem, por hipótese, que as escalas das variáveis latentes são contínuas ou intervalares. No entanto, este tipo de modelo pode ser ampliado para acomodar escalas nominais ou ordinais (*cf.* Muthén, 1984, 1987). Os *softwares* comerciais mais comuns empregados na estimação dos parâmetros são: LISREL, AMOS, LISCOMP, SEPATH e muitos outros. O principal método de estimação dos parâmetros é o da Máxima Verossimilhança, ou abreviadamente, ML (*Maximum Likelihood*). Os modelos CSM combinam modelos de medidas para as variáveis latentes com modelos de causas e efeitos entre elas.

Os modelos baseados no PLS maximizam o poder (acurácia) preditivo das variáveis latentes. O PLS é composto por um conjunto de procedimentos baseados em métodos de Mínimos Quadrados Ordinários, desenvolvidos para estimar as relações entre as variáveis do modelo. Cabe mencionar que este método é distinto do método PLS de estimação de regressões múltiplas, muito utilizado quando as variáveis predictoras são intercorrelacionadas. Este método também é distinto do método CSM em vários aspectos, como será mostrado adiante, embora ambos possam ser empregados com a mesma finalidade. Os *softwares* mais comuns

empregados na estimação dos parâmetros são: LVPLS, PLS-SEM / PLS-GRAPH, etc.

O capítulo está organizado da seguinte maneira. A seção 3.2 aborda o problema dos erros de medidas. A seção 3.3 faz uma breve introdução sobre os Modelos de Equações Estruturais e apresenta as principais características dos modelos CSM e PLS. A seção 3.4 discute a estimação e a inferência paramétrica nesses tipos de modelos e ilustra com um exemplo de estimação da Satisfação dos Consumidores. O capítulo termina na seção 3.5 com uma breve discussão sobre os critérios que devem ser adotados para uma escolha desses modelos para estimar a Satisfação do Consumidor. Por fim é apresentada uma tabela destacando as principais características dos modelos CSM e PLS.

3.2

Efeitos dos Erros de Medidas

Variáveis latentes ou traços latentes, são grandezas que não podem ser medidas diretamente, como as grandezas físicas. Isto implica que a sua magnitude deverá ser inferida indiretamente, através de outras variáveis que podem ser medidas, sendo, por isso, chamadas de indicadores das latentes. Torgerson (1958) chamou este tipo de avaliação de “medição por meio de sensação”. Com base em conhecimentos *a priori* ou em procedimentos mais sofisticados, são atribuídos significados naquilo que se observa baseado nos relacionamentos presumidos entre a variável latente e os seus indicadores. Na verdade, acredita-se (ou espera-se) que esses indicadores sejam correlacionados com a variável latente.

Todavia, os indicadores e as respectivas variáveis latentes não devem ser tratados como sendo a mesma coisa. Na verdade, existem os erros de medidas que normalmente estão presentes, e se forem ignorados, poderão conduzir a resultados enganosos. Uma razão é que as estimativas das correlações entre os indicadores e a respectiva variável latente podem estar enviesadas. Por exemplo, existem alguns tipos de análises de correlações que normalmente são empregados nas pesquisas de Satisfação do Consumidor, tais como: a Análise de Desempenho⁵ (cf. Oliver,

⁵A Análise de Desempenho, também conhecida como *Key Driver Analysis*, avalia os relacionamentos entre os vários fatores com o objetivo de identificar o mais importante deles.

1996) e a Análise de Preferências⁶ (cf. Hauser, 1991), que mostram que os efeitos dos erros das medidas têm implicações práticas.

A Confiabilidade de uma medida é o grau pelo qual a medida está livre dos erros aleatórios (cf. Bollen, 1989). Quando a Confiabilidade de uma medida não é levada em consideração, as estimativas de suas correlações com outras variáveis poderão estar enviesadas e as diferenças nas comparações entre as medidas poderão permanecer ocultas, ou seja, não serem reveladas. Por exemplo, num problema onde se pretende empregar uma regressão linear com o objetivo de avaliar a importância dos atributos da Satisfação do Consumidor, a variável Satisfação será regredida em um conjunto de variáveis preditoras (normalmente variáveis que avaliam o Desempenho Percebido do produto ou serviço adquirido). Como consequência, os erros das medidas nas variáveis preditoras poderão causar distorções (vieses⁷) nos coeficientes estimados da regressão, mesmo se os valores esperados dos erros forem iguais a zero. Este problema, particularmente relevante na análise dos dados da Satisfação, também é bem conhecido na Econometria e é descrito como o problema dos “erros nas variáveis” (cf. Maddala, 1977).

O modelo de medidas permite separar os erros das medidas daquilo que realmente se pretende estimar. Um modelo de medida é um dispositivo que conecta as variáveis observadas, ou indicadores, em uma ou mais variáveis latentes de tal forma que os valores verdadeiros poderão ser estimados separados dos respectivos erros das medidas. Um exemplo típico deste tratamento pode ser visto na Psicometria através do modelo clássico de medida (cf. Lord & Novick, 1968), representado pela equação 3.1.

$$X_{\text{OBSERVADO}} = X_{\text{VERDADEIRO}} + \zeta \quad (3.1)$$

A variável $X_{\text{OBSERVADO}}$ é o valor medido em uma escala contínua para uma simples medida ou observação, a variável $X_{\text{VERDADEIRO}}$ é o valor verdadeiro da variável (i.e. da variável latente) para aquela observação e ζ é o erro da medida, que também é observado. Normalmente assume-se, por hipótese, que $E[\zeta] = 0$ e $\text{COV}(X_{\text{VERDADEIRO}}, \zeta) = 0$. A Confiabilidade da variável $X_{\text{OBSERVADO}}$ é definida

⁶A Análise de Preferências, também conhecida como *Revealed Preference Analysis*, avalia os determinantes do processo de escolha.

⁷Viés ou Bias expressa o erro sistemático ou a tendenciosidade na medida.

como a razão das variâncias de $X_{\text{VERDADIRO}}$ e $X_{\text{OBSERVADO}}$. Segundo Bollen (1989), o erro ζ pode ser proveniente de várias fontes, mas se este erro não for corretamente levado em consideração, os estimadores poderão se tornar viesados (*cf.* Deshon, 1998).

3.3

Aspectos Gerais dos Modelos CSM e PLS

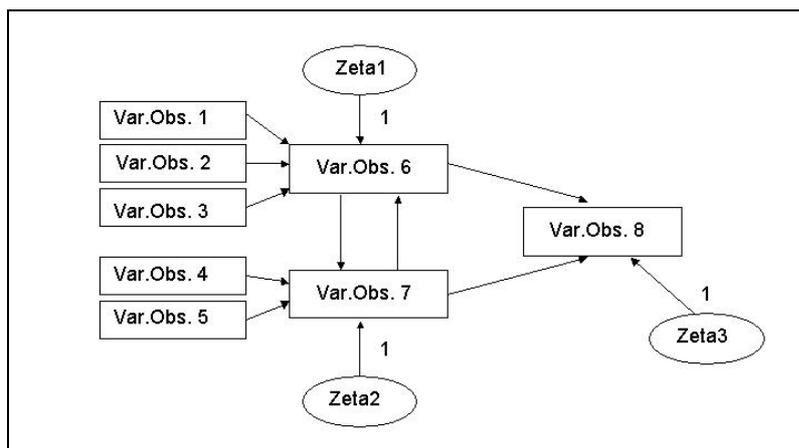
Os modelos de variáveis latentes são aqueles que incluem modelos de medidas para essas variáveis, sendo que nos Modelos de Equações Estruturais as variáveis latentes são representadas através de um esquema de relacionamentos de causas e efeitos, formando um conjunto de ligações diretas e indiretas entre essas variáveis, normalmente chamadas de Caminhos. Essas ligações descrevem as dependências entre as variáveis do modelo, onde cada ligação direta pode ser traduzida através de uma equação. O conjunto de ligações é descrito como um Diagrama de Ligações ou de Caminhos, e estabelece uma forma consistente para descrever modelos complexos em termos de relações entre as variáveis do modelo.

Existe uma variedade de modelos de variáveis latentes sendo que uma de suas diferenças básicas reside no tipo de escala das variáveis do modelo (discreta ou contínua). Na maioria das aplicações as escalas são contínuas, porém existem casos onde elas são discretas. Exemplos de modelos com variáveis discretas são os que utilizam variáveis latentes e observadas do tipo nominal, além dos modelos da Teoria de Resposta ao Item (TRI), cujas variáveis observadas são do tipo nominal ou ordinal e as variáveis latentes são do tipo ordinal ou intervalar. Heinen (1996), oferece um resumo de modelos de variáveis latentes com escalas discretas. Baumgartner & Homburg (1996), oferece uma revisão de 1977 a 1994, dos Modelos de Equações Estruturais aplicados no marketing e na pesquisa do consumidor, publicados nos quatro maiores periódicos do marketing: *Journal of Marketing*, *Journal of Marketing Research*, *International Journal of Research e Journal of Consumer Research*.

Os Modelos de Equações Estruturais também se diferem em termos conceituais. Por exemplo, existem os modelos baseados em estruturas de covariâncias, conhecidos na literatura como CSM (ou SEM – *Structural Equation*

Modeling), e os modelos baseados em mínimos quadrados parciais, conhecidos na literatura como PLS. Estes modelos possuem características particulares para a modelagem e estimação de variáveis latentes. No entanto, ambos consideram modelos de medidas para variáveis latentes contínuas e ambos tem a capacidade de estimar modelos provenientes de sistemas de equações. Os sistemas de equações nos auxiliam nos seguintes casos: (1) quando existirem variáveis intermediárias no modelo; (2) quando existirem causas diretas entre variáveis endógenas e; (3) quando existirem causas recíprocas entre variáveis. Diagramas de ligações que ilustram estas características podem ser vistos nas Figuras 3.1 e 3.2.

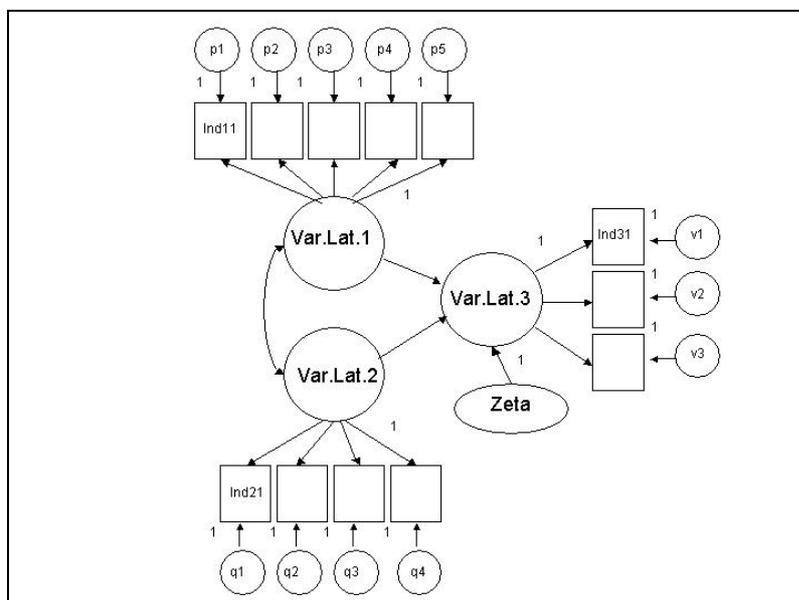
Na Figura 3.1, todas as variáveis do modelo são observadas, ou seja, não existe um modelo de medidas presente. As ligações entre as variáveis mostram as suas dependências, por exemplo, a variável observada nº 7 depende de três variáveis observadas: as de nº 4, 5 e 6. As elipses $Zeta_1$, $Zeta_2$, $Zeta_3$ representam erros de regressão. Estes valores são inferidos, isto é, são variáveis latentes. O valor 1 (um) nas ligações das variáveis Zeta representa uma restrição que faz com que as escalas das medidas dessas variáveis sejam iguais às escalas das respectivas variáveis dependentes na equação de regressão correspondente. Este modelo é chamado de “não-recursivo” porque as variáveis de nº 6 e 7 são causas recíprocas uma da outra (também conhecido como *loop*).



Fonte: Própria

Figura 3.1 – Modelo de Equações Estruturais do Tipo não-Recursivo.

O modelo da Figura 3.2 é composto por variáveis latentes e observadas sendo, portanto, um modelo híbrido. Neste caso, existe um modelo de medidas para cada uma das variáveis latentes, cujos indicadores estão representados pelos quadrados com as setas saindo das respectivas variáveis latentes. Associados aos indicadores estão os erros de medidas, representados pelos círculos, já que os mesmos também não são observados. A linha curva que liga as variáveis de nº 1 e 2 representa a covariância entre elas, ou seja, uma associação não direcional, diferente das demais ligações. Este modelo é essencialmente um modelo de regressão de variáveis latentes com duas variáveis predictoras correlacionadas. Este modelo é chamado de “recursivo” porque não existem causas recíprocas entre as variáveis latentes.



Fonte: Própria

Figura 3.2 – Modelo de Equações Estruturais do Tipo Recursivo

As diferenças entre os modelos CSM e PLS são conceituais, embora ambos os modelos possam ser empregados para estimar a Satisfação do Consumidor. Uma importante característica desses modelos é que ambos foram desenvolvidos para estimar relacionamentos lineares entre variáveis latentes. A estimação de interações ou não-linearidades entre variáveis latentes é bem mais problemática - avanços nessa área utilizando os modelos CSM podem ser vistos em Schumacker

& Marcoulides (1998) e em Arminger & Muthén (1998). Avanços com a metodologia PLS podem ser vistos em Chin, Marcolin & Newsted (1996).

O PLS foi desenvolvido para estimar somente modelos recursivos, tal como o modelo da Figura 3.2. O modelo CSM enfatiza a estimação e os testes dos parâmetros do modelo como uma forma para desenvolver e/ou avaliar teorias, através da explicação das variâncias e covariâncias entre as variáveis observadas. O PLS, por outro lado, foi desenvolvido para maximizar a acurácia preditiva do modelo (cf. Jöreskog e Wold, 1982), enquanto oferece flexibilidade para a modelagem exploratória. O PLS não necessita de hipóteses sobre a distribuição dos dados, como exige os modelos CSM, sendo chamado por isso de “*soft modeling*”. O PLS foi originalmente visto como um complemento dos modelos CSM, para ajustar Modelos de Equações Estruturais.

O ajuste dos modelos CSM envolve a minimização das diferenças entre as matrizes de variâncias e covariâncias observadas e previstas. Um algoritmo de ajuste iterativo é utilizado para estimar os parâmetros simultaneamente. O algoritmo começa a partir de um conjunto inicial de valores dos parâmetros que estão sendo estimados e então os ajusta após sucessivas iterações, até que uma medida escalar de discrepância entre o observado e o previsto seja minimizada. O método mais utilizado nesta estimação é o método da Máxima Verossimilhança. Entretanto, este método exige que os dados sejam distribuídos segundo uma Normal Multivariada e que as observações sejam independentes. Outros métodos empregados na estimação são menos restritivos, tais como: ULS (*Unweighted Least Square* - Mínimos Quadrados não-Ponderados), GLS (*Generalized Least Square* - Mínimos Quadrados Generalizados) e o critério ADF de Browne (*asymptotically distribution-free*) para o ajuste assintótico livre de distribuição – este critério exige apenas que os dados observados sejam contínuos.

Os modelos CSM devem ser primeiramente identificados⁸. Em geral, esses modelos são não-identificados, havendo, portanto, a necessidade de se impor restrições aos seus parâmetros. Essas restrições servem para reduzir o número de parâmetros a serem estimados. Um modelo será não-identificado quando não se conseguir obter soluções únicas para todos os seus parâmetros. Grosso modo, isto

⁸O problema da identificação consiste em verificar se o vetor paramétrico θ será unicamente determinado através de Σ (matriz de variância e covariância dos dados observados). Este problema deve ser resolvido antes da estimação dos parâmetros.

significa que existem mais parâmetros a serem estimados do que elementos distintos e conhecidos na matriz de variância e covariância dos dados observados. Existem heurísticas (testes empíricos) que são normalmente empregadas para determinar se um modelo CSM é ou não identificado.

A estimação das variáveis latentes (escores) nos modelos CSM segue os mesmos princípios da estimação dos Fatores Comuns na Análise Confirmatória de Fatores. No entanto, algumas restrições devem ser utilizadas para estabelecer as escalas de medidas das variáveis latentes. Os escores não são estimados diretamente como no PLS. Eles são estimados após o ajuste do modelo, através de metodologias baseadas em regressões múltiplas, cujos valores obtidos dependem do método utilizado – alguns autores alertam para o uso indiscriminado desses métodos, principalmente na Análise Exploratória de Fatores (*cf.* Lastovicka & Thamodaran, 1991).

A estimação no PLS é realizada iterando sobre uma seqüência de partes do modelo com o objetivo de minimizar a variância residual associada a todas as variáveis endógenas. Essas partes do modelo constituem-se nos modelos de medidas para as variáveis latentes, chamados de “blocos”, e no conjunto de relacionamentos que conectam as variáveis latentes. Estas partes são chamadas respectivamente de “modelos externos” (*outer*) e de “modelos internos” (*inner*), e são análogas às partes das medidas e estrutural dos modelos CSM. Os dados analisados através do PLS são tipicamente matrizes de correlações. Os algoritmos utilizados são diferentes espécies de rotinas OLS (*Ordinary Least Square* – Mínimos Quadrados Ordinários), dependendo do tipo de parâmetro que está sendo estimado no momento.

As variáveis observadas no PLS podem se relacionar com as variáveis latentes de duas formas distintas: como indicadores Reflexivos, significando que as ligações (setas) apontam da variável latente para a observada, ou como indicadores Formativos, as ligações (setas) apontam da variável observada para a latente. Os modelos CSM acomodam somente indicadores reflexivos.

As variáveis latentes no PLS têm as mesmas características das Principais Componentes, enquanto que nos modelos CSM, elas têm as mesmas características dos Fatores. No PLS, cada variável latente é definida como uma combinação linear ponderada das variáveis observadas, não existe o problema de identificação dos parâmetros e também não são necessárias hipóteses sobre a

distribuição das variáveis observadas, nem sobre as distribuições dos resíduos e nem é exigido a independência das observações.

3.4

Estimação e Inferência Paramétrica nos Modelos CSM e PLS

De acordo com Jöreskog & Sörbom (1982) e Bollen (1989), as estimativas de máxima verossimilhança dos parâmetros dos modelos CSM, assumindo a hipótese de que o modelo esteja correto e que o tamanho da amostra seja suficientemente grande, são não-enviesadas e de mínima variância. Estimativas de Máxima Verossimilhança possuem várias características desejáveis sob condições de regularidade e são relativamente robustas contra violações dessas hipóteses.

A consistência das estimativas do PLS depende do tamanho da amostra e do número de indicadores definidos para cada variável latente. Do contrário, as estimativas das cargas (*loadings*) poderão ficar enviesadas. De acordo com Lohmöller (1989) e Chin (1998), ainda não existe uma solução adequada capaz de estimar o tamanho dos vieses nas estimativas do PLS. Lohmöller (1989) analisou este problema através de modelos com uma ou duas variáveis latentes, considerando correlações iguais entre os indicadores. Os vieses nas estimativas do PLS foram definidos como sendo a razão das estimativas do PLS em relação às estimativas de Máxima Verossimilhança. Contudo, esta solução é útil somente se o modelo que for utilizado para se obter as estimativas de Máxima Verossimilhança estiver de fato correto. Chin (1998) argumenta que, em geral, os efeitos dessa inconsistência serão de tal forma que as estimativas das Cargas tenderão a ficar enviesadas para cima e as relações das variáveis latentes (parâmetros de ligação) tenderão a ficar enviesadas para baixo. Além disso, vieses nas estimativas do PLS farão com que parâmetros diferentes pareçam os mesmos e que parâmetros verdadeiramente iguais pareçam diferentes.

Um estudo sobre os vieses nas estimativas do PLS pode ser visto no trabalho de Ryan e Rayner (1998). Neste trabalho, as estimativas do PLS e do modelo CSM foram comparadas através de conjuntos de dados simulados. Várias amostras de dados de tamanhos diferentes foram geradas a partir de modelos não-recursivos, variando-se os valores dos parâmetros do modelo. Os modelos analisados eram constituídos de quatro variáveis latentes exógenas e uma

endógena. Os resultados indicaram que as estimativas dos parâmetros produzidas pelo PLS ficaram em média mais distantes dos valores verdadeiros do que as estimativas do CSM, no entanto o RMSE (*root mean square error* - uma espécie de desvio padrão generalizado) do PLS foi em média menor do que o do CSM.

O PLS é uma metodologia mais flexível do que o CSM porque ele não necessita de hipóteses sobre a distribuição das variáveis observadas. Logo, de certa forma, podem ser utilizadas quaisquer variáveis indicadoras para medir as variáveis latentes. Entretanto, não existem testes estatísticos disponíveis no PLS para avaliar a significância dos parâmetros, nem do ajuste do modelo e nem das diferenças entre modelos. Inferências para estas estatísticas são realizadas através de dois métodos: *Jackknife* ou *bootstrapping*.

Os modelos CSM fornecem estimativas dos erros padrões dos parâmetros estimados além de várias estatísticas para verificar o ajuste do modelo aos dados. Comparações estatísticas de modelos aninhados e a capacidade de testar restrições lineares e não-lineares nos valores dos parâmetros também são tratadas pelo CSM. No entanto, isto requer a adoção de hipóteses sobre a distribuição dos dados. Todavia, mesmo se essas hipóteses não forem sustentáveis, inferências ainda poderão ser realizadas via *bootstrapping*.

Os modelos CSM ainda podem ser aplicados em dados provenientes de múltiplos grupos simultaneamente, podendo-se testar as diferenças entre os modelos. Nestes modelos também é permitido estimar valores médios e interceptos das variáveis latentes. Trabalhos recentes de modelagem de dados multi-nível (hierárquica ou por agrupamentos) utilizando os modelos CSM, incluíram estimativas de parâmetros a nível individual (*cf.* Kaplan & Elliott, 1997 e Muthén, 1994). Muthén (1998) forneceu uma especificação no *software* LISCOMP para estimar modelos CSM com variáveis observadas medidas em escalas discretas. Seu trabalho mais recente estende a metodologia do LISCOMP para acomodar modelos de mistura finta.

Para fins de ilustração das diferenças básicas entre os modelos CSM e PLS na estimação da Satisfação do Consumidor, segue um problema de regressão linear bivariada entre duas variáveis latentes. O modelo pode ser visto na Figura 3.3 e os resultados da estimação na Tabela 3.1 (*cf.* Bacon, 1998). Os dados são provenientes de uma pesquisa com 200 consumidores de uma empresa prestadora de serviços automotivos. Os serviços foram avaliados através de quatro atributos

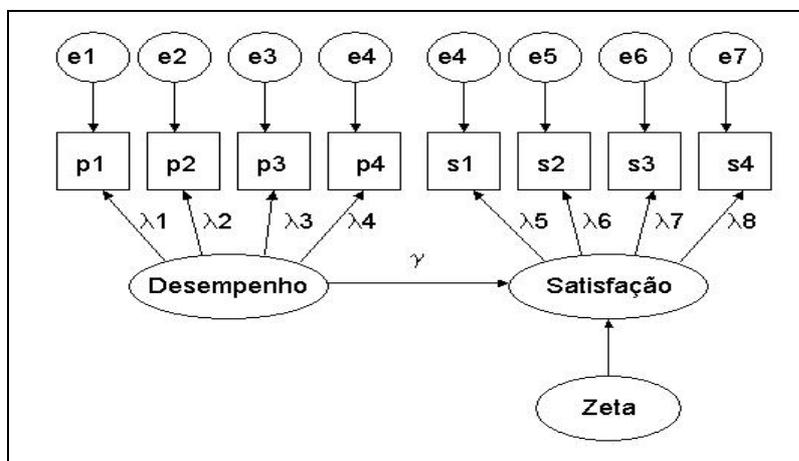
de Desempenho e de Satisfação. Os atributos do Desempenho foram: cortesia, acessibilidade, tempo de conclusão do serviço e limpeza do local de serviço, todos avaliados através de uma escala unipolar de sete pontos. Os atributos da Satisfação foram: satisfação geral com o serviço, satisfação com a visita, satisfação com outros tipos de serviços automotivos fornecidos e um julgamento de confiança sobre que tipo de serviço seria mais satisfatório para uma próxima visita. As escalas de avaliação foram do tipo bipolar com pontos extremos variando de muito insatisfeito a muito satisfeito, exceto para o último atributo que se tratou de uma nota direta.

No diagrama da Figura 3.3, os indicadores de Desempenho estão nomeados de p1 a p4 e os da Satisfação de s1 a s4. As escalas de medidas das variáveis latentes Desempenho e Satisfação foram restringidas pelas cargas Fatoriais de p1 e s1, isto é, iguais a 1 (um). Como mencionado anteriormente, esta técnica foi empregada para estabelecer uma escala de medida para as variáveis latentes. Neste caso, as escalas das latentes Desempenho e Satisfação serão as mesmas das observadas p1 e s1, respectivamente. Uma restrição análoga foi utilizada para fixar a escala de medida da variável Zeta (resíduo da regressão).

O método de estimação utilizado no modelo CSM foi o da Máxima Verossimilhança, devido à constatação prévia de que a distribuição dos dados observados não fugia da multi-normalidade. O software empregado foi o LISREL e os dados observados foram informados na forma de uma matriz de variância e covariância das oito variáveis observadas, ou seja, uma matriz 8x8. A estatística de ajuste do modelo χ^2 foi de 23,09 com 19 graus de liberdade. O valor do RMSE foi de 0,064 (na prática, valores entre 0,05 e 0,08 indicam um ajuste moderado). Os resíduos também foram analisados através de um *qq-plot* (não exibido aqui) que indicou uma aproximação da distribuição normal. As outras medidas de ajuste não foram consideradas.

Na Tabela 3.1, observa-se que a correlação estimada entre as variáveis latentes foi de $\gamma = 0.85$ e que as cargas padronizadas variaram de 0,51 a 0,80. Com base nos valores estimados dos erros-padrões das Cargas (não exibidos nesta tabela), pode-se afirmar que as Cargas são diferentes de zero. O valor padronizado da Carga pode ser interpretado como a raiz quadrada do Coeficiente de Confiabilidade da respectiva variável observada associada. Neste exemplo,

somente 26% da variação em “p4” está associado com a latente Desempenho, assumindo que o modelo esteja correto.



Fonte: cf. Bacon (1998).

Figura 3.3 – Modelo Ilustrativo da Satisfação

Tabela 3.1 – Parâmetros Estimados do Modelo (Valores Padronizados)

Parâmetros Estimados		
Parâmetro	SEM	PLS
λ_1	0,76	0,82
λ_2	0,68	0,79
λ_3	0,72	0,81
λ_4	0,51	0,63
λ_5	0,80	0,85
λ_6	0,66	0,76
λ_7	0,68	0,77
λ_8	0,66	0,76
γ	0,85	0,68

A Tabela 3.1 também mostra os parâmetros obtidos através do PLS utilizando o programa LVPLS (Lohmöller). As estatísticas de ajuste indicaram um razoável nível de ajuste do modelo, cujo valor de $R^2 = 0,58$. A raiz da Covariância Média Quadrática (RMSCOV) entre os resíduos da variável latente e observada foi de 0,04. Tal como o RMSE, quanto menor este valor melhor o ajuste.

Comparando as estimativas dos parâmetros através das duas metodologias, percebe-se que as Cargas estimadas pelo PLS parecem ser melhores do que as do CSM. A outra diferença mostra que a relação entre as variáveis latentes é menor no PLS do que no CSM. De acordo com Chin (1996), essas diferenças são comuns nas comparações entre as metodologias, entretanto, não se sabe qual conjunto de estimativas está mais próxima dos parâmetros verdadeiros. O mesmo padrão de diferenças também pode surgir devido aos vieses nas estimativas do PLS.

3.5

Cr terios de Escolha Quanto aos Modelos CSM e PLS para Estimar a Satisfa o do Consumidor

As caracter sticas particulares de cada um dos deles servem para subsidiar a escolha do modelo para estimar a Satisfa o do Consumidor. Se o modelo a ser empregado   do tipo n o-recursivo, ent o o modelo CSM seria a  nica escolha. Se o interesse   obter previs es mais acuradas dos escores das vari veis latentes, ent o o PLS pode ser o modelo mais indicado, j  que o mesmo fornece um m todo direto para se calcular os escores. Se o interesse est  na compara o das Cargas ou das rela es entre as vari veis latentes, o modelo CSM poder  ser a melhor escolha, assumindo que as hip teses necess rias sejam sustent veis ou que seja poss vel obter estimativas via *bootstrap*. No PLS, as estimativas das Cargas e das rela es entre as vari veis latentes ser o sempre enviesadas.

Fatores que favorecem a escolha dos modelos CSM:

- a) Se o pesquisador desejar fazer uma infer ncia estat stica tradicional.
- b) Se o pesquisador tem s lidos conhecimentos te ricos sobre os relacionamentos causais que ele pretende quantificar, justificando inclusive a compara o de modelos.
- c) A constru o da validade das medidas   altamente relevante.
- d) O pesquisador pretende avaliar modelos para diferentes grupos de amostras.

A grande vantagem do PLS é a sua capacidade de acomodar indicadores, independente do seu tipo de medida. Esta característica do PLS pode torná-lo mais conveniente para ser aplicado em dados de Satisfação que já foram coletados ou que ainda serão coletados sem considerar as questões apropriadas dos modelos CSM. De acordo com Rigdon (1998), existe um alto risco de falhas na aplicação de um método analítico quando o seu uso não foi previamente considerado. Independente dos recentes progressos que vem fazendo do CSM um método mais flexível, ele é ainda um método de análise confirmatório. No caso de utilizar em conjunto os modelos CSM e PLS, ainda será necessário algum tipo de julgamento. Ambos são procedimentos complexos de modelagem e ainda existe pouco consenso entre os especialistas sobre qual é o melhor método para ser aplicado num problema prático (i.e., não trivial).

Nenhum dos modelos será de grande utilidade se os dados não contiverem medidas de variáveis importantes ou se os modelos não foram especificados de forma correta. Oliver (1996) notou que parece existir um descuido geral na aplicação das regras e na importância das variáveis do processo na maioria das pesquisas de Satisfação do Consumidor, e apontou para uma prática comum de serem coletados somente dados referentes ao Desempenho Percebido dos atributos dos produtos. Logo, variáveis importantes não devem ser ignoradas sob pena de o modelo ficar mal especificado. Por exemplo, existe evidências empíricas de que a Desconfirmação parcial ou total das Expectativas intermedia os efeitos do Desempenho Percebido na Satisfação. Logo, se esta variável for omitida do modelo então os parâmetros estimados poderão refletir de forma errônea o verdadeiro impacto das variações do Desempenho percebido na Satisfação. Portanto, conclui-se que a teoria e o prévio conhecimento do assunto devem ser utilizados para guiar tanto a coleta dos dados quanto à especificação do modelo.

Ambos os modelos podem acomodar multicolinearidade⁹ nas variáveis preditoras da Satisfação. A presença da multicolinearidade distorce o verdadeiro valor das relações entre as variáveis do modelo. Normalmente utiliza-se uma variável latente para representar um grupo de variáveis observadas interrelacionadas. Nos modelos CSM pode-se ainda estimar as covariâncias entre

⁹Multicolinearidade é definida como a dependência linear entre duas ou mais variáveis. A presença da multicolinearidade dificulta a separação dos efeitos preditivos das variáveis, resultando em estimativas instáveis.

as variáveis latentes ou entre as variáveis indicadoras ou ainda utilizar um método genérico de Fatores.

O PLS permite a modelagem de variáveis multicolineares, reunidas como variáveis preditoras interdependentes de uma ou mais variáveis latentes, entretanto elas devem ser modeladas como indicadores reflexivos. Se forem modeladas como indicadores formativos, poderão ocorrer problemas, pois não existe uma forma de avaliar a interdependência entre essas variáveis, causando instabilidade na estimação. Além disso, o PLS não permite a estimação de correlações indiretas, isto é, associações entre variáveis que não tem uma direção.

Resumindo, os modelos CSM e PLS oferecem vantagens e desvantagens para se analisar dados da Satisfação do Consumidor. Contudo, ambos fornecem meios para reconhecer erros de medidas e controlar seus efeitos nas estimativas de outras quantidades. Ambos permitem estimar sistemas de relacionamentos, provendo uma forma de expressar as múltiplas dependências apontadas na literatura da Satisfação do Consumidor. Ambos permitem modelar multicolinearidade de forma que os seus efeitos possam ser atenuados na estimação.

Por fim cabe destacar que o PLS foi desenvolvido para aplicações que dispõem de poucas informações teóricas sobre o problema, ao contrário dos modelos CSM que foram desenvolvidos principalmente para se avaliar teorias. Portanto, mantendo a filosofia de que nenhum modelo é totalmente correto, então, se for possível, pode ser instrutivo aplicar ambos os modelos, de forma que as discrepâncias encontradas entre os resultados possam ser examinadas e, talvez, até reconciliadas.

A Tabela 3.2 resume as principais diferenças entre os modelos CSM e PLS.

Tabela 3.2 – Características dos Modelos CSM e PLS

CARACTERÍSTICA	MODELO CSM	PLS
Hipóteses de distribuição das variáveis observadas	Variáveis observadas são Normais Multivariadas na estimação de Máxima Verossimilhança. Observações devem ser independentes	Nenhuma
Tipo de modelos que podem ser ajustados	Recursivos e não recursivos	Recursivos
Tipos de variáveis observadas	Contínuas ou discretas ordinais (utilizando correlações policóricas como entradas, assumindo robustez ou utilizando a especificação do LISCOMP)	Contínuas, discretas ordinais ou não ordenadas
Tipos de variáveis latentes que podem ser modeladas	Contínuas	Contínuas
Tipos de indicadores	Reflexivos	Reflexivos ou formativos
Identificação dos parâmetros	Deve ser considerada. Existem heurísticas disponíveis para alguns tipos de modelos. Para outros são requeridas procedimentos empíricos	Não é considerado no PLS padrão
Quantidade de indicadores por fator	No mínimo 1 (um) se houver restrições para o erro do indicador. De outro modo, a quantidade dependerá das condições para a identificação dos parâmetros	Um ou mais
Fatores por indicador	Uma variável observada poderá indicar mais de uma variável latente	Variáveis observadas podem indicar somente uma variável latente
Correlações entre variáveis latentes podem ser estimadas indiretamente	Sim	Não
Correlações entre erros de medidas podem ser modelados	Sim	Não
Estimação das médias e dos interceptos das variáveis latentes	Sim	Não
Tipo de algoritmo de ajuste	Estimação simultânea dos parâmetros através da minimização das discrepâncias entre a matriz de variâncias e covariâncias observadas e previstas. Métodos de informação completa	Procedimento iterativo de múltiplo-estágio. O modelo é dividido em blocos e os parâmetros são estimados separadamente. Método de informação limitada
Consistência dos estimadores	Consistentes, dado a certeza do modelo e da sustentabilidade das hipóteses	As estimativas se tornam consistentes em função do tamanho da amostra e do número de indicadores por variável latente
Testes estatísticos disponíveis para as estimativas	Disponíveis e válidas desde que as hipóteses do modelo sejam sustentáveis. De outro modo, as inferências são via <i>bootstrapping</i>	Inferências requerem <i>Jackknifing</i> ou <i>bootstrapping</i>