What Does Treatment Of Depression Really Cost?

Although treating a patient with depression is expensive, the bulk of the cost is for care of accompanying illnesses.

by Thomas W. Croghan, Robert L. Obenchain, and William E. Crown

This DataWatch presents estimates of the health care charges for adults who are diagnosed and treated for depression in primary care. More than nine out of ten of these adults sought care for at least one nondepressive illness during the year following treatment initiation. On average, these conditions accounted for more than 70 percent of the total charges. Attempts to manage the costs of caring for depressed persons must consider the impact of nondepressive illness.

Depressive illness is among the most common disorders seen in primary care and is associated with high rates of chronic disability and other functional impairment. Depressed patients frequently suffer from other mental and general medical disorders or may have somatic symptoms that are easily attributed to physical illness.[1] Alternatively, medical conditions may produce symptoms resembling depression or may precipitate a depressive episode.[2] These considerations make depression very difficult to recognize in the general medical setting; in fact, only half of all patients with depression are correctly diagnosed.[3]

Depressed patients also are said to be high users of medical resources, in both mental health specialty and general medical settings.[4] Those with depressive illness consume two to four times more general medical resources than do patients without mental illness. In spite of years of study, however, it is unclear whether appropriate treatment of depressive or other mental illness is associated with more than very modest offsets in treatment costs associated with somatic illness.[5] However, care for depressive illness has evolved over the past decade. New innovations such as selective serotonin reuptake inhibitors (SSRIs) and managed care appear to be associated with increases in use of antidepressants and reduc-

Thomas Croghan is a health economics research physician in the Division of Health Services and Policy Research at Eli Lilly in Indianapolis and an assistant professor in the School of Medicine at Indiana University. Robert Obenchain is a research scientist in the Department of Statistical and Mathematical Sciences at Lilly. William Crown is director of outcomes research and econometrics at the MEDSTAT Group in Cambridge, Massachusetts.

tions in the cost of high-quality care.[6] Could these new treatments be associated with similar reductions in the cost of treating comorbid physical illness?

This DataWatch is the first analysis of the costs of various medical and psychiatric comorbidities in patients with depressive illness as it is commonly treated in primary care settings. We focus on a group of patients who received antidepressants as their initial treatment. Our simultaneous analysis of the cost of multiple diseases allows assessment of the relative economic burden imposed by specific categories of illness and has important implications for cost-effectiveness studies.

Data Sources And Methods

We analyzed 3,439 patients whose medical and pharmacy claims are included in the MarketScan database (MEDSTAT Group, Ann Arbor, Michigan) for 1990–1994. MarketScan contains comprehensive information of the medical and pharmaceutical insurance claims of approximately 700,000 persons during these years. These benefits were offered by twenty employers from around the United States and included a mix of indemnity and managed care plans. In addition to the charge data used as the outcome variables of interest, these files contain information regarding the demographic and medical characteristics of each patient, including age and sex, diagnoses, procedures, residence, location and type of service delivery, and information regarding hospital and ancillary medical care use.

Episodes of care. The main objective of this study was to estimate the expected expenditures attributable to depressive illness as opposed to those comorbid conditions that complicate depression. Our approach was to construct episodes of care for that subset of depressed patients who were initially identified and treated in primary care settings. By analyzing this subset of patients, we restricted to some degree the heterogeneity of the sample and thus the generalizability of the study. However, the majority of depressed patients initially receive treatment in a primary care setting, and this group is the major focus of recent guidelines.[7]

We began constructing episodes of care by identifying an index prescription for an antidepressant that was temporally linked to a claim for a depressive illness by no more than thirty days. To identify new episodes of treatment with antidepressants, we identified a six-month pretreatment period free of any indication of diagnosis or treatment for depressive illness. Indicators of comorbid conditions could occur at any time during the one-year treatment period following the index prescription. Total health care charges, including charges for physician visits, other outpatient visits, laboratory tests,

radiological tests, hospital stays, and prescriptions, were collected over the twelve-month period after the index prescription for each patient.

Multivariate models. For this analysis we estimated several sets of multiple regression expenditure models. These estimates were adjusted for age and sex, type of depressive illness, and indicators of the presence of comorbid conditions. Types of depressive illness included major depressive disorder, first episode (*International Classification of Diseases*, Ninth Revision, Clinical Modification [ICD-9-CM] code 296.2x); major depressive disorder (MDD), recurrent (296.3x); neurotic depression (300.4); brief depressive reaction (309.0); prolonged depressive reaction (309.1); and depressive disorder, not elsewhere classified (NEC) (311). Comorbid conditions, collected at the level of ICD-9-CM codes, were classified according to major diagnostic categories (MDCs) and entered in the expenditure models.

We had observed in prior research that although total expenditures might be similar, various classes of antidepressant treatment resulted in differential distribution of service use. For example, we had observed that fluoxetine users have higher expenditures for ambulatory care but lower inpatient expenditures than do users of other antidepressants. We thus were curious to see whether this difference had something to do with the conditions for which patients might receive care. For example, one explanation for reduced hospital charges might be a reduced need for treatment complications of cardiovascular diseases. We thus split our sample into SSRI users (fluoxetine, sertraline, and paroxetine) and users of one of several older tricyclic antidepressants (TCAs), with appropriate adjustments for differences in antidepressant selection.[8] The results are reported as the predicted expenditure for the year following initiation of antidepressant treatment.[9]

Results

Characteristics of patients with depression. The average age of the patients was 40.9 years, and there were many more women than men (2,488 versus 951) (Exhibit 1). The most common depression diagnosis was depression, NEC, followed by neurotic depression. Major depressive episode was relatively infrequent, even though this diagnosis is the focus of major guidelines.[10] Patients taking SSRIs were younger than TCA users.

More than 95 percent of patients had an indicator for at least one nondepression diagnosis at some time during the twelve-month follow-up period. Nondepressive mental illness was the most frequent (94.5 percent). Disorders of the ear, nose, mouth, and throat; the musculoskeletal system; and the skin also were common. TCA

Age (years)			
18–30	15% (25.0)[a]	16% (25.0)	11% (25.1)
31–44	50 (38.4)	51 (38.4)	48 (38.4)
45–59	31 (50.4	30 (50.1)	34 (51.1)
60–64	4 (61.8)	3 (61.8)	7 (61.9)
Total	100 (40.9)	100 (40.4)	100 (42.2)
Sex			
Female	72.3%	73.2%	70.4%
Male	27.7	26.8	29.6
Depression diagnosis			
MDD, single episode	21.0	20.4	22.5
MDD, recurrent episode	12.9	12.4	14.0
Neurotic depression	27.8	28.3	26.7
Brief depressive reaction	7.1	6.2	9.2
Prolonged depressive reaction	0.8	0.7	1.1
Depressive disorder, NEC	30.3	31.9	26.5
Disorder or condition (MDC)[b]			
Any disorder or condition	95.8	96.0	95.5
Neurological (1)	23.2	21.4	27.4[c]
Eye (2)	10.2	9.8	11.2
Ears, nose, mouth, throat (3)	49.9	50.5	48.6
Respiratory (4)	32.2	32.4	31.6
Circulatory (5)	35.7	35.0	37.3
Digestive (6)	32.3	31.3	34.8
Hepatobiliary (7)	6.5	6.4	6.5
Musculoskeletal (8)	53.5	53.8	52.8
Skin, subcutaneous tissue (9)	50.7	51.0	50.2
Endocrine (10)	35.1	35.1	35.2
Kidney and urinary tract (11)	21.5	20.5	23.6
Male reproductive system (12)	4.4	4.0	5.1
Female reproductive system (13)	32.2	32.5	31.5
Pregnancy/childbirth (14)	2.0	2.1	1.8
Newborns (15)	0.4	0.4	0.5
Blood/immunological (16)	16.0	15.2	17.9
Myeloproliferative diseases (17)	4.7	4.7	4.5
Infections and parasitic (18)	12.0	11.2	13.8
Mental illness (nondepression) (19)	94.5	94.2	95.3
Alcohol and drug use (20)	0.0	0.0	0.0
Injury, poisonings, toxic effects of drugs (21)	14.5	14.5	14.3
Burns (22)	0.4	0.3	0.6
Factors influencing health (23)	45.5	45.9	44.4
Multiple significant trauma (24)	0.0	0.0	0.0
HIV infections (25)	0.2	0.3	0.0

SOURCE: Authors' analysis of the MarketScan® database (MEDSTAT Group, Ann Arbor, Michigan), 1990–1994.
NOTES: SSRI is selective serotonin reuptake inhibitor. TCA is tricyclic antidepressant. MDD is major depressive disorder. NEC is not elsewhere classified. MDC is major diagnostic category. HIV is human immunodeficiency virus.
[a] Mean values are in parentheses.
[b] Category numbers are in parentheses.
[c] $p < 0.01$ relative to SSRIs.

users were more likely to have an indicator for a neurological disorder than were SSRI users. There was a trend toward TCA use in patients with musculoskeletal, kidney, and blood disorders and with certain infections.

The most common nondepressive mental illnesses were anxiety and adjustment disorders (Exhibit 2).[11] Nonarticular rheumatic disorders such as fibromyositis and lumbago were common musculoskeletal disorders. In addition, neck pain syndromes were common neurological conditions, which further emphasizes the degree of musculoskeletal pain experienced by patients with depression. The most common cutaneous disorders were diseases of the breast, in-

[a]		
Neurological conditions (1)	784.0	Headache
	346.9	Migraine
	723.4	Brachial neuritis or radiculitis, unspecified
	723.2	Cervicocranial syndrome
	780.3	Convulsions
Circulatory disorders (5)	401.9	Essential hypertension, unspecified
	786.5	Chest pain, unspecified
	414	Coronary atherosclerosis
	427.9	Cardiac dysrhythmia, unspecified
	401.1	Benign hypertension
Musculoskeletal disorders (8)	724.00	Spinal stenosis
	724.2	Lumbago
	729.1	Myalgia and myositis, unspecified (fibromyositis, unspecified)
	710.0	Systemic lupus erythematosus
	716.9	Arthropathy, unspecified
Skin, subcutaneous tissue (9)	610.1	Diffuse cystic mastopathy
	174.9	Malignant neoplasm, breast
	702	Other dermatoses
	706.2	Sebaceous cyst
	696.1	Other psoriasis
Endocrine disorders (10)	250.00	Diabetes mellitus without complication
	244.9	Unspecified hypothyroidism
	272.4	Unspecified hyperlipidemia
	259.9	Unspecified endocrine disorder
	272.0	Pure hypercholesterolemia
Mental illness (nondepression) (19)	300.00	Unspecified anxiety states
	309.28	Adjustment reaction with mixed emotional features
	309.24	Adjustment reaction with anxious mood
	309.9	Unspecified adjustment reaction
	309.89	Other adjustment reaction

SOURCE: Authors' analysis of the MarketScan® database (MEDSTAT Group, Ann Arbor, Michigan), 1990–1994.
NOTE: MDC is major diagnostic category.
[a] Category numbers are in parentheses.

cluding breast cancer, which is classified as a cutaneous disorder in the MDC system.

Cost impact of comorbid conditions. Predicted expenditures for patients without comorbid medical or psychiatric conditions are much lower than they are for patients with one or more comorbid conditions (Exhibit 3). To be more specific, on average, treatment of depressive illness alone represents only about 28 percent of the total charges for the depressed patients with comorbid conditions in this population ($2,279 versus $8,037). There were no significant differences between treatment groups.

Age	$ 8	$ 13	$ −2
Male gender	−127	69	−382
MDD, single episode	650[a]	591	874
MDD, recurrent	889[a]	955	437
Neurotic depression	510[a]	385	398
Brief depressive episode	−132	−261	−88
Prolonged depressive episode	2,305[a]	1,590	2,449
Depressive disorder, NEC	−783[a]	−569	−1,010
Neurological	2,194[a]	1,978	2,316
Eye	632[a]	550	752
Ears, nose, mouth, throat	614[a]	603	453
Respiratory	1,110[a]	1,095	1,070
Circulatory	1,412[a]	1,327	1,412
Digestive	1,316[a]	1,400	939
Hepatobiliary	742[a]	638	1,109
Musculoskeletal	1,505[a]	1,474	1,276
Skin, subcutaneous tissue	795[a]	715	767
Endocrine	682[a]	614	713
Kidney and urinary tract	579[a]	746	363
Male reproductive system	1,177[a]	1,545	442
Female reproductive system	779[a]	677	678
Pregnancy/childbirth	1,697[a]	1,197	2,960
Newborns	1,625	2,417	−149
Blood/immunological	611[a]	763	277
Myeloproliferative diseases	1,495[a]	1,361	1,792
Infections and parasitic	551[a]	418	681
Mental illness (nondepression)	1,629[a]	1,392	1,868
Injury, poisonings, toxic effects of drugs	675[a]	781	615
Burns	261	326	12
Factors influencing health	39	299	591

SOURCE: Authors' analysis of the MarketScan® database (MEDSTAT Group, Ann Arbor, Michigan), 1990–1994.
NOTES: SSRI is selective serotonin reuptake inhibitor. TCA is tricyclic antidepressant. MDD is major depressive disorder. NEC is not elsewhere classified.
[a] Value significantly different from zero at $p \leq 0.05$.

The most frequently occurring comorbid conditions, mental illnesses other than depression ("mental illness" in Exhibit 3), account for approximately $1,600 of the total charges of those episodes in which they occur. Other characteristics associated with large incremental spending included neurological disorders ($2,194 per episode), pregnancy and postpartum conditions ($1,697 per episode), and musculoskeletal conditions ($1,505 per episode). By contrast, incremental charges for most depression diagnoses were more modest, and the presence of a diagnosis of depression, NEC (the most frequently occurring depression diagnosis), is associated with lower-than-average spending. Based on these results, costs for an uncomplicated episode of depression, NEC, would be about $1,500.

After we adjusted for covariates, age and sex did not predict expenditures. To understand whether treatment choice had an effect on expected charges related to patient characteristics, we compared fluoxetine users with users of TCA as their initial treatment. There were no significant differences between SSRI and TCA users. Finally, we split our sample into those whose depressive episodes occurred in 1990–1992 (1,242 persons) and those in the larger sample. Although the larger sample size available for 1990–1994 resulted in more stable estimates, there were no substantive differences between the two analyses. We did not have complete data on benefit plans and thus were not able to study the impact of managed care on expenditure patterns.

Discussion

So, what does treatment of depression really cost? The answer depends on one's perspective. Treating an uncomplicated episode of depressive illness in this primary care population averaged a little more than $2,000, and treating the most common form was significantly less. However, depressed patients consume a disproportionate amount of medical resources: an overall average of nearly $8,000 during the year following the index prescription for an antidepressant. Clearly, treating patients with depression is very expensive, but depression-related expenditures represent only about a quarter of the total amount spent.

The number of comorbid conditions and the substantial impact these have on costs explain a great deal of the expenditure pattern. In this DataWatch we have described the impact of various medical and mental illnesses that occur in the year following initiation of antidepressant therapy on the costs of care for depressed patients who receive antidepressant medication. The results suggest that the costs related to care for depression itself make up only a small portion of the overall cost. Mental illnesses other than depression

"Our analysis suggests a strategy that health plans could use to manage depressed patients more efficiently."

account for nearly as much of the total cost as the cost of the underlying depressive illness.

The antidepressant-treated population described here has a very high prevalence of identified medical and psychiatric comorbidity. Most epidemiological surveys of mental illness, including the National Comorbidity Survey and the Epidemiologic Catchment Area study, have examined prevalence of comorbid conditions among persons with a mental disorder, and they have documented the frequency of treatment in this population. However, the prevalence of comorbid conditions in a treated population has not been adequately studied. Gregory Simon partially addresses the issue, finding that depressed patients in primary care tend to have more chronic diseases, but our study is the first to estimate the magnitude of the issue.[12] Further study will be required to determine the degree to which insurance claims for these disorders correspond to the actual presence of disease and the effect of alternative insurance arrangements. Our preliminary work in a state Medicaid program suggests that far fewer depressed patients seek care for comorbid anxiety compared with the privately insured population included in this report.[13]

Implications for health plans. Even if our estimates for the prevalence of comorbid mental illness are imprecise, depression and other mental disorders appear to go hand in hand in patients treated with antidepressants in primary care. Although the actual rate of coexistence of depression and other disorders is quite high, our findings suggest several interpretations of relevance to health plans. First, physicians often may confuse diseases with overlapping symptoms. This confusion, if true, undoubtedly would have detrimental effects on treatment choice. Second, there is some evidence that physicians alter the stated diagnosis depending on the motivation for reporting the diagnosis. Thus, insurance claims, medical records, and physicians' diaries may differ for the same patient.[14]

Our analysis suggests a strategy that health plans could use to manage depressed patients more efficiently. Frequently occurring patient characteristics that are predictive of higher costs could be used to help identify patients early in the course of their illness. Because these patients are already incurring high costs, there is little for a health plan to lose by applying intensive services earlier in the course of an illness. For example, depressed patients with comorbid anxiety might be identified for early intensive patient management.

Two important points can be made regarding the very high impact of nondepressive mental illnesses. First, although early, intensive management of patients with anxiety and adjustment disorders may not result in cost savings, it may offer the opportunity for improving the outcome of that care. Second, although our exclusion of patients with psychosis and substance abuse disorders limits the generalizability of our results regarding mental illness, one might predict similar results in dually diagnosed patients with these conditions.

Such intensive, early intervention might resemble collaborative treatment models in which a mental health specialist resides in a primary care clinic. Various forms of these models are now being investigated. In general, the early results suggest that improvements in care occur at some moderate increase in cost, but these results have not been stratified according to comorbid conditions.[15] Our results suggest that targeting collaborative intervention would improve the cost side of the cost-effectiveness equation, but the effect on health status remains to be seen.

It is frequently said that adequate treatment of mental illness and depression should result in reductions in costs related to physical symptoms that might be related to the underlying psychiatric disorder. Our analyses suggest that most of the expected costs to depressed patients cannot be differentially altered by currently available antidepressant medications. Thus, whereas new treatments appear to reduce the cost and improve the quality of depression care, there does not appear to be a similar reduction in costs for comorbid medical illnesses.[16]

Implications for cost-effectiveness research. Many health plans now use cost-effectiveness as one of several criteria for making decisions regarding coverage of new technologies. Cost-effectiveness studies of depression care are usually based on extrapolations from clinical trials or otherwise limited patient populations.[17] These studies are unlikely to capture the full range of depressed patients, especially those who are the highest cost. Because these costs appear to be related to comorbid conditions, cost-effectiveness studies based on more limited populations are likely to overstate potential cost savings and therefore overstate cost-effectiveness.

WE HAVE ATTEMPTED to disaggregate charges associated with various diagnostic and demographic characteristics of depressed patients. The patients described here have a wide variety of comorbid conditions, which add considerably to their "high-utilizer" status. Based on the results presented here, we doubt that any single program (or "magic bullet") is likely to result in significant cost containment. By providing more accurate esti-

mates of the impact of certain patient characteristics, our results should, however, provide a useful benchmark in the design of programs to contain costs and improve medical care.

The authors thank Richard Frank, Tim Hylan, and Cathy Melfi for their comments, and Ron Rogers and Mary Werts for assistance with preparation of the manuscript. The research for this DataWatch was funded by Lilly Research Laboratories. This paper was presented in part at the annual Association for Health Services Research meeting, 8–10 June 1996, in Atlanta.

NOTES

1. K.B. Wells, J.M. Golding, and M.A. Burnam, "Chronic Medical Conditions in a Sample of the General Population with Anxiety, Affective, and Substance Use Disorders," *American Journal of Psychiatry* 146, no. 11 (1989): 1440–1446; J.L. Coulehan et al., "Medical Comorbidity of Major Depressive Disorder in a Primary Medical Practice," *Archives of Internal Medicine* 150, no. 11 (1990): 2363–2367; K.B. Wells et al., "How the Medical Comorbidity of Depressed Patients Differs across Health Care Settings: Results from the Medical Outcomes Study," *American Journal of Psychiatry* 148, no. 12 (1991): 1688–1696; K.B. Wells, J.M. Golding, and M.A. Burnam, "Psychiatric Disorder in a Sample of the General Population with and without Chronic Medical Conditions," *American Journal of Psychiatry* 145, no. 6 (1988): 976–981; W. Katon, "Depression: Somatic Symptoms and Medical Disorders in Primary Care," *Comprehensive Psychiatry* 23, no. 3 (1982): 274–287; A. LaBruzza, "Physical Illness Presenting in Psychiatric Disorder: Guidelines for Differential Diagnosis," *Journal of Operational Psychiatry* 12, no. 1 (1981): 24–31; W. Katon, "Depression: Relationship to Somatization and Chronic Medical Illness," *Journal of Clinical Psychiatry* 45, no. 3, part 2 (1984): 4–11; and P.F.M. Verhaak, "Variations in the Diagnosis of Psychosocial Disorders: A General Practice Observational Study," *Social Science and Medicine* 23, no. 6 (1986): 595–604.
2. H.C. Schulberg, M. McClelland, and B.J. Burns, "Depression and Physical Illness: The Prevalence, Causation, and Diagnosis of Comorbidity," *Clinical Psychology Review* 7, no. 7 (1987): 145–167; and N. Frasure-Smith, F. Lesperance, and M. Talajic, "Depression following Myocardial Infarction: Impact on 6-Month Survival," *Journal of the American Medical Association* 270, no. 15 (1993): 1819–1825.
3. K.B. Wells et al., "Detection of Depressive Disorder for Patients Receiving Prepaid or Fee-for-Service Care: Results from the Medical Outcomes Study," *Journal of the American Medical Association* 262, no. 23 (1989): 3298–3302; W.W. Zung et al., "Recognition and Treatment of Depression in a Family Medicine Clinic," *Journal of Clinical Psychiatry* 44, no. 1 (1983): 3–6; M. Von Korff et al., "Anxiety and Depression in a Primary Care Clinic," *Archives of General Psychiatry* 44, no. 2 (1987): 152–156; and J.E. Barrett et al., "The Prevalence of Psychiatric Disorders in a Primary Care Practice," *Archives of General Psychiatry* 45, no. 12 (1988): 1100–1106.
4. M. Von Korff et al., "Disability and Depression among High Utilizers of Health Care," *Archives of General Psychiatry* 49, no. 2 (1992): 91–100; G. Simon et al., "Health Care Costs Associated with Depressive and Anxiety Disorders in Primary Care," *American Journal of Psychiatry* 152, no. 3 (1995): 352–357; and J. Johnson, M. Weissman, and G. Klerman, "Service Utilization and Social Morbidity Associated with Depressive Symptoms in the Community," *Journal of the American Medical Association* 267, no. 11 (1992): 1478–1483.
5. K.B. Wells et al., "Cost-Sharing and the Use of General Medical Physicians for

Outpatient Mental Health Care," *Health Services Research* 22, no. 1 (1987): 1–17; R.G. Frank and M.S. Kamlet, "Determining Provider Choice for the Treatment of Mental Disorder: The Role of Health and Health Status," *Health Services Research* 24, no. 1 (1989): 83–103; C.A. Jackson, W.G. Manning, and K.B. Wells, "Impact of Prior and Current Alcohol Use on Use of Services by Patients with Depression and Chronic Medical Illnesses," *Health Services Research* 30, no. 5 (1995): 687–705; G.E. Simon, M. Von Korff, and W. Barlow, "Health Care Costs of Primary Care Patients with Recognized Depression," *Archives of General Psychiatry* 52, no. 10 (1995): 850–856; T.W. Croghan et al., "Effect of Antidepressant Therapy of Health Care Utilization and Costs in Primary Care," *Psychiatric Services* 48, no. 11 (1997): 1420–1426; and W.E. Crown et al., "Application of Sample Selection Models to Outcomes Research: The Case of Evaluating the Effects of Antidepressant Therapy on Resource Utilization," *Statistics in Medicine* (forthcoming).

6. R.G. Frank, S.H. Busch, and E.R. Berndt, "Measuring Prices and Quantities of Treatment for Depression," *American Economic Review* 88, no. 2 (1998): 106–111; and E.R. Berndt, R.G. Frank, and T.G. McGuire, "Alternative Insurance Arrangements and the Treatment of Depression: What Are the Facts?" *American Journal of Managed Care* 3, no. 2 (1997): 243–250.
7. American Psychiatric Association, "Practice Guideline for Major Depressive Disorder in Adults," *American Journal of Psychiatry* 150, no. 4 (Supplement 1993): 1–26; Depression Guideline Panel, *Depression in Primary Care: Treatment of Major Depression,* Clinical Practice Guideline no. 5, AHCPR Pub. no. 93-0550 (Rockville, Md: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research, 1993).
8. J. Heckman, "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic Social Measurement* 5, no. 4 (1976): 475–492; J. Heckman, "Varieties of Selection Bias," *American Economic Review* 80, no. 2 (1990): 313–318; and C.A. Melfi and D.M. Waldman, "Limited and Discrete Dependent Variable Models: Use, Applications, and Comparisons," in *Econometric Modeling in Economic Education Research,* ed. W.E. Becker (Boston: Kluwer-Nijhoff, 1987).
9. An internal technical report is available from the authors upon request. Write to Thomas W. Croghan, Division of Health Services and Policy Research, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, Indiana 46285.
10. APA, "Practice Guideline for Major Depressive Disorder in Adults;" and Depression Guideline Panel, *Depression in Primary Care,* vol. 2.
11. Patients with either psychotic or substance abuse disorders were excluded from the analysis.
12. Simon et al., "Health Care Costs of Primary Care Patients."
13. C.A. Melfi and T.W. Croghan, "Access to Treatment of Depression in a Medicaid Population," *Journal of Healthcare for the Poor and Underserved* (forthcoming).
14. R.A. Browne et al., "Why the Quality of Antidepressant Claims Is Uncertain and What Can Be Done to Improve It," *Drug Benefit Trends* (forthcoming); and K. Rost et al., "The Deliberate Misdiagnosis of Major Depression in Primary Care," *Archives of Family Medicine* 3, no. 4 (1994): 333–337.
15. W. Katon et al., "Collaborative Management to Achieve Treatment Guidelines," *Journal of the American Medical Association* 273, no. 13 (1995): 1026–1031.
16. Frank et al., "Measuring Prices and Quantities of Treatment for Depression;" and Berndt et al., "Alternative Insurance Arrangements and the Treatment of Depression."
17. M. Hotopf, G. Lewis, and C. Normand, "Are SSRIs a Cost-Effective Alternative to Tricyclics?" *British Journal of Psychiatry* 168, no. 4 (1996): 404–409.

Statistical Methods used in "What does treatment of depression really cost?"

by

Robert L. Obenchain

These notes document the statistical methods used to compute incremental charge estimates (and con…dence bounds) for Exhibit 3 of Croghan, Obenchain and Crown(1998). A relatively wide variety of methods were dictated by the diversity of issues encountered in this retrospective study of health-care claims datasets extracted from the MarketScan® database (MEDSTAT Group, Ann Arbour, Mich.) See the table at the top of page 2.

KEY WORDS: skewness reducing transformations; smearing retransformation estimates; ill-conditioned multiple linear regression model; normal theory maximum likelihood shrinkage; extent and shape of shrinkage; Bayesian highest posterior density intervals; estimation of incremental e¤ects.

[0]Robert L. Obenchain is a Research Scientist providing statistical support of the USA Health Outcomes Evaluation Group, Lilly Research Laboratories, Lilly Corporate Center, Indianapolis, IN 46285-1850; (317) 276-3150; ochain@lilly.com.

Three Classes of Problems	Methodology
Multicollinearity	Shrinkage Regression Estimation and
	Bayesian Highest Posterior Density Intervals
Skewness of Charges Distribution	LOG( ) Transformation, Prediction,
	EXP() Retransformation and “Smearing”
Estimation of Incremental E¤ects	Classi…cation into Three or More Mutually
	Exclusive / Exhaustive Categories

1 Multicollinearity (Correlated Predictor Variables)

Croghan et al.(1998) used multiple linear regression models with more than 30 di¤erent regressor variables to predict the response variable, y = natural logarithm of total annual health-care charges. (There were no observed $0 values for total, yearly charges.) Since we initially had data on n = 1242 patients (and ultimately on 3439 patients), using 30 to 40 predictor variables may seem quite reasonable in the sense that we had more than 1200 (or 3400) “degrees of freedom” left over to estimate “error” and/or lack-of-…t in these models.

The truth is that even relatively small intercorrelations between thirty or more

predictor variables can lead to truly "serious" multicollinearity (ill-conditioning.) Speci…cally, least squares estimates for regression coe¢cients then have a tendency to be unstable relative to very small numerical changes in the data. [This is typically illustrated by adding random noise after the last reported decimal place in the data and/or by deleting a few observations that are not outliers!] In particular, some least squares estimates may have numerical signs that are opposite to the sign of the (marginal) correlation exhibited when log(charge) is plotted versus only that predictor variable. The sign of one of these two statistics (either the marginal correlation or the …tted coe¢cient) will usually make common sense, while the other will then seem "wrong" or "unbelievable.." Problems with numerical signs in the multiple regression models of Croghan et al.(1998) are summarized in the table below.

VARIABLE	n = 1242			n = 3439		
	Corr	OLS	Ridge	Corr	OLS	Ridge
GENDER	+0.0462	¡0.05775 *	+0.00065	+0.0721	¡0.01595	¡0.00994
DEP_DX4	¡0.0360	+0.04921 *	¡0.03384	+0.0038 *	+0.02570	¡0.00604
GEO_2	+0.1032	¡0.02806	¡0.00493	+0.0411	¡0.00982	¡0.01422
PROZAC	¡0.0293	+0.07305 *	¡0.01295			
SSRI				¡0.0300	+0.00239 *	¡0.00072
SWIT_AUG				+0.0842	¡0.00565 *	+0.00086

In the above table, asterisks (*) mark "wrong" sign problems that were successfully treated using maximum likelihood "ridge regression" methodology described below.

Here, one might expect that the DEP_DX4 indicator of “brief depressive episode” would be associated with low annual health care costs, but ordinary least squares estimates do not necessarily support this expectation. Similarly, treatment of depression using Prozac (‡uoxetine; initial n=1242 dataset) or any of the three major SSRIs (‡uoxetine, sertraline or paroxetine; ultimate n=3439 dataset) might well be expected to be more cost-e¤ective than using a TCA/HCA ...rather than, say, equally or less cost-e¤ective. Finally, the e¤ect of GENDER (0=male,1=female) strikes me as being most curious. The GENDER indicator is highly correlated with “male and female reproductive system” problems (STDMDC12 and STDMDC13, respectively) and with “factors in‡uencing health” (STDMDC23) that tend to be more highly correlated with log(charges) than is GENDER alone; here are these intercorrelations within the …nal dataset (n=3439.)

GENDER	+1.0000			
STDMDC12	¡0.3263	+1.0000		
STDMDC13	+0.4193	¡0.1350	+1.0000	
STDMDC23	+0.2983	¡0.0435	+0.2425	+1.0000
log(charges)	+0.0721	¡0.0535	+0.1580	+0.1775	+1.0000

In other words, women consistently tended to incur higher log(charges) than men, but the GENDER indicator received “negative” credit as it battled with STDMDC codes 12, 13 and 23 for least-squares “recognition.” Maximum likelihood shrinkage methodology failed (for the n=3439 dataset) to “correct” the sign of the GENDER

coe¢cient but at least succeeded in preventing this male-female di¤erence from being labelled “statistically signi…cant.”

1.1 Regression Model Equations

A multiple linear regression model for y = log(charge) is expressed in matrix notation as

$$E(y|X) = 1^{1} + X^{-} \text{ and } Var(y|X) = \%^{2}I \ , \qquad (1)$$

where y is a n £ 1 vector of observed response values; 1 is a n £ 1 vector of ones; X is a given n £ p matrix of non-constant regressor coordinates; I is an n £ n identity matrix; ^{1} is the unknown intercept scalar; $^{-}$ is a p £ 1 vector of unknown regression coe¢cients; and the responses are stochastic and uncorrelated with constant, unknown variance, $\%^{2}$:

The shrinkage/ridge regression methods we used to treat observed symptoms of multicollinearity require a three-step estimation process:

(1) place the data in a canonical form that minimizes the effects of one's choice for scaling of predictor variables on principal axes and coordinates,

(2) estimate standardized coefficients using maximum likelihood methods, and

(3) compute the corresponding coefficients for the original choice of axis scaling.

In step (1), the data are first "centered" by subtracting off column means, so that $1'y = 0$ and $1'X = 0'$. In other words, centering allows the above model to be written succinctly as $E(y|X) = X\beta$ and $Var(y|X) = \sigma^2(I - 11'/n)$, where the intercept term from equation (1) is implicitly $\alpha = \bar{y} - \bar{x}'\beta$. Thus, in shrinkage/ridge regression, one's estimate for the α scalar changes because the estimate of the β vector is shrunken. This convention assures that every shrinkage/ridge fitted hyperplane always passes through $y = \bar{y}$ at $x = \bar{x}$, as does the fitted ordinary least squares hyperplane.

Next, step (1) also requires that each variable be rescaled by dividing its coordinates by the observed standard deviation of those coordinates. For example, the standard deviation of the given response = log(charge) variable is $s_y = \sqrt{\sum (y_i - \bar{y})^2/(n-1)}$. The canonical form for variables is thus equivalent to "z-scores" of the form $z_i = (y_i - \bar{y})/s_y$. It follows that the sample sum-of-squares for each canonical-form variate is always $z'z = (n - 1)$.

The equation for expected values in (1) can now be rewritten as

$$E\left[\frac{(y_i - \bar{y})}{s_y} \mid x_i^0\right] = \sum_{j=1}^{P} \frac{(x_{ij} - \bar{x_j})}{s_j} \beta_j \tag{2}$$

where i is the index for observations ($1 \le i \le n$), j is the index for predictor variables ($1 \le j \le P$), $x_i^0 = (x_{i1}, \ldots, x_{iP})$ is the i-th row of the original predictors X matrix, and s_j denotes the original standard deviation in values in the j-th column of this X matrix.

The ordinary least squares estimate of β is of the form $\hat{\beta} = b^o = (Z_X^0 Z_X)^{-1} Z_X^0 (y - \bar{y}1)/s_y = Gc$, where Z_X denotes the z-score scaling of the original X matrix, G is the (orthogonal) "direction cosines" matrix for the principal axes of Z_X , and c contains the "uncorrelated components" of b^o. Furthermore, $Var(b^o)$ is given by the matrix expression $\sigma^2 (Z_X^0 Z_X)^{-1} = \sigma^2 G \Lambda^{-1} G^0$, where Λ is the diagonal matrix of principal axis eigenvalues. Note that $\sigma^2 = (1 - R^2)$, where R^2 is the familiar R-squared statistic.

1.2 Shrinkage Regression Equations

Shrinkage/ridge estimates of β are of the form $\hat{\beta} = b^* = G \Gamma c$, where $\Gamma = \mathrm{diag}(\gamma_1, \ldots, \gamma_P)$ is a diagonal matrix of principal axis "shrinkage factors," each on the range $0 \le \gamma_i \le 1$. The classical estimate of $Var(b^*)$ is then $\sigma^2 \Gamma G \Gamma^2 \Lambda^{-1} G^0$, while its Bayesian posterior variance is $\sigma^2 \Gamma G \Gamma \Lambda^{-1} G^0$, which tends to be somewhat larger than the classical dispersion whenever the γ_i are strictly less than 1 (and thus $\gamma_i^2 < \gamma_i$.)

There is insufficient space here to give full details about how normal-distribution-

theory methods are used to estimate ¢ so as to maximize the likelihood that the resulting shrinkage/ridge estimator, b¤, exploits potential variance-bias trade-o¤s and achieves minimum mean-squared-error risk. Shrinkage/ridge methods were …rst introduced by Hoerl and Kennard(1970a,1970b); details of my maximum likelihood approach are given in Obenchain(1975,1995). This methodology selects both the "shape" of the shrinkage path, Goldstein and Smith(1974), and the "extent" of shrinkage along that path. The methodology has the potential to work very well in actual practice, as demonstrated by the simulation results of Gibbons(1981). To form con…dence regions centered at shrunken estimates, the Bayesian highest posterior density methodology outlined by Lindley and Smith(1972) is used. A variety of free software implementations of my maximum likelihood shrinkage/ridge regression algorithms can be downloaded from the internet, Obenchain(1998a).

1.3 Di¤erences in Coe¢cient Estimates Between Datasets

Again, the regression models considered in Croghan et al.(1998) are "ill-conditioned" in the sense that signi…cant correlations exist between predictor variables. On the other hand, this ill-conditioning was considerably weaker in the second, larger dataset (3439 patients from the MedStat II study) than in the …rst (1242 patients from the MedStat I study.) To illustrate numerical instability in …tted regression coe¢cients, consider Figures 1 and 2, below.

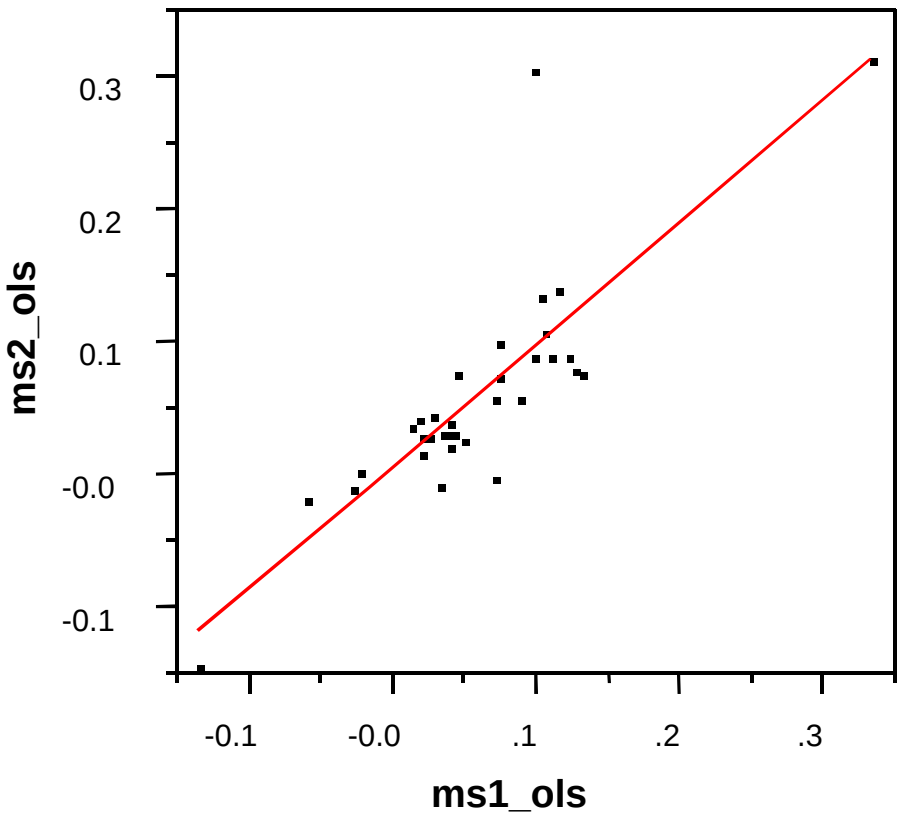


Figure 1: Ordinary least squares coe¢cient estimates from the 2nd MedStat study may be predicted from those from the 1st MedStat study by the linear relationship ms2_ols = .006 + 0.922 ms1_ols, but the R-square for this relationship is only 69.7%.

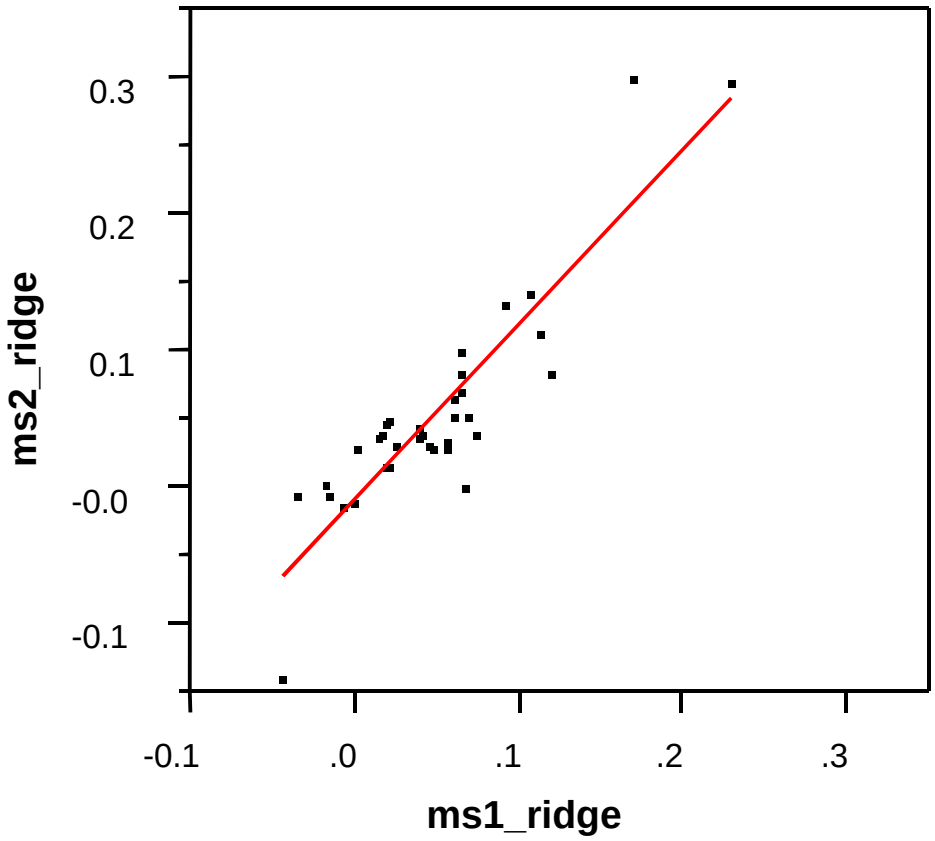


Figure 2: Maximum likelihood shrunken coe¢cient estimates from the 2nd MedStat study may be predicted from those from the 1st MedStat study by the linear relationship ms2_ridge = -0.010 + 1.277 ms1_ridge, where the R-square for this relationship has increased to 80.7%.

1.4 TRACE Displays, Etc., Etc.

The appendix to this paper contains graphical displays that show how the relative magnitudes and numerical signs of …tted coe¢cients change upon shrinkage as well as tables of the resulting elasticities (and 95% con…dence intervals for them.) Results are tabulated no only for log(charges) but also for "smeared" exp(predictions), as explained next.

2 Skewness in the Distribution of Charges

Health care charges cannot be negative. And, while the vast majority of patients may well incur relatively low charges, a few patients may incur very high charges. As a result, typical charge distributions are highly (positively) skewed in the sense that, instead of having any point of symmetry, they tend to have a very long, heavy, right-hand "tail." In other words, the mean charge (arithmetic average) incurred is typically much, much higher than the median charge (50% point) or mode charge (most frequently observed value.) For example, the distribution of charges speci…cally for hospitalization may have median = mode = $0 because fewer than half of the patients under study have a hospitalization in any given year.

Monotonic, nonlinear transformations of charges are commonly applied in order to produce a statistical distribution that is much more nearly symmetric; i.e. one in which the mean, median and mode are all nearly equal. The transformed distribu-

tion can then be much more realistically approximated by a "normal" (bell-shaped) distribution. A square-root transformation is sometimes used for this, but all transformed values then remain strictly non-negative (i.e. bounded from below rather than unbounded.) Natural logarithms (base e = 2.71828...) can be applied as long as all observed charges are strictly positive; log transformations map the value 0 to minus in…nity, ¡ 1 : Furthermore, a log transformation applied to strictly positive charges yields an unbounded variable; one that can assume all values strictly between ¡ 1 and + 1 , like a true normal distribution.

While statistical models for log(charges) can be more "realistic" than models of charges, the …nal step in the analysis is usually to re-express results back in terms of the original charge units. The main problem that then arises is that statistics of the form predicted charge = exp[predicted log(charge)] typically tend to be severely biased downward. For example, suppose that the true distribution of y = log(charges) actually was normal with speci…ed mean μ and variance σ^2. The mean value of exp(y) is then

$$E[\exp(y)] = e^{\mu+\sigma^2/2} = \exp(\mu) \pounds \exp(\sigma^2/2) > \exp(\mu) \text{ whenever } \sigma^2 > 0 . \quad (3)$$

In other words, the mean of the charge = exp(y) distribution then depends upon the variance as well as upon the mean of the distribution of y = log(charge).

Duan(1983) describes this phenomenon as "unbiased and consistent quantities on the transformed (y) scale usually do not retransform into unbiased or consistent

quantities on the untransformed (charges) scale." To adjust for this downward bias, the nonparametric "smearing estimate" of Duan(1983) is of the form

$$\begin{aligned} \text{smeared}[y|x_o] &= \text{smeared prediction at } x = x_o, \\ &= \exp(x_o' b) \times \frac{1}{n} \sum_{i=1}^{n} \exp(e_i) \ . \end{aligned} \tag{4}$$

In other words, the smeared prediction is obtained by simply multiplying exp(predicted y at x_o) by a factor that does not depend upon x_o. This factor is the average value of exp(residual) over all n observations in the regression, where residuals are defined as usual by e_i = observed y_i minus the predicted y at $x = x_i$ (rather than predictions at the "target" regressor combination, $x = x_o$.) Now, note that the average value of e_i is 0, where exp(0) = 1. Thus negative residuals contribute $\exp(e_i)$ terms that fall somewhere in the range from 0 to 1, while positive residuals contribute $\exp(e_i)$ terms in the unbounded range from 1 to + 1 . Thus the Duan(1983) smearing factor is typically greater than one.

The table below lists numerical values for this smearing factor in the multiple regression models discussed both in the original manuscript (1996-1997) and in the revised/published version of Croghan et al.(1998).

Original 1990-92 Sample	Number of Patients	$\frac{1}{n}\sum_{i=1}^{n}\exp(\beta_i)$
Fluoxetine	799	1.490
HCA/TCA	443	1.418
Combined	1242	1.487
Ultimate 1990-94 Sample	Number of Patients	$\frac{1}{n}\sum_{i=1}^{n}\exp(\beta_i)$
SSRI	2406	1.485
HCA/TCA	1033	1.596
Combined	3439	1.518

3 Equations for Predictions and Forecasts

The dispersion information from equation (1) can be combined with equation (2) by writing

$$\frac{(y_i - \bar{y})}{s_y} = \sum_{j=1}^{p} \frac{(x_{ji} - \bar{x}_j)}{s_j} c_j + \varepsilon_i \qquad (5)$$

where the error terms, $\varepsilon_1, \ldots, \varepsilon_n$, are interchangeable random variables with mean zero and variance $\sigma^2(n-1)/n$. (The ε_i variates cannot be statistically independent because they always sum to exactly zero.)

Equation (5) can be rearranged to make predictions and forecasts for a given set of predictor values $x_o^0 = (x_{o1},\ldots,x_{on})$ using any given coefficient estimates $b^0 =$

$\beta_1, \ldots, \beta_P$. For this purpose, we multiply both sides of (5) by s_y and move the $\bar{y}$ term to the right hand side of the equation to get...

$$\hat{y}_0 = \bar{y} + s_y \cdot \sum_{j=1}^{P} \frac{(x_{oi} - \bar{x}_j)}{s_j} \cdot \beta_j + s_y \cdot \varepsilon \qquad (6)$$

where the ε term will always be zero when making point predictions of averages and forecasts for individual patients.

On the other hand, the variability in $\hat{y}_0$ needs to be accounted for when forming con...dence limits. Thus, when forming $100(1 - \alpha)\%$ upper or lower con...dence limits on log(charge) forecast for an individual patient (outside the current sample of n), the contribution from the $s_y \cdot \varepsilon$ term will be $\pm$ tscore $\cdot s_y \cdot \sqrt{1}$, where tscore is the upper $100(1 - \alpha/2)\%$ point of the student's t distribution with $(n - P - 1)$ degrees-of-freedom.

The uncertainty in the linear combination of β estimates de...ned by the predictor variable z-score vector $z_{xo}' = [\frac{(x_{o1} - \bar{x}_1)}{s_1}, \ldots, \frac{(x_{oP} - \bar{x}_P)}{s_P}]$ is estimated using the quadratic form

$$\widehat{Var}(z_{xo}) = \sum_{i=1}^{P} \sum_{k=1}^{P} \frac{(x_{oi} - \bar{x}_i)}{s_i} \cdot \frac{(x_{ok} - \bar{x}_k)}{s_k} \cdot \widehat{cov}(\beta_i, \beta_k) = z_{xo}' \widehat{Var}(\beta) z_{xo} . \qquad (7)$$

and the uncertainty in $\bar{y}$ is given by the familiar formula $Var(\bar{y}) = s_y^2/n$. In particular, note that for predictions at the X variable centroid ($x = \bar{x}$), that $z_{\bar{x}} = 0$, the predicted log(charge) is $\bar{y}$, and that $\widehat{Var}(z_{\bar{x}}) = 0$.

In summary, we have

$$\begin{aligned} \hat{y}(x_o) &= \text{prediction or forecast of log(charges) at } x = x_o \\ &= \bar{y} + s_y \cdot \sum_{j=1}^{P} \frac{(x_{oi} - \bar{x}_j)}{s_j} \cdot b_j \ . \end{aligned} \tag{8}$$

Furthermore, the upper and lower 100(1 - α)% con...dence limits on $\hat{y}(x_o)$ as a prediction of average log(charges) are of the form

$$\hat{y}(x_o) \pm \text{tscore} \cdot s_y \cdot \sqrt{1/n + \widehat{\text{Var}}(z_{xo})} \ , \tag{9}$$

where tscore is again the upper 100(1 - α/2)% point of the student's t distribution with (n - P - 1) degrees-of-freedom.

The upper and lower 100(1 - α)% con...dence limits on $\hat{y}(x_o)$ as a forecast for an individual patient can be much wider

$$\hat{y}(x_o) \pm \text{tscore} \cdot s_y \cdot \sqrt{1/n + \widehat{\text{Var}}(z_{xo}) + 1} \ . \tag{10}$$

Finally, remember that these estimates and limits are to be retransformed back to the charges scale and smeared as in equation (4) to yield

$$\exp[\hat{y}(x_o) \pm \text{whatever}] \cdot \frac{1}{n} \sum_{i=1}^{n} \exp(e_i) \ . \tag{11}$$

4 Incremental E¤ects for Mutually Exclusive Classes

To identify di¤erences in treatment and/or covariate characteristics between patients, multiple regression models commonly employ so-called "dummy" or "indicator" X variables. For each individual patient, these variables assume one of only two possible numerical values. Speci…cally, zero => absence while one => presence of a characteristic or treatment. Thus, if the …rst column of X, namely x_1, is an indicator variable of this form, then $\bar{x}_1$ is the fraction of patients for which the corresponding treatment or characteristic is present. Furthermore, the sample standard deviation of an indicator variable is of the special form $s_1 = \sqrt{\bar{x}_1 \cdot (1 - \bar{x}_1)} \cdot \sqrt{n/(n - 1)}$. In other words, the z-scores associated with $x_1 = 0$, $x_1 = \bar{x}_1$, and $x_1 = 1$ are $z = -\sqrt{\bar{x}_1/(1 - \bar{x}_1)} \cdot \sqrt{n/(n - 1)}$, $z = 0$, and $z = +\sqrt{(1 - \bar{x}_1)/\bar{x}_1} \cdot \sqrt{n/(n - 1)}$, respectively. Predictions and forecasts for speci…ed values of indicator variables are thus easily made using equations (8) through (11).

Special considerations may have to be applied, however, when a multiple linear regression model contains several indicator variables that, together, de…ne mutually exclusive and collectively exhaustive classes of patients.

First of all, if there are a total of K such classes, then regression models that contain an overall intercept term will usually contain only (K-1) such indicator variables. This convention allows β to be "estimable" by ordinary-least-squares. (Speci…cally, if all K indicator variables were included, then the sum of those K columns in the X

matrix would be a column of ones, the $Z'Z$ matrix would be singular, and $(Z'Z)^{-1}$ would not exist!)

When K=2, it is clear that only one indicator variable is needed because the second indicator would always equal (1 minus the …rst indicator.) In other words, there would always be a "perfect" correlation of ¡1 between these two indicator variables. And the prediction or forecast at $x_2 = x_o$ is always the same as the prediction or forecast at $x_1 = 1 - x_o$.

However, when $K \geq 3$ mutually exclusive and collectively exhaustive classes are to be included, things get more complicated. Suppose that the indicator variables for the …rst (K-1) out of K classes are the ones included in the multiple regression model. In this case, it is clear how to make the prediction or forecast when any one $x_k = 1$ within $1 \leq k < K$. After all, $x_k = 1$ means that all other $x_j = 0$ for $j \neq k$. Similarly, it is clear how to made the prediction or forecast for $x_1 = x_2 = \cdots = x_{K-1} = 0$; this is the prediction or forecast for a patient diagnosed into the …nal category, the one with no indicator variable included in the multiple regression model.

The problem is... "What is the prediction or forecast most appropriate when all we know is that $x_k = 0$ for the k-th class?" In other words, the only information we are given is that one, single diagnosis does NOT apply to that patient.

The solution stems from realizing that, if $x_k = 0$, then exactly one $x_j = 1$ for $j \neq k$, while all other such $x_j = 0$. Since we don't know which x_j is one, we form a

weighted average of predictions, where the weights are the estimated conditional probabilities that $x_j = 1$ given that $x_k = 0$ for all $j \neq k$:

$$\text{Prediction or Forecast at } (x_k = 0) = \sum_{j \neq k} \frac{\bar{x}_j}{(1 - \bar{x}_k)} \times \text{Prediction or Forecast at } (x_j = 1) , \tag{12}$$

where $\bar{x}_K = 1 - \bar{x}_1 - \cdots - \bar{x}_{K-1}$

The overall "incremental effect" of a single class, k, within $K \geq 3$ mutually exclusive and exhaustive classes is then...

$$\begin{aligned} \text{Incremental Effect of } x_k &= \text{Prediction at } (x_k = 1) - \text{Prediction at } (x_k = 0) \qquad (13) \\ &= \exp\left[\bar{y} + s_y \sum b_i \times \bar{x}_i / s_i\right] \times \left[f_k - \sum_{j \neq k} \frac{\bar{x}_j}{(1 - \bar{x}_k)} \times f_j\right], \\ &= \exp\left[\bar{y} + s_y \sum b_i \times \bar{x}_i / s_i\right] \times \left[\frac{f_k - \sum_{j=1}^{K} \bar{x}_j \times f_j}{(1 - \bar{x}_k)}\right], \end{aligned}$$

where $f_j = \exp(b_j / s_j)$ for $1 \leq j < K$ and $f_K = 1$.

5 Summary

The calculations described here tend to be somewhat complicated, but they are actually quite easy to implement in a spreadsheet program such as Microsoft Excel®.

References

Croghan, T. W., Obenchain, R. L. and Crown, W. E. (1998). "What does treatment of depression really cost?" (Although treating a patient with depression is expensive, the bulk of the cost is for accompanying illnesses.) Health A¤airs 17(4), 198-207.

Duan, N. (1983). "Smearing estimate: a nonparametric retransformation method." Journal of the American Statistical Association 78, 605-610.

Gibbons, D. G. (1981). "A simulation study of some ridge estimators." Journal of the American Statistical Association 76, 131-139.

Goldstein, M. and Smith, A. F. M. (1974). "Ridge-type estimators for regression analysis." Journal of the Royal Statistical Society B 36, 284-291

Hoerl, A. E. and Kennard, R. W. (1970a). "Ridge regression: biased estimation for non-orthogonal problems." Technometrics 12, 55-67.

Hoerl, A. E. and Kennard, R. W. (1970b). "Ridge regression: applications to non-orthogonal problems." Technometrics 12, 69-82. (errata: 723.)

Lindley, D. V. and Smith, A. M. F. (1972). "Bayes estimates for the linear model." Journal of the Royal Statistical Society B 34, 1-72.

Obenchain, R. L. (1975). "Ridge analysis following a preliminary test of the shrunken hypothesis." Technometrics 17, 431-441. (with discussion 443-445.)

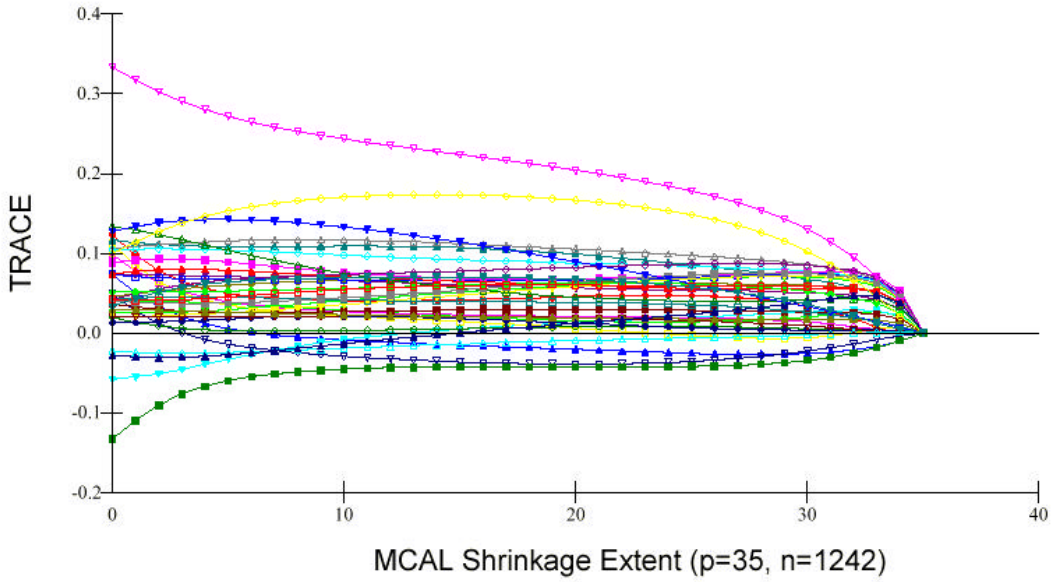
Obenchain, R. L. (1995). "Maximum likelihood ridge regression." Stata Technical Bulletin 26, 33-55.

Obenchain, R. L. (1998a). "Shrinkage Regression" web pages. http://www.iquest.net/~softrx.

Obenchain, R. L. (1998b). "RXridge: Microsoft Windows API Software. (unfinished; alpha-test only.)

Appendix 1 - First MEDSTAT Study

Coefficient ridge trace display for path Q-shape = -3.5 on the MEDSTAT I dataset (1242 patients with episodes of depression in 1990-92.) The extent of shrinkage most likely to achieve minimum MSE risk here is MCAL = 13. This rank deficiency suggests that the 35 intercorrelated predictor variables are about as adequate as 35 - 13 = 22 uncorrelated predictors.



MedStat One Study (1242 Patients; Depression Episodes 1990-1992)
==========

R-squared Statistic = 0.5070
OLS Residual Variance = 0.507303

Maximum Likelihood Shrinkage Regression Estimation
Shrinkage Path Q-shape = -3.5
Shrinkage Extent = MCAL = 13.0 (rank deficiency)

NOTE: # symbol marks a sign conflict (correlation vs coefficient)

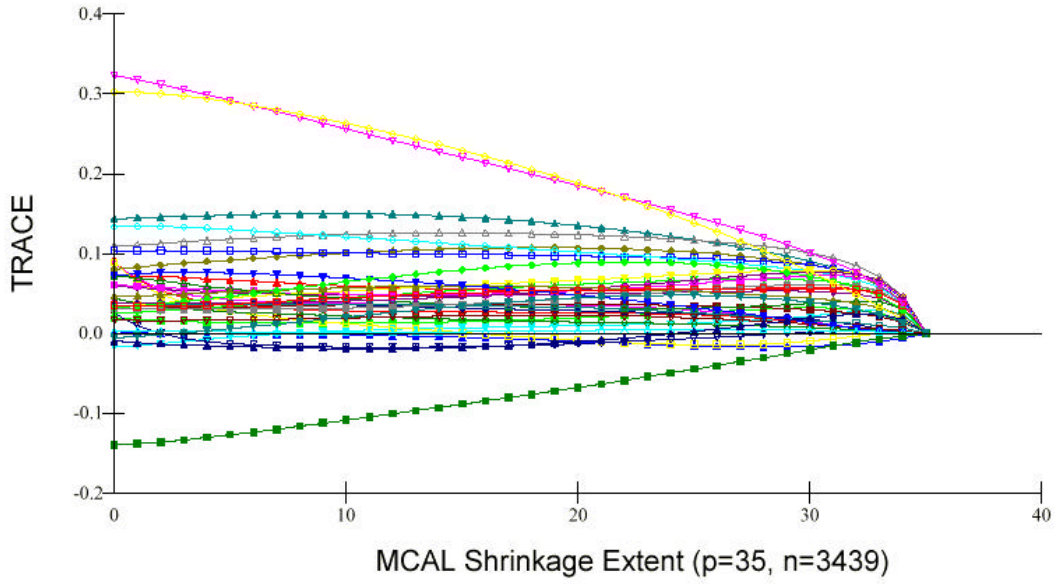
Var.Name	Marginal Correlation	OLS	Low-Ridge		Ridge	High-Ridge	
AGE	0.0975	0.04140	-0.00545		0.02380	0.05306	
GENDER	0.0462	#-0.05775	-0.02642		0.00065	0.02772	
DEP_DX1	0.0993	0.12382	0.01110		0.04053	0.06997	*
DEP_DX2	0.0663	0.09820	-0.00785		0.02397	0.05578	
DEP_DX3	0.0605	0.11193	-0.01027		0.01745	0.04517	
DEP_DX4	-0.0360	# 0.04921	-0.06477		-0.03384	-0.00292	*
DEP_DX5	0.0159	0.01969	-0.02947		0.00454	0.03855	
STDMDC1	0.2504	0.11627	0.07542		0.10830	0.14118	*
STDMDC2	0.0950	0.03651	-0.00578		0.02816	0.06210	
STDMDC3	0.1826	0.07298	0.03043		0.06278	0.09513	*
STDMDC4	0.1985	0.04588	0.03437		0.06679	0.09921	*
STDMDC5	0.2958	0.10593	0.08413		0.11446	0.14478	*
STDMDC6	0.2453	0.07358	0.03529		0.06633	0.09736	*
STDMDC7	0.1950	0.02537	0.01506		0.04703	0.07900	*
STDMDC8	0.2748	0.10440	0.06274		0.09398	0.12521	*
STDMDC9	0.2100	0.07427	0.03559		0.06720	0.09881	*
STDMDC10	0.2697	0.08818	0.04101		0.07072	0.10043	*
STDMDC11	0.1956	0.02061	0.01050		0.04270	0.07490	*
STDMDC12	0.0275	0.01301	-0.01258		0.01981	0.05220	
STDMDC13	0.1996	0.13393	0.03420		0.06276	0.09132	*
STDMDC14	0.0868	0.04226	0.00819		0.04049	0.07279	*
STDMDC15	0.0297	0.02063	-0.01205		0.02117	0.05439	
STDMDC16	0.2533	0.04150	0.04442		0.07591	0.10739	*
STDMDC17	0.0691	0.02962	-0.01262		0.02133	0.05527	
STDMDC18	0.2135	0.04063	0.01744		0.04978	0.08212	*
STDMDC19	0.1771	0.12729	0.08834		0.12232	0.15630	*
STDMDC21	0.1607	0.05110	0.02481		0.05847	0.09213	*
STDMDC22	-0.0195	-0.02214	-0.05007		-0.01577	0.01854	
STDMDC23	0.1898	0.04237	0.02690		0.05829	0.08967	*
STDNADM	0.4896	0.33367	0.20488		0.23119	0.25750	*
STDPADM	0.2948	0.09905	0.14627		0.17334	0.20040	*
GEO_2	0.1032	#-0.02806	-0.03579	#	-0.00493	0.02594	
PROZAC	-0.0293	# 0.07305	-0.03833		-0.01295	0.01243	
PROZINTR	-0.1289	-0.13322	-0.06804		-0.04310	-0.01816	*
PROZSWIT	0.1050	0.03349	0.03792		0.07035	0.10278	*

Exhibit 3 - Preliminary (1242 Patients, 1990-92)
Incremental Charges and Confidence Limits following exp() and smearing

Variable	OLS		LOW	HIGH	qshape=-3.5 mcal=13		LOW	HIGH
age	25		-2	52	15		-3	32
gender	-840	*	-1559	-167	9		-376	380
dep_dx1	1395	*	911	1975	600	*	386	836
dep_dx2	1437	*	711	2384	379		-4	845
dep_dx3	591	*	384	795	82	*	37	109
dep_dx4	-303		-617	109	-861	*	-1009	-680
dep_dx5	448		-2042	5709	175		-1573	3341
dep_dx6	-1974	*	-935	-3289	-244		216	-791
no comorbid	-3892		-2550	-4577	-3815		-2832	-4405
stdmdc1	1840	*	1138	2592	1702	*	1153	2282
stdmdc2	786		-85	1775	598		-116	1393
stdmdc3	868	*	367	1372	747	*	361	1133
stdmdc4	607	*	45	1193	893	*	452	1349
stdmdc5	1438	*	820	2086	1561	*	1129	2006
stdmdc6	1020	*	402	1672	916	*	478	1370
stdmdc7	619		-393	1811	1197	*	360	2143
stdmdc8	1239	*	723	1757	1115	*	744	1487
stdmdc9	885	*	371	1403	800	*	423	1180
stdmdc10	1150	*	551	1773	915	*	524	1316
stdmdc11	308		-320	983	651	*	156	1174
stdmdc12	494		-954	2394	769		-439	2261
stdmdc13	1798	*	1167	2458	816	*	439	1202
stdmdc14	1899	*	19	4418	1808	*	326	3665
stdmdc15	1771		-1317	7220	1824		-823	6020
stdmdc16	735		-37	1590	1401	*	790	2060
stdmdc17	948		-315	2504	668		-362	1894
stdmdc18	735		-35	1588	910	*	307	1562
stdmdc19	2283	*	1662	2826	2212	*	1688	2679
stdmdc21	1026	*	192	1962	1187	*	480	1963
stdmdc22	-1496		-3230	1644	-1115		-2804	1691
stdmdc23	512		-10	1042	706	*	324	1093
prozac	884	*	282	1469	-161		-481	153
stdnadm	11521	*	9057	14394	6632	*	5615	7739
stdpadm	4145	*	1911	7043	9161	*	7087	11576
geo_2	-367		-959	200	-64		-470	331
prozintr	-1591	*	-2162	-1004	-530	*	-830	-225
prozswit	1061		-261	2702	2457	*	1216	3922

Appendix 2 - Second MEDSTAT Study

Coefficient ridge trace display for path Q-shape = -3.0 on the MEDSTAT II dataset (3439 patients with episodes of depression in 1990-94.) The extent of shrinkage most likely to achieve minimum MSE risk here is MCAL = 3. This rank deficiency suggests that the 35 intercorrelated predictor variables are about as adequate as 35 - 3 = 32 uncorrelated predictors. In particular, the MEDSTAT II data are much less severely ill-conditioned than the MEDSTAT I data.



MedStat Two Study (3439 Patients; Depression Episodes 1990-1994)
==========

R-squared Statistic = 0.5789
OLS Residual Variance = 0.42548

Maximum Likelihood Shrinkage Regression Estimation
Shrinkage Path Q-shape = -3.0
Shrinkage Extent = MCAL = 3.0 (rank deficiency)

NOTE: # symbol marks a sign conflict (correlation vs coefficient)

Var.Name	Margin Correlation	OLS	Low-Ridge		Ridge	High-Ridge	
AGE	0.0816	0.01907	-0.00712		0.01500	0.03713	
GENDER	0.0721	#-0.01595	-0.03192	#	-0.00994	0.01203	
DEP_DX1	0.0622	0.09070	0.02516		0.04504	0.06492	*
DEP_DX2	0.0672	0.08853	0.02924		0.04941	0.06958	*
DEP_DX3	0.0291	0.08955	0.01961		0.03934	0.05906	*
DEP_DX4	0.0038	0.02570	-0.02672	#	-0.00604	0.01465	
DEP_DX5	0.0402	0.04358	0.00941		0.03080	0.05219	*
STDMDC1	0.3214	0.14386	0.12433		0.14650	0.16868	*
STDMDC2	0.0933	0.03030	0.01054		0.03215	0.05376	*
STDMDC3	0.1711	0.05965	0.03195		0.05402	0.07610	*
STDMDC4	0.2680	0.08024	0.06568		0.08787	0.11006	*
STDMDC5	0.3188	0.10914	0.09184		0.11445	0.13705	*
STDMDC6	0.2807	0.10293	0.08122		0.10359	0.12597	*
STDMDC7	0.1814	0.02521	0.00810		0.03025	0.05240	*
STDMDC8	0.2854	0.13374	0.11085		0.13306	0.15527	*
STDMDC9	0.2009	0.07256	0.04769		0.07002	0.09235	*
STDMDC10	0.2228	0.06064	0.03364		0.05621	0.07878	*
STDMDC11	0.2255	0.03170	0.01803		0.04057	0.06311	*
STDMDC12	0.0535	0.03733	0.01664		0.03847	0.06030	*
STDMDC13	0.1580	0.07199	0.03998		0.06243	0.08487	*
STDMDC14	0.0675	0.03193	0.01463		0.03645	0.05828	*
STDMDC15	0.0384	0.01670	-0.00572		0.01587	0.03747	
STDMDC16	0.2187	0.03806	0.01531		0.03794	0.06056	*
STDMDC17	0.1454	0.04651	0.02768		0.04939	0.07110	*
STDMDC18	0.1747	0.03131	0.00808		0.03033	0.05258	*
STDMDC19	0.0850	0.07227	0.05483		0.07636	0.09789	*
STDMDC21	0.1932	0.03238	0.01805		0.03997	0.06190	*
STDMDC22	0.0084	0.00307	-0.01873		0.00275	0.02423	
STDMDC23	0.1775	0.03523	0.01170		0.03415	0.05660	*
STDNADM	0.4584	0.32316	0.28332		0.30548	0.32764	*
STDPADM	0.3873	0.30229	0.27582		0.29753	0.31924	*
GEO_2	0.0411	#-0.00982	-0.03603	#	-0.01422	0.00760	
SSRI	-0.0300	# 0.00239	-0.02236		-0.00072	0.02091	
SUBCLIN	-0.1453	-0.13979	-0.15517		-0.13348	-0.11180	*
SWIT_AUG	0.0842	#-0.00565	-0.02088		0.00086	0.02261	

Exhibit 3 - Final (3439 Patients, 1990-94)
Incremental Charges and Confidence Limits following exp() and smearing

	ALL	N=3439			SSRI	N=2406			TCA	N=1033		
	Low	Elast	High		Low	Elast	High		Low	Elast	High	
ssri	-280	-9	256									
age	-4	8	21		-1	13	27		-24	-2	21	
gender	-412	-127	152		-230	69	361		-827	-382	47	
dep_dx1	358	650	953	*	265	591	929		-12	421	874	
dep_dx2	513	889	1284	*	525	955	1410		-106	437	1022	
dep_dx3	251	510	774	*	105	385	671		-6	398	817	
dep_dx4	-563	-132	333		-760	-261	288		-705	-88	598	
dep_dx5	618	2305	4476	*	-202	1590	4032		33	2449	5985	
dep_dx6	-259	-783	-994	*	-146	-569	-784		-282	-1010	-1341	
no comorbidity	-2957	-3647	-4125	*	-2039	-3077	-3762		-1431	-3159	-4070	
stdmdc1	1831	2194	2571	*	1563	1978	2410		1764	2316	2895	
stdmdc2	201	632	1091	*	62	550	1077		63	752	1514	
stdmdc3	363	614	865	*	323	603	884		39	453	868	
stdmdc4	821	1110	1405	*	773	1095	1425		598	1070	1560	
stdmdc5	1124	1412	1706	*	1009	1327	1651		960	1412	1876	
stdmdc6	1021	1316	1617	*	1065	1400	1743		483	939	1409	
stdmdc7	190	742	1343	*	30	638	1311		183	1109	2180	
stdmdc8	1254	1505	1755	*	1194	1474	1754		864	1276	1687	
stdmdc9	541	795	1049	*	432	715	998		354	767	1182	
stdmdc10	405	682	964	*	309	614	925		264	713	1175	
stdmdc11	253	579	917	*	371	746	1137		-132	363	885	
stdmdc12	482	1177	1951	*	705	1545	2500		-470	442	1515	
stdmdc13	494	779	1070	*	367	677	994		232	678	1140	
stdmdc14	626	1697	2960	*	107	1197	2506		918	2960	5681	
stdmdc15	-486	1625	4677		-275	2417	6606		-2308	-149	3661	
stdmdc16	241	611	999	*	337	763	1215		-268	277	862	
stdmdc17	795	1495	2271	*	592	1361	2226		597	1792	3221	
stdmdc18	143	551	984	*	-44	418	912		48	681	1373	
stdmdc19	1221	1629	2001	*	919	1392	1820		1166	1868	2463	
stdmdc21	297	675	1072	*	351	781	1236		2	615	1281	
stdmdc22	-1475	261	2799		-1818	326	3862		-2083	12	3541	
stdmdc23	134	391	649	*	15	299	584		168	591	1018	
stdnadm	9964	11244	12630	*	9580	11036	12635		6149	7711	9480	**
stdpadm	10724	12149	13699	*	9422	11062	12884		8384	10176	12200	
geo_2	-644	-247	129		-617	-149	288		-969	-359	207	
subclin	-1794	-1556	-1315	*	-1718	-1448	-1171		-1802	-1408	-1009	
swit_aug	-577	25	699		-653	98	967		-580	270	1259	