Quantile Regression in Pharmaceutical Marketing Research

George Mu, IMS Health; Yong Cai, IMS Health; Yi Han, Ipsen Pharmaceuticals, Inc

ABSTRACT

In pharmaceutical marketing research, healthcare data heterogeneity presents a challenge to researchers. It is not always possible to identify a simple function that fits well for all individuals data points. Quantile Regression (QR) is an efficient and robust way to tease out the different patterns existing in the healthcare market. This paper demonstrates the value of applying QR to solve pharmaceutical marketing research problems. To illustrate the methodology, we use SAS® QUANTREG procedures to compare physicians' new products uptake patterns. Quantile regression, when applied to the right studies, can distill more market insights from the same data sets.

INTRODUCTION

Quantile regression was first introduced by Koenker and Basset in their econometric paper "Regression Quantiles" in 1978. Since then, it has gradually drawn attentions from both academic and practitioners. In essence, QR is a statistical model estimating conditional quantile functions given observed covariates. QR is especially useful to inspect the rate of change by quantiles.

In recent years, QR has been applied to different fields. Koenker and Bilias (2001) used QR to study reemployment bonus in micro economics. Levin (2001) applied QR analysis in education to study the class size and peer effects. There are lots of literatures of QR application in ecology as well, for example, Cade and Noon (2003), Chamaillé-Jammes and Blumstein (2012) and Lemke K. (2012).

The emerging of specialized QR software enabled researchers to apply QR methodology conveniently solving real world problems. SAS Inc. released the QUANTREG procedure in SAS/STAT 9.1 (Chen 2005). In analogy to REG procedure which fits OLS regression, QUANTREG is a comparable module for fitting Quantile Regression.

In this paper, we demonstrated how to use SAS QUANTREG procedure to build QR models for pharmaceutical marketing research. The solutions from QR can help researchers gain more market insights and make informed business decisions.

PHARMACEUTICAL MARKETING APPLICATION

Pharmaceutical industry captures data from various sources such as pharmacies, doctors' offices, health insurance companies etc. Many data are naturally suited for QR analysis. For example, physicians can be grouped into deciles by the number of prescriptions they wrote; patients can be segmented by the degree of compliance, etc. It is often of interest to see if physicians' prescribing patterns are different by deciles; and which factor effectively drives patients' medication compliances. QR provides a convenient tool to tease out heterogeneity and to compare patterns by quantile group. And it is also a valuable tool that can be added to market researchers' daily toolbox.

Some practitioners would think a two-step approach such as segment-then-regression would yield similar results. But it can be dangerous and problematic. Koenker and Hallock (2001) already pointed out this faulty notion that QR could be achieved by segmenting the data and building unconditional ordinary least square regression on the subsets. It might yield "disastrous results" due to the improper data truncation on the dependent variable. Heckman (1979) provided more theoretical discussions along the same line. Unlike aforementioned regression-by-segment method, QR utilizes all data to estimate quantile functions and therefore avoid the sample selection bias. Another advantage of QR is that it provides robustness and efficient estimates under non-Gaussian conditions (Buhai 2005).

QUANTILE REGRESSION BASICS

Linear regression estimates the mean response for a given set of predictor variables. It addresses the overall importance of predictor variables to the response variable. QR approach has significant advantages in investigating the relative level of influences for predictor variables. As an extension to linear conditional mean regression, QR is estimated by minimizing the asymmetrically weighted absolute residuals at given quantile (Chen 2005):

$$\hat{\beta}_\tau = argmin_\beta \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta)$$

where, ρ_τ is a piecewise linear loss function. For $\tau \in [0,1]$ and $I(\cdot)$ being the indication function,

$$\rho_\tau(u) = u(\tau - I(u < 0))$$

$\hat{\beta}_\tau$ is the estimated τ-th regression quantile. The interpretation is similar to the coefficients in simple linear regression: one unit change of x will cause β unit change of y in τ-th quantile.

The above formulas can also be expressed as:

$$\hat{\beta}_\tau = argmin_\beta \left\{ \sum_{i|y_i \geq x_i'\beta} \tau|y_i - x_i'\beta| + \sum_{i|y_i < x_i'\beta} (1-\tau)|y_i - x_i'\beta| \right\}$$

At $\tau = 0.5$, this is the median model, i.e.

$$\hat{\beta}_\tau = argmin_\beta \sum_i |y_i - x_i'\beta|$$

QR APPLICATION TO INVESTIGATE EARLY ADOPT PHYSICIANS

The cholesterol and triglyceride regulating drug Crestor™ was launched in US market August, 2003. It was interesting to see how quickly doctors started to prescribe Crestor post launch and how steep the uptake pattern was. In this example, we investigated key contributing factors based on Rx data from top 5,000 Crestor prescribers within the first 24 months post launch. The data were captured from prescription transactions through retail pharmacies.

Following quantile regression code was used to analyze early adopter patterns:

```
ods html;
ods graphics on;
proc quantreg data=crestor4 order=FREQ;
   class gender spec;
   model Rx = time time_sq gender spec/ quantile = 0.05 to 0.95 by 0.1
         seed=1812 plot=quantplot;
run;
ods graphics off;
ods html close;
```

Table 1 presents one of the standard outputs from SAS at quantile 0.55 as an example. The outputs are similar to simple regression with parameter estimates and typical statistical tests.

Quantile and Objective Function	
Quantile	0.55
Objective Function	636617.8017
Predicted Value at Mean	23.1648

Parameter Estimates								
Parameter	Levels	DF	Estimate	Standard Error	95% Confidence	Limits	t Value	Pr > |t|
Intercept		1	-4.5779	0.0298	-4.6363	-4.5195	-153.56	<.0001
Time		1	4.1018	0.0079	4.0863	4.1173	517.88	<.0001
Time_sq		1	-0.1113	0.0004	-0.112	-0.1105	-280.98	<.0001
Gender	M	0	1.4026	0.0536	1.2924	1.5129	24.93	<.0001
Gender	F	1	0	0	0	0	.	.
SPEC	CARD	1	0.5873	0.0235	0.5412	0.6335	24.97	<.0001
SPEC	OTHR	1	0.5873	0.0721	0.446	0.7287	8.14	<.0001
SPEC	PCP	0	0	0	0	0	.	.

Table 1. Parameter estimates at quantile = 0.55

In Figure 1, the full spectrum of quantile plot is presented. Time as a covariate is adjusted by model. Among all variables, Gender showed big rate of change in physicians' new drug adoption. The male doctors at high quantile adopted Crestor much faster than the rest of the doctors. Compare to PCP's, Cardiologists and other physicians tended to adopt Crestor in a much fast rate at high 80% to 100% quantiles.

At low quantiles, PCP and non-cardiologists specialties act similarly. Lower quantile physicians had similar adoption patterns within each specialty group. Quantile analysis reveals that Non-PCP's at high Rx Quantiles and all Cardiologists were the main driving force to Creators adoptions. Interestingly, male doctors at high quantile also adopted Crestor at a much faster rate.
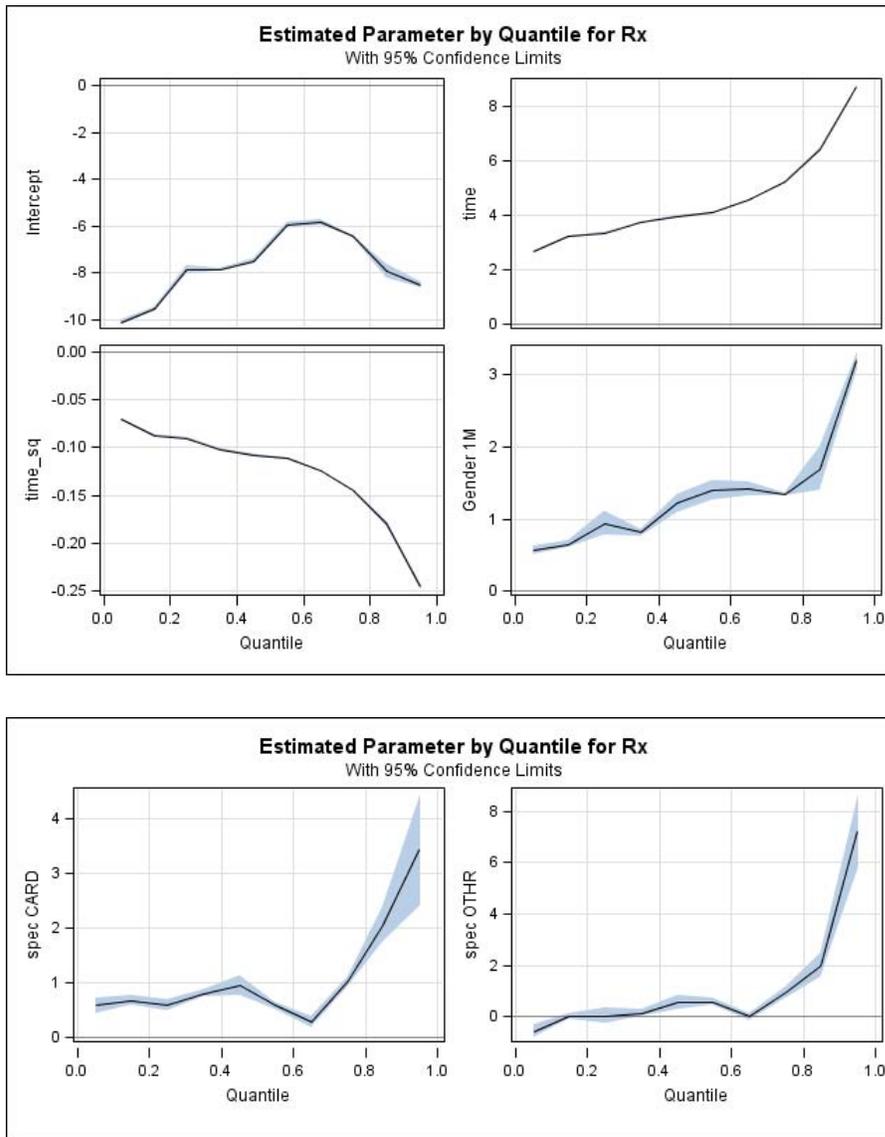

Figure 1. Quantile plot for time, specialty and gender effects

Table 2 compares modeling results between QR and OLS regression. OLS parameter estimates are listed in the second column followed by the QR estimates at 0.05 incremental quantile.

The OLS estimates only reflected the average impact at aggregated level. Comparatively, QR was able to differentiate the uptake pattern among physicians with different practice size. For example, "spec-OTHR" had positive impact on average but QR showed at 0.05 quantile the impact was actually negative with value -0.57. It is clear that QR can provide better understanding of the underlying relationships than OLS.

Parameter	OLS	QR (0.05)	QR (0.15)	QR (0.25)	QR (0.35)	QR (0.45)	QR (0.55)	QR (0.65)	QR (0.75)	QR (0.85)	QR (0.95)
Intercept	-9.81	-10.1	-9.54	-7.88	-7.87	-7.48	-5.98	-5.85	-6.45	-7.93	-8.50
Time	4.90	2.67	3.23	3.35	3.73	3.96	4.10	4.56	5.26	6.41	8.74
Time_sq	-0.14	-0.07	-0.09	-0.09	-0.10	-0.11	-0.11	-0.12	-0.14	-0.18	-0.25
Gender-M	2.57	0.56	0.65	0.94	0.81	1.22	1.40	1.41	1.34	1.69	3.21
Gender-F	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Spec- CARD	1.44	0.59	0.68	0.60	0.81	0.94	0.59	0.29	1.02	2.06	3.43
Spec- OTHR	1.51	-0.57	0.00	0.03	0.14	0.55	0.59	0.00	0.93	2.00	7.23
Spec-PCP	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 2. Comparison to OLS parameter estimates

CONCLUSION

Healthcare marketing research data are not always well distributed. Simple mean or median models may not always correctly identify key drivers for the events under investigation. The example in this paper demonstrated that QR could pick up key covariate undetected by OLS. We recommend investigators to pay extra attention to the data distributions. QR technique can be a valuable tool to provide extra insights for both model construction and model diagnosis.

Disclaimer

All the views expressed in this publication are personal opinions and do not convey any views from IPSEN Biopharmaceuticals, Inc. and/or IMS Health.

REFERENCES

Buhai, I. Sebastian, 2005 "Quantile Regression: Overview and Selected Applications." Ad Astra. 4, 2005, 1-17.

Cade, B. and B. Noon, 2003 "A Gentle Introduction to Quantile Regression for Ecologists." Frontiers in Ecology and the Environment, 1, 412-420.

Chamaillé-Jammes, S. and Blumstein, D. 2012. "A case for quantile regression in behavioral ecology: getting more out of flight initiation distance data." Blumstein Behavioral Ecology and Sociobiology. 66 – 6: 985-992

Chen, Colin. 2005. "An Introduction to Quantile Regression and the QUANTREG Procedure." SAS. SUGI. 213-230.

Heckman, J. J. 1979. "Sample Selection Bias as a Specification Error." Econometrica. 47-1: 153-61.

Lemke K. 2010 "Modeling Healthcare Expenditures Using PROC QUANTREG", NESUG 2010, 1-7

Levin, J. 2001, "For whom the reductions count: A quantile regression analysis of class size and peer effects on scholastic achievement", Empirical Economics 26, 221–246.

Koenker, R. and Bassett, G.W. 1978. "Regression Quantiles." Econometrica. 46: 33-50.

Koenker, R. and Hallock, K. 2001. "Quantile Regression." Journal of Economic Perspectives. 15: 143-156.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

George Mu
IMS Health Inc.
1 IMS Drive,
Plymouth Meeting, PA 19462
Work Phone: 484-567-6214
Fax: 484.567.6984
E-mail: gmu@us.imshealth.com
Web: http://www.imshealth.com/portal/site/ims

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.