THE PHILIPPINE REVIEW OF ECONOMICS
VOL. XLVI NO. 1 JUNE 2009 PP. 189-219

School inputs and student performance in public elementary schools in Palawan: a quantile regression analysis*

Jan Carlo B. Punongbayan**
University of the Philippines School of Economics

This study investigates the role of school resources in different measures of student performance in public elementary schools in the province of Palawan. We contend that it is not enough to identify which school resources matter the most, but that it would be more informative for policy purposes to identify which student types may benefit the most from the provision of a given school resource. This way, we may be able to target our allocations toward more productive educational investments. Using quantile regression analysis, we find that in the case of Palawan, improvements in pupil-teacher and pupil-toilet ratios may benefit high-performing schools the most. We also find that class size and pupil-room ratio improvements, along with the provision of guidance counselors and science laboratories, may benefit low-performing schools the most. Our results also give some evidence that conventional ordinary least squares (OLS) procedures may be both insufficient and imprecise in estimating education production functions, and that educational policies based on least squares methods alone may be misguided if not accompanied by other techniques, such as quantile regression, which can offer more valuable insights into education production processes in general.

JEL classification: I21, I28, C19
Keywords: Quantile regression, academic performance, education resources

---

* This paper is based on the author's undergraduate thesis which won the Gerardo P. Sicat Award for Best Undergraduate paper.

**Jan Carlo B. Punongbayan is valedictorian and summa cum laude graduate of the University of the Philippines School of Economics, Class of 2009.

1. Introduction

However important education is in determining the long-run economic growth and development of a country, the Philippines' public educational system is continually deteriorating. Whether in terms of dropout, participation, and achievement rates, the country's educational indicators are seriously lagging behind those of its Asian neighbors. Although the Department of Education (DepEd) receives the lion's share in the annual national budgets, the lack of a sound population policy has spread public resources so thinly that per-student allocations cannot effectively lead to adequate instruction and learning. And notwithstanding the severe shortages and strain on public funds, corruption and misallocation of educational resources have worsened the local education scene.

A more efficient allocation on the part of government may help alleviate the severe shortages of educational resources. Furthermore, while school inputs matter, not all of them may be of equal significance or potency in affecting student performance. Indeed, it may be good for the government to spend its funds not on every conceivable educational investment that comes along, but only on those that can be shown to improve student performance the most. Thus, there may be a case for pinpointing not only which of these resources contribute the most in improving student outcomes, but also to which types of students they matter, and then focusing educational spending on them.

This study aims to investigate the nature of school inputs (particularly school resources) in the Philippine public education system by investigating the case of the province of Palawan, for which both school and community input data are available at the school and barangay levels. Figuring out which students may benefit the most from a given resource is addressed by using the emergent quantile regression method. By investigating the role of resources in the learning process, this study aims to give potential policy advice for a more efficient allocation of educational resources, as well as suggestions for further study, toward improving educational outcomes in the country.

2. Review of literature

There has been a long-standing debate on whether school resources matter in shaping educational outcomes as measured by student performance. Specifically, because of its political appeal as a convenient policy tool, the issue of class size has been the focus of many empirical studies. Linking several educational inputs and measures of student performance in education production functions, Hanushek [1998] leads the studies finding class size as vital. On the other hand, using revised methodologies, Krueger [2003] and Hedges, Laine, and Greenwald [1994] lead studies finding little significance in class size. Experimental studies, as well as evidence from different continents, also find mixed results for other school inputs.

Nevertheless, many studies find that the provision of certain aides, facilities, and other school resources (holding for family background and individual characteristics) seems to benefit poor-performing students the most, as well as those from minorities, developing countries, and lower socioeconomic classes. In the Philippines, Quimbo [2003] finds that provision of basic learning materials and strengthening parental education, among others, may significantly boost science and math performance of public school students. Bacolod and Tobias [2006], using rich data from the Cebu Longitudinal Health and Nutrition Survey, also find that individual characteristics, as well as the provision of even the most basic school facilities like electricity, may matter more than either class size or teacher training in improving math test scores.

To date, only one study has analysed education production by means of quantile regression in the Philippines. Using nationwide National Achievement Test (NAT) and National Diagnostic Test (NDT) data from the DepEd for the school year 2002-2003, Orbeta [2008] finds in a recent working paper that while class size and pupil-seating ratios are not a significant determinant of student achievement, pupil-teacher ratios are important for low-performing public elementary schools, and pupil-classroom ratios are important for high-performing high schools. However, due to time constraints, the said study was not able to control for either household or community characteristic variables (e.g., poverty incidence).

Before Orbeta's [2008] study, quantile regression was already used in three education production studies in the United States. By using socioeconomic and school-input variables, as well as data on the standardized test performance of primary and high school students, the three earlier studies found that alternative inputs, such as peer group effects, school enrollment, and parental educational attainment, not only have an impact on students in general, but also have varying impacts on students along different levels of achievement—a finding that conventional least squares analyses in the past have missed.

3. Conceptual framework

At the heart of education economics lie education production functions. Analogous to microeconomic firm theory, they link various possible inputs to the learning process with measures of educational outcomes. In general, education production functions are of the form

$$y = y(i, f, s) \tag{1}$$

where y is the measure of student achievement, i is a vector of individual traits, f is a vector of family and social characteristics, and s is a vector for school inputs.

Todd and Wolpin [2003] recommend that education production models view student outcomes as the result of a cumulative process, and that family and school inputs be incorporated in models as historical rather than contemporaneous inputs. However, due to the unavailability or rarity of such historical data especially in developing countries, most studies have had to use contemporaneous data in regressions instead. Possible resultant problems such as omitted variable bias may be mitigated by using lagged dependent variables, where achievement data are available for at least two periods (as in Bacolod and Tobias [2006]).

Educational outcomes vary greatly, like aptitudes in leadership, creativity, or music. For research purposes, however, these are inferior to test score data, which are easier to collect and analyse. Educational inputs, meanwhile, come in even more varied forms, under three major categories: (1) individual characteristics, (2) social and family background, and (3) school inputs. The spatial and temporal variability of such inputs largely explains the observed differences in student achievement. Such variability also explains why there is no consensus among education researchers on which group of factors is most important in determining outcomes in education.

Amid the varying interpretations, however, a vast majority of the researchers on education production have solely relied on ordinary least squares (OLS), investigating the effects of school resources on performance on average. In effect, they have discarded any possible effects of school resources not only on students in general, but also on different types of students, i.e., high- versus low-performing students. In other words, they have focused solely on estimating the effects of school inputs on the conditional mean function, when they could very well have varying effects at different points along the conditional test score distribution. Therefore, investigating the conditional effects (and not only the average effects) of school inputs can provide a richer analysis on the educational input-output process than is offered by conventional OLS procedures, as is commonly used. This is exactly what is addressed by the emergent technique called quantile regression.

4. Methodology and empirical strategy

This section discusses quantile regression, the sources and limitations of the data, the empirical methodology and regression model used, and the variables employed in this study.

4.1. Quantile regression

Virtually all past studies on the effects of school resources on student performance have solely relied on classical linear regression, sometimes along with its derivatives like two-stage least squares, instrumental variables, or weighted OLS. Least squares methods are in fact already able to explain much about educational production, by estimating the magnitude and direction of different parameter effects—like those of class size or textbooks—on student performance.

However, OLS ignores the fact that school inputs may affect not only students on average (since OLS estimates only the conditional mean function), but also students at different points of the conditional achievement distribution. It may be inefficient, for instance, to provide computers to all students when in fact some specific segment of the distribution would benefit the most from such provision, thus deserving a more targeted distributional approach.

To this end, we use instead the emergent technique of quantile regression (QR), which promises to offer a richer analysis than OLS regarding education production. The seminal work of Koenker and Bassett [1978] on QR is considered among the major breakthroughs in econometrics in the past three decades. In fact, according to Koenker and Hallock [2001], QR is steadily gaining ground in several fields like labor economics, ecology, and finance.

Unlike classical linear regression models, which estimate only the conditional mean function, quantile regression models are able to simultaneously estimate the entire range of "conditional quantile functions," allowing the researcher to infer the effects of covariates on different segments of the dependent variable's distribution. Thus, while OLS rigidly assumes that the effect of a covariate is constant across the entire distribution of the dependent variable, QR is able to give a fuller picture by estimating a set of parameters catering to the different quantiles of the response variable. It is in light of this richer insight that QR is viewed as a "natural extension" of OLS, and also why QR is popular among studies where the focus is not really on the mean effects but only on specific quantiles, especially the extremes.

Koenker and Bassett [1978] formally introduced regression quantiles as follows:

$$Y_i = X_i \beta_\theta + u_{\theta i} \qquad Quant_\theta(Y_i \mid X_i) = X_i \beta_\theta \tag{2}$$

where the θ^{th} conditional quantile of Y given X for individual i is $Quant_\theta(Y_i \mid X_i)$, noting that $0 < \theta < 1$. In other words, error terms are given different weights along the Y distribution, where $\theta\%$ of errors will be negative and $(100 - \theta)$ % will be positive, and QR estimates functions for different levels of θ. It is further assumed that the error term $u_{\theta i}$ satisfies the quantile restriction $Quant_\theta(u_{\theta i} \mid X_i) = 0$.

In this framework, the OLS function—which estimates only the conditional mean function (where u_i is assumed to be homoskedastic)—is simply derived by removing the θ parameter to get $Y_i = X_i \beta + u_i$. Alternatively, OLS can be thought of as the "summary" of all quantile effects, such that $\int Quant_\theta (Y_i \mid X_i) = E(Y_i \mid X_i)$. It is due to this aggregation that OLS fails to report important variable relationships at the different quantiles, which QR could otherwise have detected.

The objective of QR also differs from that of OLS in that QR minimizes not the sums of squared residuals, but the absolute values of the residuals instead:

$$\min_{\beta} \frac{1}{n} \left\{ \sum_{i:Y_i \geq X_i'\beta} \theta \left| Y_i - X_i'\beta \right| + \sum_{i:Y_i \geq X_i'\beta} (1-\theta) \left| Y_i - X_i'\beta \right| \right\} \tag{3}$$

whose solution is the θ^{th} regression quantile of Y. As such, one can "explore," as in Koenker's words, the marginal effects of the covariates on the Y distribution at any given percentile by changing θ, to see whether there is homogeneity in the effects, as OLS assumes. Buchinsky [1998] established the minimization solution using algorithms and so-called bootstrapping methods.

Rangvid [2003] summarizes the advantages of using QR over OLS. By virtue of minimizing absolute deviations instead, QR estimates are robust to heteroskedasticity compared to OLS parameters (which require homoskedasticity). Extreme observations or outliers of the dependent variable also have a weaker impact on QR coefficients. Most important, however, QR relaxes the rather rigid assumption of OLS that the parameters are constant across the dependent variable distribution.

To date, there have been three pathbreaking studies explicitly investigating the effect of school resources on student outcomes using quantile regression analysis. Eide and Showalter's [1998] pioneering work finds that contrary to OLS estimates of previous studies, QR revealed for the first time that in fact they do matter at some segments of the conditional performance distribution.

Meanwhile, Levin [2001], studying Dutch primary school students, proposes that reducing class size may raise test scores only if accompanied by changes in class and peer composition, which were found to affect students at the lower end of the distribution the most. Lastly, Bassett, Tam, and Knight [2002] find that pupil-teacher reductions benefit low-performing students but *harm* high-performing ones, among other differential effects across the quantiles.

By use of quantile regression analysis, the present study aims to similarly determine not only whether school inputs do matter in determining public elementary student performance, but also for whom such inputs matter, in the hope of improving the allocation of educational resources and ultimately the achievement of elementary pupils in the Philippines.

4.2. Data and estimation

This study uses elementary school-level data on NAT scores (as well as school resources) in the school year 2005-2006 for the province of Palawan. Both data sets were obtained from the DepEd's Research and Statistics Division (RSD). School input data comprise a wide range of school-level characteristics, including class size, pupil-teacher ratio, and dummy variables for facilities like libraries and laboratories.

Acknowledging that performance is also a function of factors aside from school inputs, we also incorporate barangay-level data for each of the elementary schools for the year 2005. Such data were obtained from the Community-Based Monitoring System (CBMS), a constituent network of the Poverty and Economic Policy Research Network (PEP) based in Manila. As data collection is still ongoing nationwide, the data were complete for only a handful of provinces. It is for this reason that we specifically investigate only the province of Palawan, for which provincewide data on a wide range of community variables and data on school characteristics (which is the limiting dataset) are available and most complete.

We recognize, however, that not all barangays have elementary schools, that not all schools have submitted data on their NAT performance, and that not all schools have data on their school resources. Our data therefore include only schools for which data on all criteria are complete. Considering, however, that public schools are required by the DepEd to submit data yearly, and that there is little reason to believe that some schools are less likely to report their data than others (because the DepEd does not easily give out its data, anyway), the missing data are more likely the result of physical and financial constraints (such as difficulties in the transmission of data from far-flung areas) than nonrandom events (such as outright refusal or reluctance to submit data).

Since the probability of nonsubmission is not likely related to the values of the actual data, using the criterion of Howell [2002], such incomplete data may be treated as "missing completely at random" (MCAR) and may be safely deleted in the regressions. Although Howell admits that the research design might "lose power," the absence of some data may not necessarily lead to biased results. Imputation, of course, would be a questionable recourse and hence not resorted to.

We further assume that barangay-level characteristics belong to the elementary school/s found within its bounds; that students attend the school found within their barangay; and that in the event of a barangay having more than one school, the same barangay characteristics apply for those other schools. Despite these limiting assumptions, the core insight—that the general NAT performance of a school depends both on its educational resources and the characteristics of the community in which it is found—cannot be easily discounted.

With the help of a unique identification system for the schools, the NAT and school characteristics data were reconciled. The resultant dataset was then merged with the community-level data using another identification system from the CBMS. Only observations comprising complete data for the achievement, school input, and community-level variables were included in the study, resulting in a final observation count of 251 schools.

We estimate education production functions, linking school and community variables with mean total NAT performance, using OLS first and then QR. We then extend the cross-section analysis by estimating the effects of school inputs on each of the different subject components of the NAT exam: English, Mathematics, Science, Filipino, and Hekasi; both OLS and QR will also be used. We do this to determine whether any differing trends would emerge by using the component subjects of the NAT examination, than if we only use mean total performance. This way, we are able to trace not only the average effects of school resources on student achievement, but also whether they have differential effects across the entire range of quantiles of the achievement distribution, using mean total as well as mean subject performance measures.

4.3. Analytical model

With the provincewide data at hand, we specify our QR model as follows:

$$Y_i = \beta_{0\theta} + \beta_{i\theta}\sum S_i + \gamma_{i\theta}\sum C_i + \varepsilon_{i\theta} \tag{4}$$

where: Y_i is the NAT performance of school i for SY 2005-2006,
S_i is a vector of school inputs or characteristics,
C_i is a vector of community characteristics,
β_θ and λ_θ are parameters to be estimated, and
ε_i is an error term.

The variables for school performance, school characteristics, and community characteristics used to estimate the model are listed in Table 1 along with their definitions. The OLS model used can be obtained by simply removing the θ parameters. Also, for the different subjects under the NAT, we use $SCIPREV_i$, $MATHPREV_i$, $ENGPREV_i$, $HEKPREV_i$, and $FILPREV_i$ in place of $TOTPREV_i$ for the Science, Mathematics, English, Hekasi, and Filipino mean scores of school i, respectively.

While the model contains community-level variables aside from contemporaneous school-input variables, there are still other historical and unobserved exogenous variables (including family and individual inputs) that the model at hand might have missed. It is for this reason that we include a lagged dependent variable (LDV) of the test and test-component score variables for the previous school year 2004-2005 in order to account for such unobserved factors.

In the next section, we first estimate the educational production function using OLS, then compare it with the QR results. To infer the different quantile effects, we choose the 10th, 25th, 50th, 75th, and 95th quantiles (for which θ = 0.10, 0.25, 0.50, 0.75, and 0.95, respectively). Then we investigate each input to see whether its effect changes along the different conditional test score distributions, an analysis that goes unreported with conventional OLS methods.

5. Results and discussion

This section presents the regression results for both OLS and QR methods, using mean total NAT scores as well as per-subject NAT scores as measures of student performance. We present which school inputs matter for which schools, and discuss several recurring insights from the analysis.

5.1. Regression results

Table 1 presents both the OLS and QR results using mean total achievement (incorporating all five subject areas) as the measure of student performance, showing the QR coefficient estimates for the 10th, 25th, 50th, 75th, and 95th quantiles.

Table 1. Definition of variables*

Variable	*Name*	*Definition*
TOT_i	Mean performance score (MPS)	Simple average of the aggregated mean scores of all five subjects in a given NAT-year for school *i*
T_i	Pupil-teacher ratio	Total enrollment over total number of teachers in school *i*
RA_i	Pupil-room ratio	Total enrollment over total number of classrooms in school *i*
EA_i	Pupil-seating ratio	Total enrollment over total number of seats, including tables, chairs, desks, and armchairs for school *i*
OA_i	Pupil-toilet ratio	Total enrollment over total number of toilet bowls and urinals in school *i*
$LOCPER_i$	Percent of locally funded teachers**	Percent of locally funded teachers over total teaching force of school *i*
C_i	Class size	Total enrollment over total number of classes or sections in school *i*
$TOTPREV_i$	Previous MPS	MPS of school *i* for the previous school year 2004-2005
$DUMGUID_i$		Dummy for the presence of a guidance counsellor
$DUMLAB_i$		Dummy for the presence of a science laboratory
$DUMLIB_i$		Dummy for the presence of a library
$DUMMED_i$		Dummy for the presence of a medical/dental clinic
$DUMCAF_i$		Dummy for the presence of a cafeteria within school *i*
$DUMPRINC_i$		Dummy for the presence of a principal
$MALNUT_i$	Child malnutrition rate	The proportion of under-5 children suffering from malnutrition in the barangay of school *i*

Table 1. Definition of variables* (continued)

Variable	*Name*	*Definition*
$SAFEWATER_i$	Access to safe water	The proportion of households in the barangay of school *i* having access to safe water supply
$SAFETOILET_i$	Access to sanitary toilets	The proportion of households in the barangay of school *i* with access to sanitary toilet facilities
$DEATHS_i$	Child deaths	The proportion of children in the barangay of school *i* who died before their 5th birthday
$PREG_i$	Maternal deaths	The proportion of women who died during pregnancy in the barangay of school *i*
$POVTHRESH_i$	Poverty incidence	The proportion of households in the barangay of school *i* with an income below the poverty threshold at the time
$LITRATE_i$	Literacy rate	The proportion of functionally literate persons aged 10 and above, a proxy for innate abilities in the barangay of school *i*
$ENROLL_i$	Participation rate	The percentage of children aged 6 to 12 enrolled in a school for the barangay containing school *i*, an indicator of the community's valuation of the importance of education

*Based on the 2006 "Glossary of commonly used terms in education statistics" by the Inter-Agency Committee on Education Statistics (IACES).

**A teacher is considered "locally funded" if she is either a volunteer or is funded by the Local Government Unit (LGU), the Parent-Teacher-Community Association (PTCA), or through the Special Education Fund (SEF) of the province, city, or municipality.

Table 2. OLS and QR results, using mean total performance

Variable	OLS	QR .10	QR .25	QR .50	QR .75	QR .95
Pupil-teacher ratio	0.013	0.249	0.074	-0.037	-0.121	-0.366
Pupil-room ratio	-0.012	-0.055	-0.013	-0.034	0.012	0.068
Pupil-seating ratio	0.278	0.732	0.446	0.245	0.153	0.217
Pupil-toilet ratio	-0.005	0.005	-0.004	-0.007	-0.008	-0.011
Percent teachers locally funded	-2.69	3.666	-3.931	-7.927	-4.659	-11.119
Class size	-0.079	-0.074	-0.11	-0.014	-0.043	-0.056
Mean total score (lagged)	***0.252	*0.225	***0.253	0.212	***0.282	***0.3
Guidance counsellor (dummy)	9.734	**26.672	**20.73	**10.46	-0.765	***-8.715
Science laboratory (dummy)	-11.662	2.679	-8.571	*-10.065	**-18.476	**-23.48
Library (dummy)	0.379	*4.548	3.264	-3.223	-0.884	1.861
Medical clinic (dummy)	2.723	1.669	3.159	4.561	1.34	-4.592
Cafeteria (dummy)	2.187	1.894	-0.533	2.07	2	0.777
Principal (dummy)	1.853	-1.249	0.093	1.313	4.102	6.153
Percent children (0-5) malnourished	-0.055	-0.143	-0.029	-0.086	-0.002	-0.078
Percent households access to safe water	**-0.051	-0.032	-0.047	***-0.094	-0.018	-0.048
Percent households access to sanitary toilet	0.031	0.066	0.064	*0.064	-0.009	-0.086
Percent households below poverty line	*-0.072	-0.085	-0.099	-0.111	-0.052	-0.069
Percent persons (10+) literate	0.044	-0.04	0.096	0.07	0.028	0.064
Percent children (6-12) enrolled in school	-0.005	0.046	-0.075	0.002	0.004	-0.05
Constant	***43.205	25.495	**36.069	***47.035	***51.354	***71.343

*,**,*** Significant at α=10 percent, 5 percent, and 1 percent, respectively.

We note that OLS reports only a handful of statistically significant explanatory variables: the constant, lagged mean score, poverty incidence, and percent of households with access to safe water (which does not even have the correct expected sign). And neither do the estimates for the pupil-teacher ratio, pupil-seating ratio, and science laboratory dummy have the correct sign. Based on the OLS results, therefore, we are led to conclude that no school resource is in fact significant in affecting student achievement in Palawan, and that community-based variables may in fact be more important. Also, the spurious signs of some variables make the regression results all the more less convincing.

If, however, we go beyond OLS and examine the conditional quantile effects using quantile regression, we get a very different picture of education production indeed.

Figure 1 presents 20 panels of QR results, each for the 19 regressors and the constant. The broken line in each represents the mean OLS estimate for that variable, and its flatness denotes the OLS assumption that the parameter effects are constant all throughout the different quantiles of the response variable. The two faint dotted lines above and below each broken line represent the 90 percent confidence interval band for the OLS estimate. Meanwhile, the unbroken curves depict the coefficient estimates of QR (measured along the vertical axis) for each variable, along the entire range of quantiles (measured along the horizontal axis) across the dependent variable distribution.

The first thing one will notice in all the panels is that the QR parameter estimates greatly vary along the quantiles, in contrast to the constant effects estimated by OLS. One will also notice that in some of the panels there are striking trends or movements of the QR estimates along the quantiles, in contrast to the erratic trends in other panels.

Take, for example, the panel for the pupil-teacher ratio, denoted by "T" and found in the first row, second column. Starting from the 10th quantile and going to the right until one reaches the 95th quantile, there seems to be a uniformly downward trend, which can be interpreted by referring again to Table 1. It says there that OLS finds that ten more students assigned per teacher, on average, is associated with a 0.13 increase in the mean total performance of a school. QR, however, estimates that indeed this effect will not be constant among students, and that ten more students per teacher is associated with a 2.49 percentage-point improvement in mean total performance at the 10th quantile (or among low-performing students) but is associated with a decline in performance by as much as 3.66 percentage points at the 95th quantile (or among high-performing students). It appears, therefore, that higher-performing schools are made worse off by more pupils sharing a teacher, and may possibly benefit more (than low-performing students) from pupil-teacher ratio improvements, at least in Palawan.

Figure 1. QR coefficient estimates along the different quantiles, for each of the 19 explanatory variables

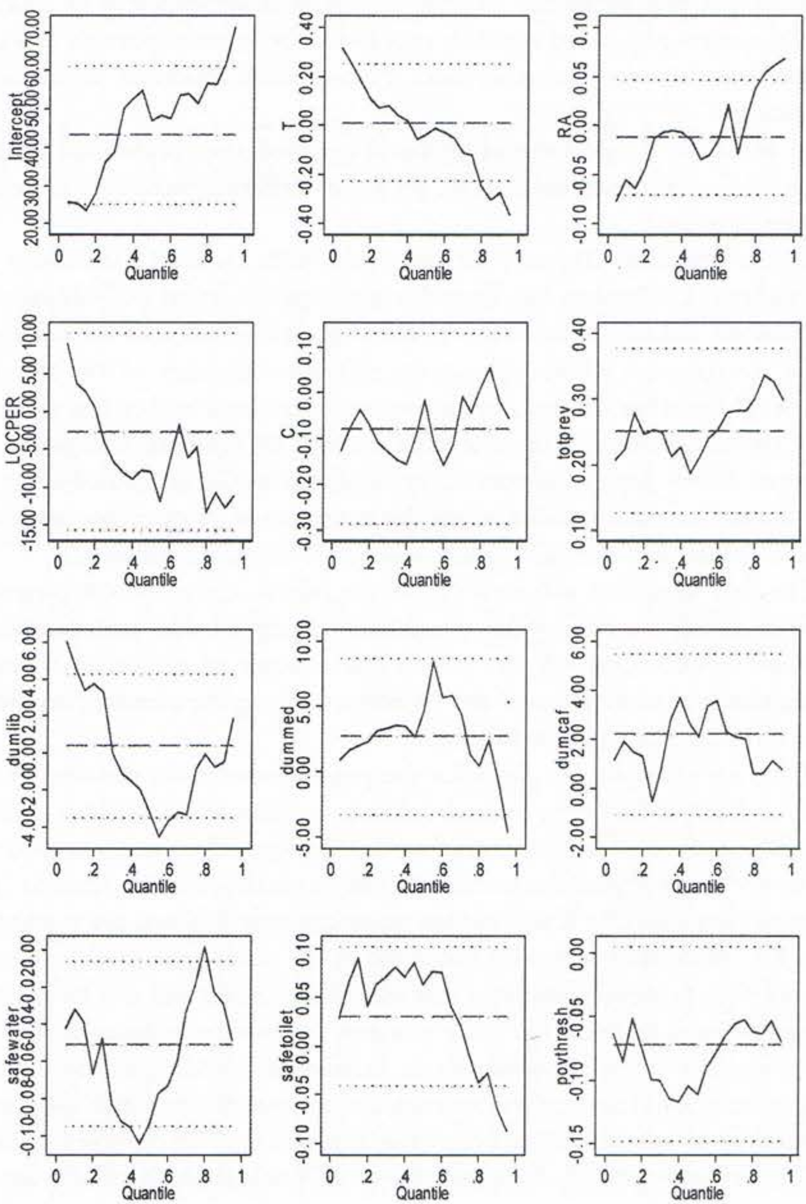
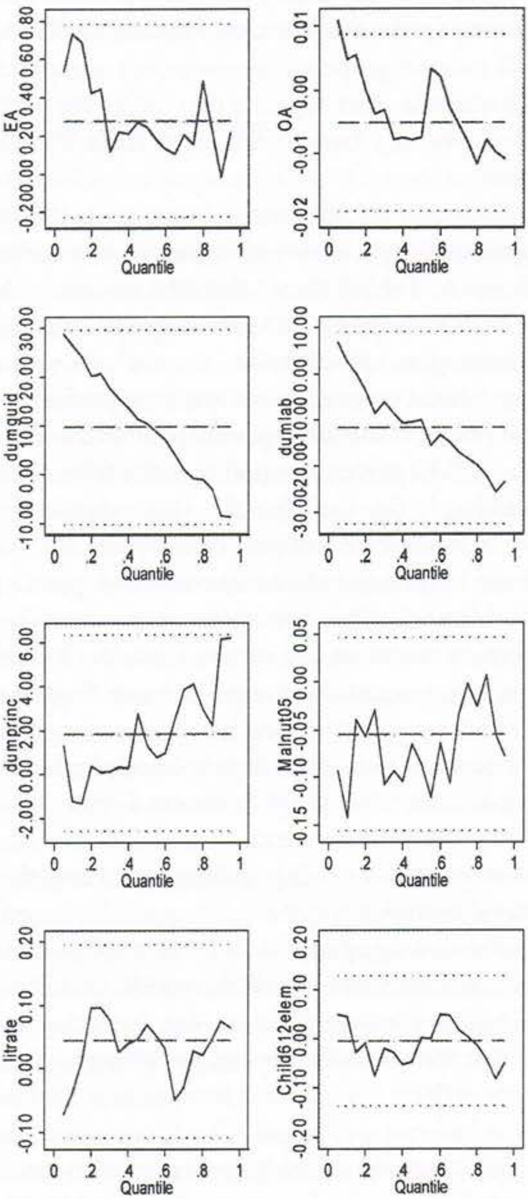


Figure 1. QR coefficient estimates along the different quantiles, for each of the 19 explanatory variables (columns 4 and 5 continued)



A similar line of interpretation may be applied to the other variables with rather striking trends in Figure 1. The pupil-room ratio panel (row 1, column 3) shows the opposite trend. Table 1 tells us that OLS associates ten more students per classroom with a 0.12 decline in mean total performance. QR, however, suggests that the effect will vary among student types: ten more students will result in a 0.5 percentage-point decline in performance at the 10th quantile but will result in a 0.68 percentage-point improvement among high-performing students at the 95th quantile. This suggests that unlike the pupil-teacher result, low-performing students may benefit the most from improvements in the pupil-classroom ratio.

Most surprising among the covariates, however, is the seeming singular importance of a guidance counselor in affecting student achievement (see the panel in row 2, column 4). Table 1 shows that OLS associates the presence of a guidance counselor with a whopping 9.73 percentage-point overall improvement in mean total performance. QR estimates, on the other hand, report even stronger effects: a guidance counselor can improve performance by as much as 26.67 percentage points at the 10th quantile, and at the same time decrease performance by about 8.72 percentage points at the 95th quantile.

Even more striking is the fact that the OLS estimate for the guidance counselor dummy is statistically insignificant, while the QR estimates are statistically significant in four out of the five selected quantiles. The panel in Figure 1 reinforces this finding by showing an unambiguously declining trend in the QR coefficient estimates as one moves from the lowest to the highest quantiles. These findings suggest that it may be more fruitful to prioritize the provision of guidance counselors among low-performing schools than high-performing ones. The same decision rule may be applied to the provision of science laboratories as seen in its panel in Figure 1 (row 2, column 5), albeit showing a less striking QR estimate trend.

The library dummy shows a rather peculiar trend. From the panel in row 3, column 1, it appears that students at the lower quantiles benefit from libraries the most (associated with a significant 4.55 percentage-point improvement at the 10th quantile) whereas students around the middle quantiles are made worse off by libraries (a 3.22 percentage-point decline in performance at the 50th quantile). The fact that OLS can only report, on average, a 0.38 improvement from library presence reflects the inherent limitation of OLS in revealing what could otherwise be significant and larger effects, as reported by QR.

The medical clinic dummy is also peculiar, in that according to row 3, column 2 in Figure 1, it appears to benefit students at the middle quantile the most (a 4.56 percentage-point improvement at the 50th quantile) while making students at the highest quantiles worse off (by as much as a 4.59 decline in mean performance at the 95th quantile). If one merely relied on the OLS effect of 2.72 improvement, on average, and immediately concluded that indeed clinics are important and that they have to be provided for all, one could have missed the even more important insight of prioritizing its provision among middle-performing schools first (because they may benefit from it the most) before any other school types.

5.2. Regression results by subject component

Using different subject components of the NAT as alternative measures of student performance, we find that certain school inputs may also matter differently for different quantiles of the student performance distribution.

First we use the Science component as the measure of student performance. While OLS reports that the presence of guidance counselors is statistically insignificant and results only in a 6.81 percentage-point mean Science improvement, QR reports estimates that are significant at the 10th, 25th, and 75th quantiles, with effects that can be as large as 20.03 and as low as –5.76 percentage points. As for child malnutrition, OLS reports that the parameter is insignificant even if it is actually significant at both the 25th and 50th quantiles.

Also, we find once more the seemingly anomalous result in which some school input dummies, particularly the guidance counselor and science laboratory dummies, have a negative sign instead of the expected positive, and that they are even significant for the better-performing schools (a significant –5.76 effect of guidance counselors at the 75th quantile, and a sizeable and significant –17.42 effect of science laboratories at the 95th quantile). It seems from the results that better-performing schools are "disadvantaged" by the presence of science laboratories.

Low-performing schools may benefit from the presence of a science laboratory insofar as it provides the basic equipment and learning environment expected from laboratory activities. High-performing schools, however, necessitate the provision of upgraded equipment and laboratory facilities, so that the "disadvantage" that accrues to them from the presence of publicly provided facilities may partly explain the "drawbacks" of not being able to learn in more modern facilities.

As for the significant yet sizeable negative effects of guidance counselors on better-performing schools, it is possible that low-performing schools benefit relatively more because they are associated with more behavioral and academic difficulties. Students in schools at higher performance quantiles may face less academic or behavioral difficulties as encountered by their counterparts in lower quantiles, hence making guidance counselors somewhat "redundant". Also, to the extent that teachers often substitute for guidance counselors (as is observed especially in provincial schools), their dual capacity as counselor and faculty member at the same time may compromise teaching efficiency and therefore "deprive", to a certain extent, their students of the time and effort that would have been in place had they been full-time teachers.

Figure 2, presenting the QR results using the Science component, also shows a suddenly sharp rise and dip in the library and principal dummies, respectively, that did not appear in the same panels in Figure 1. For example, in row 3, column 1 of Figure 2, the estimate trend for the library dummy goes down from left to right but suddenly spikes upward at near the 95th quantile. True enough, a significant 6.079 improvement at that quantile level, compared to the relatively low 1.75 improvement estimated by OLS. This implies that for the same covariates, using a different measure of educational output may yield a different conclusion regarding regression coefficient estimates.

Using the Math component instead, some QR estimates for school resources are statistically significant for the first time: the pupil-teacher ratio, pupil-seating ratio, percent of teachers that are locally funded, and class size. Moreover, the library dummy not only displays an apparent downward quantile estimate trend but also goes beyond the OLS confidence interval band at both ends of the entire quantile range. While OLS accords only very small importance to libraries, QR shows the complete picture and tells us that not only are libraries important, but that providing them to low-performing schools first than to schools of higher mean Math performance may be a more efficient resource allocation.

If we use the English component instead, several quantile estimates for the community variables become significant—namely, the malnutrition rate, access to safe water, poverty incidence, and most notably, the literacy rate. The panel for literacy rate in Figure 3 (row 4, column 4) shows for the first time a visible upward trend that crosses beyond the confidence bands. While OLS reports an insignificant 1.03 percentage-point gain from a 1 percent increase in literate persons aged ten years and above in the barangay, the effect could be as low as a 0.38 decline at the 10th quantile and a relatively large increase of 3.54 at the 95th quantile. This suggests that high-performing schools may benefit the most out of increases in community literacy, contrary to the common belief that the poor benefit the most from various literacy programs.

Figure 2. QR coefficient estimates along the different quantiles, using mean Science performance

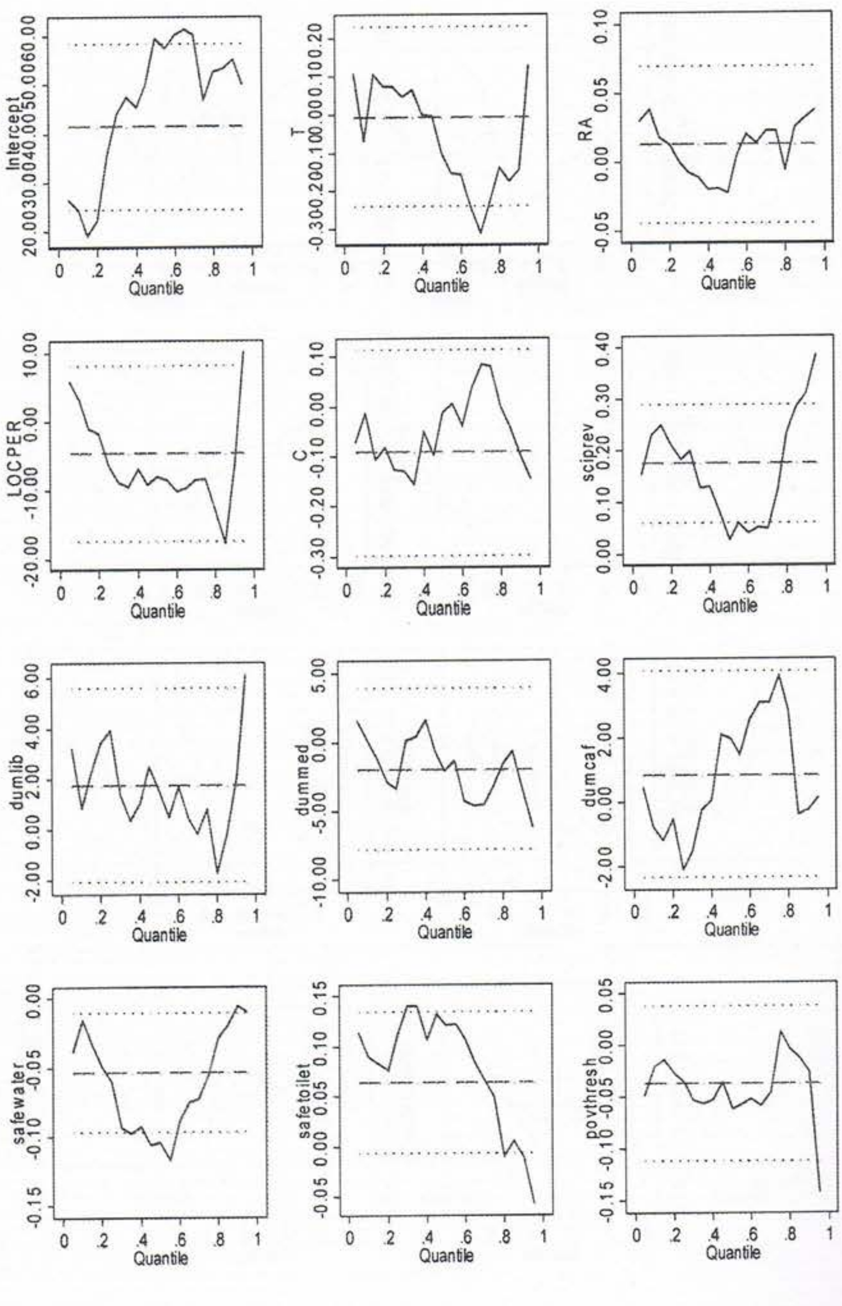


Figure 2. QR coefficient estimates along the different quantiles, using mean Science performance (columns 4 and 5 continued)

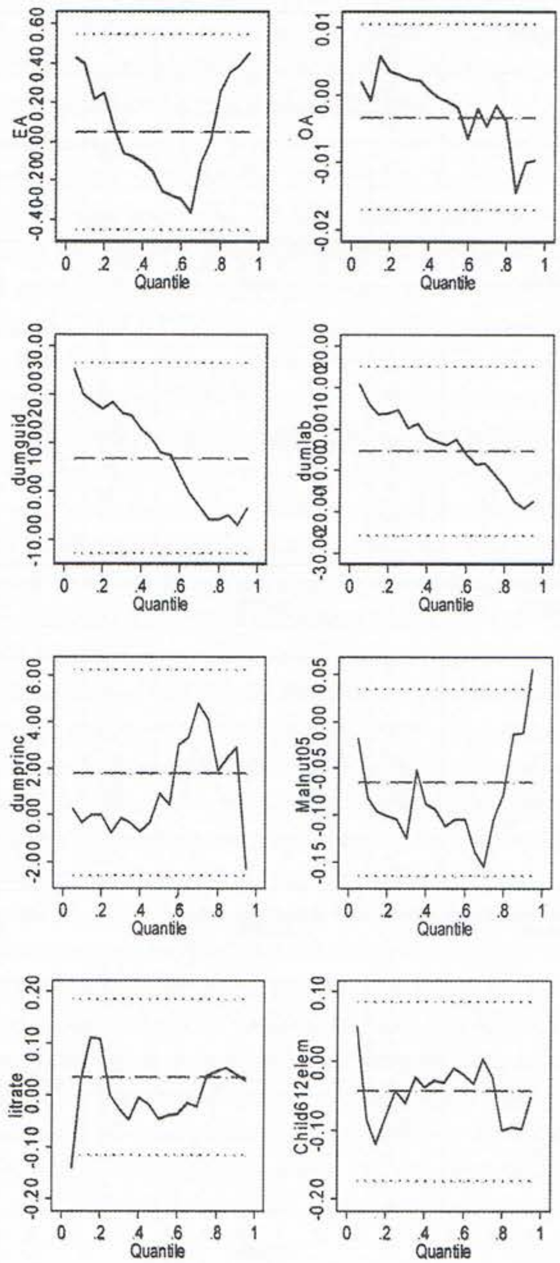


Figure 3. QR coefficient estimates along the different quantiles, using mean English performance

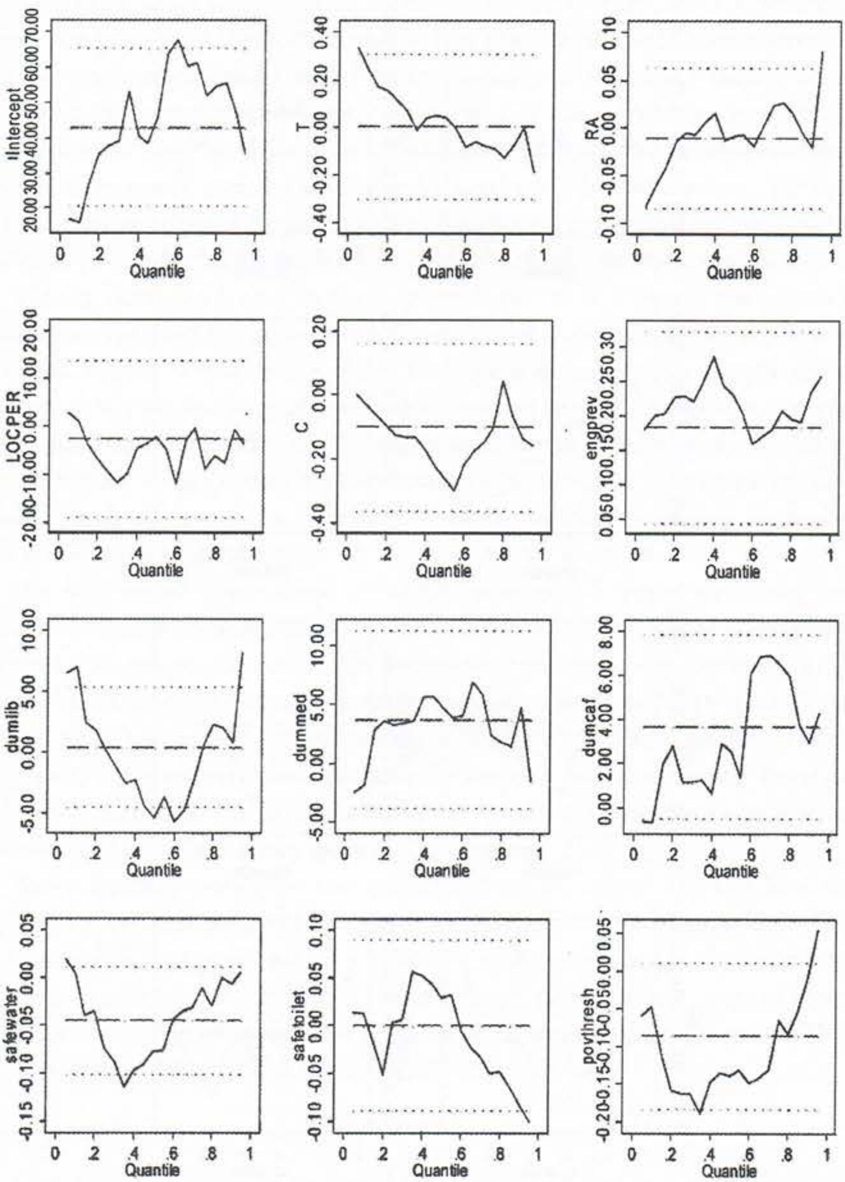
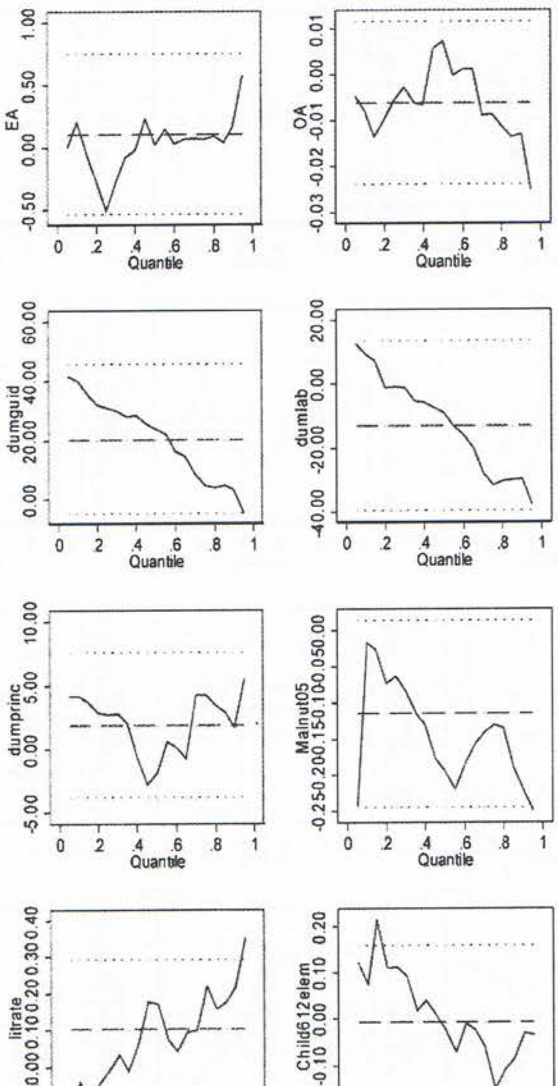


Figure 3. QR coefficient estimates along the different quantiles, using mean English performance (columns 4 and 5 continued)



Hekasi results show that class size, pupil-toilet ratio, and pupil-seating ratios all go beyond the OLS confidence intervals at both ends of the quantile ranges. Adding ten more students per seat, for instance, is associated with a 1.99 percentage-point decline in mean Hekasi performance for schools at the 10th quantile, but is associated with an improvement of as much as 11.1 percentage points for schools at the 95th quantile. It is possible that positive externalities may be responsible for increased performance with increased seating density or a larger class size, in that greater class dynamics may foster an environment conducive for healthy debate and discussion, at least for better-performing schools. Conversely, reduced seating and class density may lead to less interaction and less development of certain communication and argumentation skills among students, possibly contributing to the "disadvantage" of reducing such ratios.

Finally, nutrition seems to play a pivotal role when Filipino test scores are used, as manifested in the cafeteria dummy, safe water, and malnutrition rate variables. Figure 4 shows the panel for access to safe water supply (row 4, column 1) with quantile estimates going beyond the OLS confidence interval, suggesting that OLS may in fact be inaccurate in reporting such quantile trends. Figure 4 also shows the surprisingly discernible QR estimate trend for below-5 malnutrition rates (row 3, column 5), similarly crossing beyond the OLS confidence interval band.

An additional percentage of children ages 0-5 years suffering from malnutrition in a barangay is associated with a 0.39 percentage-point decline in mean Filipino performance. QR estimates, however, post the effect as low as a 1.5 decline in percentage points among schools at the 10th quantile, and as high as a 0.47 improvement among schools at the 95th quantile. Feeding programs, one may say, could possibly be better conducted among barangays with poor-performing schools first than in barangays with better-performing schools, as is usually done in outreach programs.

Most striking would be the cafeteria dummy, which for the first time displays a conspicuous trend that even goes beyond the OLS confidence band. OLS associates the presence of a cafeteria with a statistically significant 4.11 percentage-point improvement in mean Filipino performance, but QR reports that the effect could be as high as 7.47 at the 40th quantile (significant at $\alpha=10\%$) and as low as a decline of 2.36 at the 85th quantile.

5.3. Implications

Our discussion of the OLS and QR regression results has led to several recurring insights.

Figure 4. QR coefficient estimates along the different quantiles, using mean Filipino performance

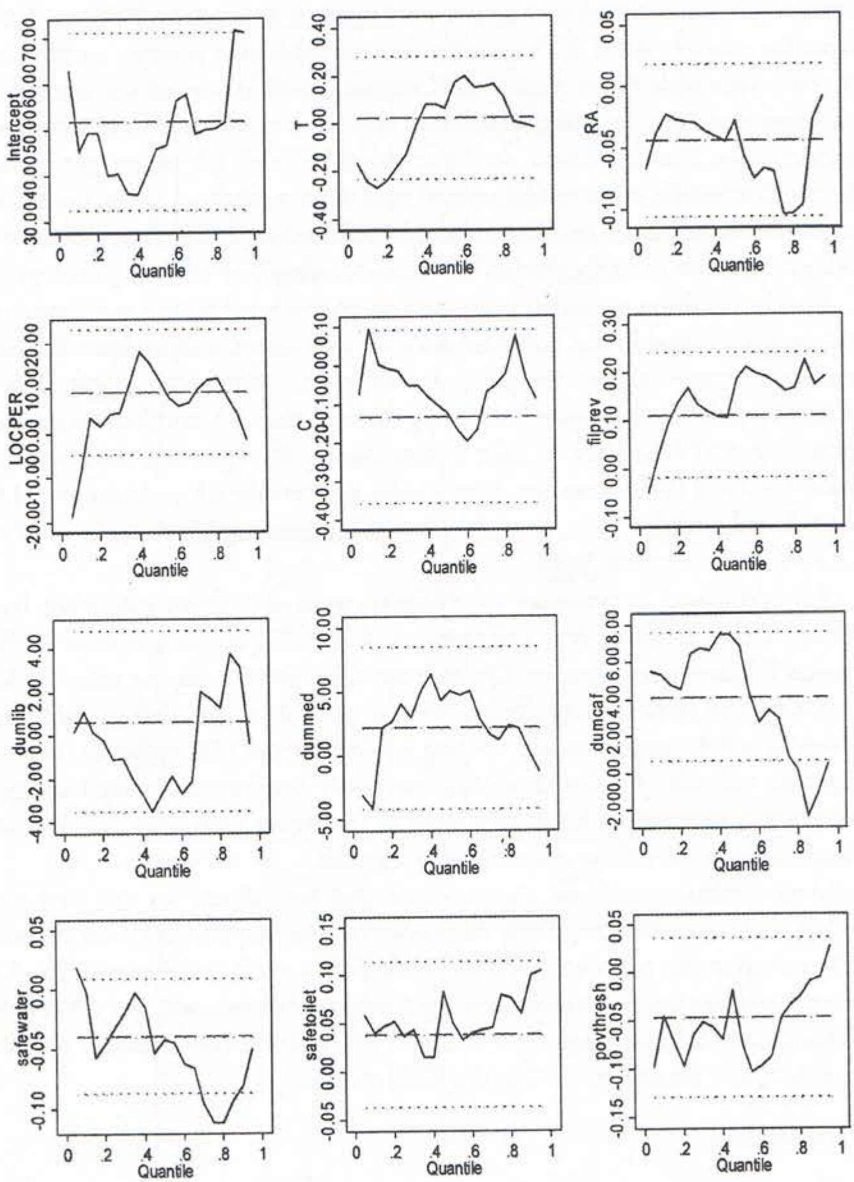
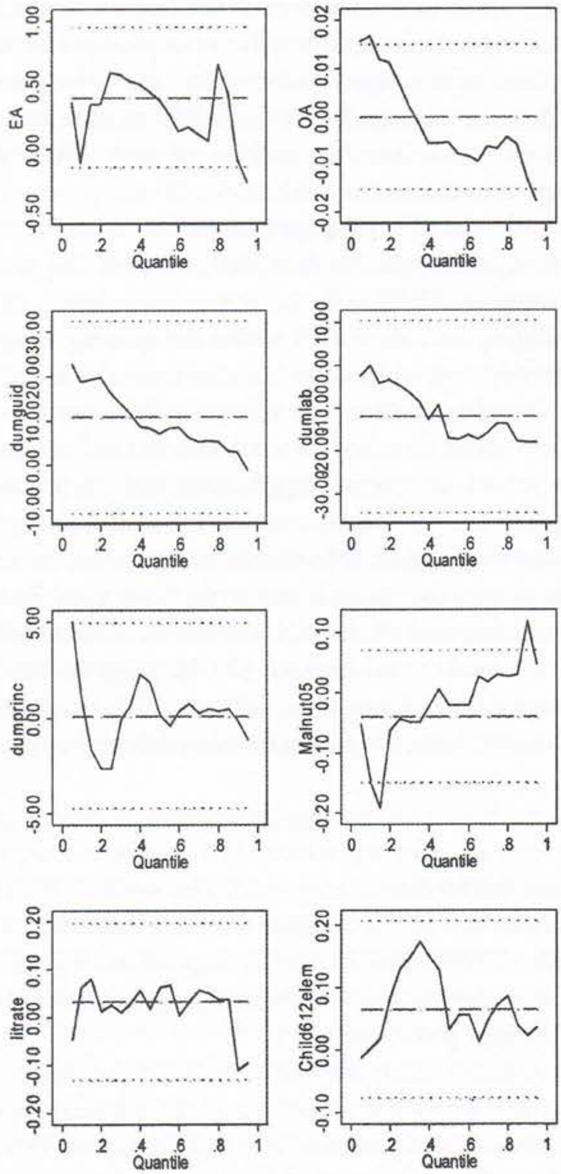


Figure 4. QR coefficient estimates along the different quantiles, using mean Filipino performance (columns 4 and 5 continued)



First, OLS cannot detect and report any variations in covariate effects across the different quantiles of the response variable, for it assumes (rather rigidly) that the parameter effects are constant across the entire quantile range.

Second, OLS may report a parameter to be statistically insignificant when in fact the different quantile parameters actually are significant, according to QR. We have addressed this issue by asserting that even though a facility's parameter is found to be statistically insignificant even in QR, it does not mean that it cannot be of any economic significance—that is, its sign and magnitude are considerable enough to warrant the provision of such a facility despite being deemed statistically insignificant in both OLS and QR.

Besides, that only a handful of quantile estimates are statistically significant is also observed in other education studies, such as the pioneering work of Eide and Showalter [1998]. Of the five school inputs explored by the study, only two display notable effects across the quantile range of the math performance dependent variable; the other covariates display little or no effect on performance, based on statistical significance. Ziliak and McCloskey [2004] caution researchers about so-called "asterisk economics", so as not to neglect parameters that are of economic significance and of practical relevance, notwithstanding conventional practices in statistical significance reporting.

As a third recurring insight, QR coefficient estimates can at times display distinctive trends across the quantile range that can even lie outside the 90 percent confidence interval of the OLS estimate. This insight, a criterion originally used by Koenker and Hallock [2001], suggests that regardless of statistical significance, OLS may at times fail not only in detecting interquantile variations in covariate effects, but also in assuring the reliability of its parameter estimates.

Fourth, either school resources or community variables may alternately prove more relevant in affecting educational outcomes, depending on the measure of student performance employed. The signs of covariate estimates may vary considerably as well, for the same reason. Thus, different performance measures may offer different possible policy suggestions regarding the efficient use of different school resources, and that keeping this in mind during policy-making processes might prove useful.

Lastly, since at times there may be a trade-off between "winners" and "losers" in the provision of some school inputs, untargeted programs intended to improve facilities ratios and provision, although thought to benefit all students in general, may have the unintended consequence of benefiting and causing disadvantage to the wrong target schools. Indeed, such "winner-loser" trade-off observed in most other quantile regression studies (see Bassett, Tam, and Knight [2002]).

We note that Bassett, Tam, and Knight [2002] found that being a student in Chicago has a positive effect for low-performing students and a negative effect in all the higher quantiles. They attribute this to self-selection, for top students in Chicago are more likely to enroll in private schools than top students outside of Chicago, so that top students in public schools in Chicago may be made worse off by not attending private schools. Bassett, Tam, and Knight [2002:24] state: "Whatever the reason, these different impacts go undetected with least squares analysis." Such results illustrate that seemingly "anomalous" findings (of "winners" and "losers" along the different quantiles) are also observed in other education studies using quantile regression, and that their being counterintuitive may not be as unusual as they initially appear.

Educational policymakers must therefore be cognizant of possible consequences arising from unintended effects of programs among schools executed in an untargeted manner. For if they rely on OLS estimates alone, they might conclude that if the parameter is shown to be significant, it should be provided equally to all student types, never realizing that it benefits *and* harms different school types in the process.

This shows yet another advantage of using QR over OLS when it comes to educational production: since QR can pinpoint exactly which quantiles are possibly affected by a certain educational input, it can offer policy guidelines that can pinpoint and direct the flow of educational resources only to winners (or those who will benefit) rather than unintentionally also provide them to students who will be "disadvantaged" by such a provision to them. Therefore, it may be less necessary to worry about the "winner-loser trade-off," given the specified and direct nature of empirical and policy suggestions QR has to offer.

6. Conclusion and recommendations

This paper investigates the role of school resources in educational outcomes in the province of Palawan. In choosing how best to allocate our educational resources, we contend that it is not enough for the government to identify which school resources matter the most, and expend all resources on them. Rather, it would be more efficient and useful for policy purposes to further identify which among the schools will benefit the most from a given school resource.

This study uses the emergent technique of quantile regression to address such problem. The results for the province of Palawan show that, in general, high-performing schools may benefit the most from improved pupil-teacher and pupil-toilet ratios, more than lower-performing schools, regardless of the performance measure employed. Conversely, class size and pupil-room ratio improvements, along with the provision of libraries and guidance counselors, may improve the mean performance of lower-performing schools the most. Meanwhile, we find that the other covariates used in this study may or may not improve schools at different quantile levels, depending on the measure of mean performance used. Table 2 presents which school inputs matter for which types of schools.

We also extract some insights regarding the use of quantile regression for education production studies. First, we find that by producing only a single parameter estimate across all quantiles, OLS may be both insufficient and imprecise in uncovering the underlying covariate relationships that may exist between school resources and performance measures across the different quantiles.

Second, we find that even though OLS may estimate that a school resource's parameter is statistically insignificant, it is possible for quantile regression to show that it is significant at different quantile estimates. Also, by using the different components of NAT as alternative measures of student performance, we see that such different measures indeed yield varying quantile estimate patterns, and that the resulting policy recommendations may also differ as a result.

Finally, we find that for any untargeted educational investment policy, there may possibly be "winners" and "losers" among the entire range of schools. But with the help of quantile regression, we can specify exactly who the "winners" and "losers" are for each school resource, allowing us to efficiently direct resources only to those who stand to gain from their provision.

Future research should aim to incorporate a wider array of explanatory variables, include an expanded time dimension into the analysis (perhaps using panel data), or attempt to engage in randomization, wherever possible. Finally, with the benefit of more complete data from the DepEd and the CBMS (whose promising datasets are still being completed), it would be fruitful to conduct a nationwide quantile regression analysis regarding education production, toward a unified and comprehensive direction for educational investment policy making. Even more ambitious would be a nationwide quantile regression study at the level of the individual student.

All of these future studies should help in building the literature of education production for the Philippines, where rigorous empirical research may assist education policymakers in improving the quality of educational planning and resource allocation in a developing country like ours. In turn, we expect such investments to help improve the educational performance of our students, in the still greater hope of our country rising from the current state of its public educational system.

Table 3. Trends of school resource effects in the province of Palawan

School resources that may benefit high-performing students the most	*Ambiguous effect**	*School resources that may benefit low-performing students the most*
Lower pupil-teacher ratio	Lower pupil-seating ratio	Lower pupil-room ratio
	Principal	Lower class size
Lower pupil-toilet ratio	Cafeteria	Guidance counsellor
	Library	Science laboratory
	Medical clinic	
	Lower percent locally-funded teachers	

*Effect depends on measure of performance used.

References

Acedo, C. [2000] *Teacher supply and demand in the Philippines*. Policy Notes Three. Manila: Asian Development Bank.

Bacolod, M. and J. Tobias [2006] "Schools, school quality and academic achievement: Evidence from the Philippines", *Economics of Education Review* **25**(6): 619-632.

Bassett, G. Jr., M.-Y. Tam, and K. Knight [2002] "Quantile models and estimators for data analysis", *Metrika* **55**(1-2): 17-26.

Buchinsky, M. [1998] "Recent advances in quantile regression models: a practical guideline for empirical research", *Journal of Human Resources* **33**(1): 88-126.

Eide, E. and M. Showalter [1998] "The effect of school quality on student performance: A quantile regression approach", *Economics Letters* **58**(3): 345-350.

Hanushek, E. A. [1998] "The evidence on class size", Occasional Paper No. 98-1. New York: W. Allen Wallis Institute of Political Economy, University of Rochester.

Hedges, L. V., R. Laine, and R. Greenwald [1994] "Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes", *Educational Researcher* **23**(3): 5-14.

Howell, D. C. [2002] Treatment of missing data. Mimeo. University of Vermont. http://www.uvm.edu/~dhowell/StatPages/More_Stuff/Missing_Data/MissingDataFinal.pdf. Accessed 2 October 2008.

Koenker, R. and G. Bassett [1978] "Regression quantiles", *Econometrica* **46**(1): 33-50.

Koenker, R. and K. Hallock [2001] "Quantile regression", *Journal of Economic Perspectives* **15**(4): 143-156.

Krueger, A. B. [2003] "Economic considerations and class size", *Economic Journal* **113**(485): 34-63.

Levin, J. [2001] "For whom the reductions count: a quantile regression analysis of class size and peer effects on scholastic achievement", *Empirical Economics* **26**(1): 221-246.

Orbeta, A. Jr. [2008] "Achievement test scores and school characteristics: evidence from Philippine public elementary and secondary schools." PowerPoint presentation. 6th National Social Science Congress. Philippine Social Science Center, Quezon City. http://www.pssc.org.ph/Downloads/nssc/AchievementTest.pdf. Accessed 27 October 2008.

Quimbo, S. [2003] "Explaining math and science achievement of public school children in the Philippines", *Philippine Review of Economics* **40**(2): 59-72.

Rangvid, B. S. [2003] "Educational peer effects: quantile regression evidence from Denmark with PISA 2000 data", Ph.D. thesis, Aarhus School of Business, Institute for Local Government Studies: Copenhagen, Denmark.

Todd, P. and K. Wolpin [2003] "On the specification and estimation of the production function for cognitive achievement", *Economic Journal* **113**(485): F3-F33.

Ziliak, S. and D. McCloskey [2004] "Size matters: the standard error of regressions in the American Economic Review", *Econ Journal Watch* **1**(2): 331-358.