

The lagged-value-added model incorporates the assumption that current knowledge is an additive function of all current and past inputs and an i.i.d. stochastic error term. This can be written formally as:

$$Y_{i,t} = \alpha_t X_{i,t} + \alpha_{t-1} X_{i,t-1} + \cdots + \alpha_1 X_{i,1} + \epsilon_{i,t} \quad (1)$$

where $Y_{i,t}$ is child i 's test scores at time t and $X_{i,t}$ is a vector of child, school, and village controls at time t . As data on all past inputs are usually unavailable, however, they are generally subsumed into a lagged dependent variable included as a control. In this case the lagged-value-added model can be re-written as:

$$Y_{i,t} = X_{i,t} \alpha + Y_{i,t-1} \beta + \epsilon_{i,t} \quad (2)$$

Where $Y_{i,t-1}$ is assumed to capture all past inputs and unobservable heterogeneity across students. The specification given by Equation 2 is the model primarily employed in this paper.

The primary interest of this analysis is on the differential “value-added” by government and private schools. This parameter is the coefficient on a dummy variable in $X_{i,t}$ for whether the child attends a private school, where positive values suggest a higher “value-added” (a positive test-score gap) between private and government schools.

Value-added estimates for teachers – used in in Section ?? – are generated by adding dummies for each teacher to Equation 2. One dummy variable is added for each teacher, and the coefficient associated with each teacher’s dummy is their “value-added.”

Three aspects of the lagged-value-added specification are worth emphasizing. First, while the inclusion of a lagged dependent variable effectively controls for unobserved differences that affect differences in test *levels*, it cannot control for unobserved heterogeneity that affects learning *rates*. It is for this reason that while superior to other available methods, value-added analyses can not fully overcome selection issues.¹

Second, the β term can be interpreted as the “persistence parameter” in that it estimates the degree to which past learning may carry forward. A value of one is equivalent to assuming that children do not forget past lessons, while a value of zero corresponds to students forgetting all past lessons each year. While imposing a persistence parameter of one may seem reasonable – it amounts to regressing the difference in test scores between time t and time $t - 1$ on controls – a growing literature has shown that the test score gains of short term interventions often “die out” over time, suggesting that not all that is learned is retained (????). ? shows that imposing the restriction that $\beta = 1$ biases learning estimates.

Finally, lagged test scores are generally measured with error, which leads to an often significant attenuation bias in the estimate of β and biased estimates of other coefficients (???). Thus keeping with best practices, lagged test scores from all three subjects – English, Urdu, and math – are included in all specifications to instruments for the primary lagged test score of interest.

¹Some analysis have turned to second-differencing the data and focusing on students who change schools (?), but these analyses have their own limitations, among them limited sample sizes (given that changes between types of school are relatively infrequent in most surveys) and the assumption that school changes are not the result of some unobserved shock (i.e. that school switches are not accompanied by contemporaneous with other changes – a potentially problematic assumption given the relative infrequency with which students change schools).

Note further that all child-level results are presented using heteroskedastic-robust standard errors clustered at the level of the village.