Having established that the determinants of school choice vary dramatically between caste-homogeneous and caste-heterogeneous villages in Section **??**, this section now turns to an analysis of how differences in sorting impact the test-score gap between government and private schools. If private schools outperform government schools primarily due to differences in the quality of instruction, then the government-private test-score gap should be relatively stable across villages with different caste compositions. If, however, private schools appear to outperform government schools primarily due to differences in the composition of their students, then villages that are subject to different sorting processes should also see differences in the government-private test-score gap. Specifically, the degree to which private schools scores exceed those in government schools should decline as one moves from caste-homogeneous villages (where students sort on academic potential) to caste-heterogeneous villages (where sorting is driven by considerations of caste politics).

0.1 Measuring Learning

To measure learning, this analysis employs a lagged-value-added model. Lagged-value-added models have increasingly become the norm in the education research (**???**) due to their potential to take into account not only observational differences between students, but also the potential to control for some unobserved differences, and the fact that learning is not entirely persistent (things learned in the past are often forgotten).

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, and can be expressed as:

$$Y_{i,t} = X_{i,t}\alpha + Y_{i,t-1}\beta + \epsilon_{i,t} \tag{1}$$

where $Y_{i,t}$ is child i's test scores at time t and $X_{t,i}$ is a vector of child, school, and village controls at time t (a full discussion of the lagged-value-added model, specifications employed, and identifying assumptions can be found in Appendix **??**).

Note that while the inclusion of a lagged dependent variable effectively controls for unobserved heterogeneity that affect differences in test score *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] In the lagged-value-added model, coefficients on independent variables are interpreted as the contribution of each variable to learning.

0.2 Convergence in Government-Private Test Scores

Table 1 presents lagged-valued-added estimates of learning as a function of various demographic controls and village fractionalization. It shows that the effect of caste fractionalization on the government-private test-score gap (row 2) is negative and significant for English and negative (albeit insignificant) for math and Urdu. Further, as shown in columns (2), (5), and (8) of Table 1, the inclusion of various demographic controls such as a child wealth index and dummies for parental education along with the village fixed effects has no significant effect on the results.

---

[1] Some 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

Table 1: Child Test Scores

|	English			Urdu			Math		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Private School	0.62***	0.58***	0.57***	0.27***	0.26***	0.21***	0.21*	0.20*	0.16
	(0.072)	(0.078)	(0.078)	(0.073)	(0.070)	(0.073)	(0.12)	(0.11)	(0.11)
Fractionalization * Private	-0.46***	-0.44***	-0.44***	-0.18	-0.18	-0.11	-0.12	-0.15	-0.11
	(0.12)	(0.13)	(0.12)	(0.12)	(0.11)	(0.11)	(0.18)	(0.16)	(0.15)
Lagged English Scores	0.37***	0.36***	0.39***	0.15***	0.14***	0.14***	0.16***	0.15***	0.16***
	(0.018)	(0.018)	(0.018)	(0.012)	(0.012)	(0.012)	(0.015)	(0.015)	(0.016)
Lagged Math Scores	0.069***	0.072***	0.071***	0.12***	0.12***	0.12***	0.37***	0.38***	0.40***
	(0.0080)	(0.0080)	(0.0083)	(0.0083)	(0.0086)	(0.0086)	(0.013)	(0.014)	(0.014)
Lagged Urdu Scores	0.15***	0.15***	0.15***	0.38***	0.38***	0.40***	0.23***	0.22***	0.22***
	(0.011)	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)
Child's Wealth Index		0.017***	0.015***		0.0073**	0.0068**		0.014***	0.016***
		(0.0033)	(0.0035)		(0.0030)	(0.0030)		(0.0040)	(0.0042)
Educated Parent		0.057***	0.052***		0.052***	0.049***		0.045***	0.043***
		(0.012)	(0.013)		(0.011)	(0.011)		(0.014)	(0.014)
Biraderi Fractionalization			0.23***			0.090			0.15
			(0.084)			(0.068)			(0.11)
Village: Pct Adults Literate			0.00014			-0.00061			0.00024
			(0.00097)			(0.00086)			(0.0013)
Log Number of Households			0.020			0.015			0.011
			(0.013)			(0.014)			(0.018)
Village Land Gini			0.049			0.060			-0.25*
			(0.11)			(0.100)			(0.13)
Constant	0.25	0.40*	0.57**	0.55**	0.69***	0.79***	0.12	0.31	0.61*
	(0.26)	(0.22)	(0.25)	(0.22)	(0.25)	(0.28)	(0.34)	(0.32)	(0.35)
Village Fixed Effects	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
District Fixed Effects	No	No	Yes	No	No	Yes	No	No	Yes
Observations	37147	26141	26141	37147	26141	26141	37147	26141	26141

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Controls for age, age squared, gender, and class omitted from table. Standard errors clustered at village level.

To aid in interpretation, Figure 1 plots the government-private performance differential as a function of caste fractionalization (these plots correspond to columns (2), (5), and (8) respectively). In all three cases, the rise in fractionalization is associated with a near 50% decline in the private school premium, although this is by far most striking in the case of English.

The differential impact of village fragmentation on English and other subjects is not surprising. English is generally considered the path to upward mobility in Pakistan, and is often the focus of private schools in Punjab. For example, while only 6% of government schools use English as one of their languages of instruction, this is the case in 28% of private schools (**?**, p. 49). Indeed, a consistent pattern in the data is that – possibly as a result of this specialization – the English test-score gap is consistently the largest among the three subjects tested.

Figure 1: Private School Test Score Premium with Lagged Scores
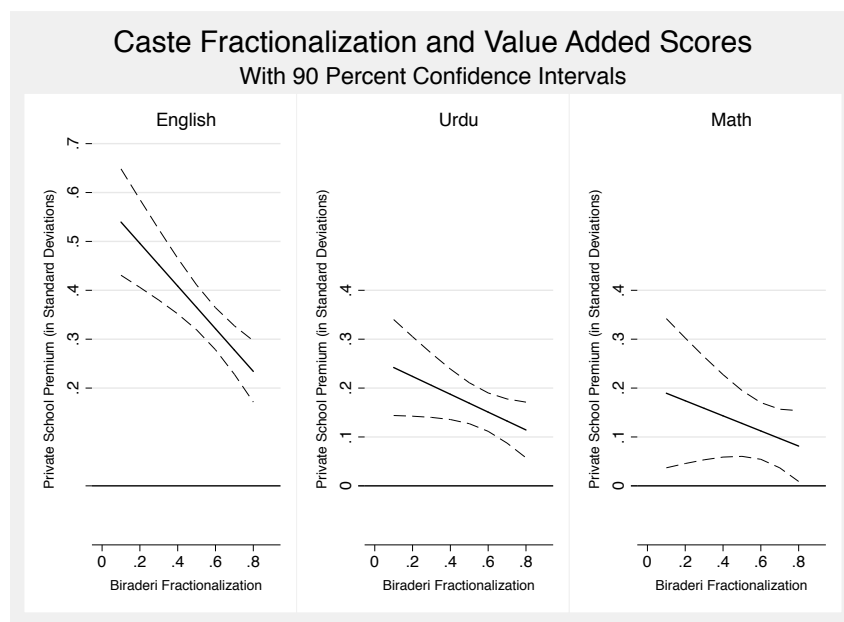
0.3 Decomposition of Convergence

Further evidence that the convergence in government-private test scores is driven differences in student sorting – not differences in actual learning outcomes – comes from the fact that while the government-private test-gap decreases, overall learning remains relatively unchanged across villages. As shown in Table 2 – which examines the relationship between overall test scores and village heterogeneity – test scores do not vary with caste composition. English scores are slightly higher in more fractionalized villages in Column 1, but the magnitude of this difference is relatively small, and once more demographic controls are added in Column 2 this effect disappears. No relationship exists for other subjects. Government scores increase and private scores decline with fractionalization, in other words, but those changes are almost perfectly offsetting.

This is also illustrated in Column 3 of Table 1, where village fixed effects are replaced with district fixed effects, allowing for the inclusion of a caste-heterogeneity variable. Even when controlling

---

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).

for the relatively small improvement in overall English scores, the estimated effect of village caste composition on the government-private test-score gap remains stable, and the level change remains much smaller than the magnitude of test-score convergence. This provides additional evidence that as the determinants of school choice change, it is the *distribution* of academically inclined students that changes, not the performance of the schools themselves.

0.4   Caste and Residual Academic Potential

For it to be the case that sorting by caste reduces the degree to which private schools enroll dispro-portionately academically-inclined students, residual academic potential – potential that cannot be explained by things like parental education and wealth – must be equally distributed across different castes (or be distributed slightly in favor of lower status *biraderis*). If not, and even the least talented "high status" students were more talented than the most talented "low status" students, then the concentration of "high status" students in private schools would result in *divergence*, rather than *convergence*, of test scores. As shown in Table 3, however, there is no evidence that those from higher social status *biraderis* have higher residual talent than those from low status *biraderis*. As evident from the top row of coefficients, after controlling for other observational factors, student caste does not appear to have any consistent effect on test scores.

Table 2: Child Test Scores and Fractionalization

	English		Urdu		Math	
	(1)	(2)	(3)	(4)	(5)	(6)
Private School	0.31***	0.29***	0.14***	0.14***	0.11***	0.088**
	(0.028)	(0.028)	(0.025)	(0.024)	(0.035)	(0.034)
Biraderi Fractionalization	0.14*	0.10	0.080	0.057	0.13	0.12
	(0.075)	(0.073)	(0.066)	(0.064)	(0.094)	(0.089)
Lagged English Scores	0.40***	0.39***	0.16***	0.14***	0.17***	0.16***
	(0.018)	(0.018)	(0.012)	(0.012)	(0.016)	(0.016)
Lagged Math Scores	0.067***	0.070***	0.12***	0.12***	0.39***	0.40***
	(0.0082)	(0.0083)	(0.0083)	(0.0086)	(0.013)	(0.014)
Lagged Urdu Scores	0.15***	0.15***	0.39***	0.40***	0.23***	0.22***
	(0.011)	(0.012)	(0.011)	(0.012)	(0.013)	(0.013)
Village: Pct Adults Literate	0.00058	0.00017	-0.00027	-0.00060	0.00031	0.00025
	(0.00097)	(0.00098)	(0.00086)	(0.00085)	(0.0014)	(0.0013)
Log Number of Households	0.017	0.019	0.012	0.015	0.0083	0.011
	(0.013)	(0.013)	(0.015)	(0.014)	(0.019)	(0.017)
Village Land Gini	-0.0043	0.036	0.0079	0.057	-0.29**	-0.26*
	(0.13)	(0.12)	(0.11)	(0.098)	(0.13)	(0.13)
Child's Wealth Index		0.015***		0.0068**		0.016***
		(0.0035)		(0.0030)		(0.0042)
Educated Parent		0.053***		0.049***		0.043***
		(0.013)		(0.011)		(0.014)
Constant	0.45	0.64**	0.68***	0.81***	0.34	0.62*
	(0.28)	(0.25)	(0.25)	(0.28)	(0.37)	(0.35)
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	37147	26141	37147	26141	37147	26141

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Controls for age, age squared, gender, and class omitted from table. Standard errors clustered at village level.

Table 3: Child Social Status and Residual Talent

	English			Urdu			Math		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
High Status Zaat	-0.040	-0.072	-0.11**	-0.037	-0.10	-0.081	0.0014	-0.025	0.037
	(0.045)	(0.070)	(0.046)	(0.058)	(0.071)	(0.052)	(0.066)	(0.085)	(0.067)
Private School	0.32***	0.22*	0.21*	0.28**	0.26*	0.14	-0.037	-0.060	-0.11
	(0.12)	(0.13)	(0.12)	(0.12)	(0.14)	(0.13)	(0.11)	(0.14)	(0.15)
Fractionalization * Private	-0.040	0.081	0.048	-0.21	-0.19	-0.067	0.28	0.33	0.29
	(0.20)	(0.21)	(0.20)	(0.19)	(0.23)	(0.20)	(0.19)	(0.24)	(0.24)
Lagged English Scores	0.32***	0.31***	0.41***	0.16***	0.16***	0.16***	0.18***	0.19***	0.18***
	(0.042)	(0.046)	(0.040)	(0.031)	(0.036)	(0.029)	(0.045)	(0.046)	(0.040)
Lagged Math Scores	0.060**	0.048	0.045	0.11***	0.076*	0.089**	0.30***	0.27***	0.37***
	(0.030)	(0.035)	(0.032)	(0.038)	(0.042)	(0.040)	(0.050)	(0.050)	(0.048)
Lagged Urdu Scores	0.17***	0.19***	0.17***	0.34***	0.35***	0.41***	0.25***	0.26***	0.26***
	(0.038)	(0.046)	(0.042)	(0.039)	(0.045)	(0.046)	(0.046)	(0.053)	(0.052)
Child's Wealth Index		-0.0067	-0.011		0.0020	0.0010		-0.014	-0.0094
		(0.013)	(0.012)		(0.010)	(0.0090)		(0.017)	(0.014)
Educated Parent		0.14***	0.14***		0.14***	0.14***		0.17***	0.18***
		(0.043)	(0.039)		(0.044)	(0.035)		(0.055)	(0.046)
Biraderi Fractionalization			-0.029			-0.063			-0.042
			(0.10)			(0.13)			(0.17)
Village: Pct Adults Literate			0.00037			-0.0032**			-0.0013
			(0.0018)			(0.0016)			(0.0027)
Log Number of Households			0.021			0.025			0.050
			(0.036)			(0.033)			(0.061)
Village Land Gini			0.027			0.11			-0.15
			(0.18)			(0.21)			(0.32)
Constant	0.57	0.71	0.75	1.18*	2.35***	2.22***	1.57**	2.83***	2.19**
	(0.56)	(0.71)	(0.60)	(0.64)	(0.74)	(0.67)	(0.74)	(0.92)	(0.94)
Village Fixed Effects	Yes	Yes	No	Yes	Yes	No	Yes	Yes	No
District Fixed Effects	No	No	Yes	No	No	Yes	No	No	Yes
Observations	1790	1323	1323	1790	1323	1323	1790	1323	1323

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01