PSYCHOMETRIKA—VOL. 89, NO. 3, 1074–1103
SEPTEMBER 2024
https://doi.org/10.1007/s11336-024-09979-0

TEMPORALLY DYNAMIC, COHORT-VARYING VALUE-ADDED MODELS

GARRITT L. PAGE

BRIGHAM YOUNG UNIVERSITY

ERNESTO SAN MARTÍN

MILLENNIUM NUCLEUS ON INTERGENERATIONAL MOBILITY MOVI

LIDAM/CORE, UNIVERSITÉ CATHOLIQUE DE LOUVAIN

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

DAVID TORRES IRRIBARRA

SCHOOL OF PSYCHOLOGY, PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

SÉBASTIEN VAN BELLEGEM

LIDAM/CORE, UNIVERSITÉ CATHOLIQUE DE LOUVAIN

We aim to estimate school value-added dynamically in time. Our principal motivation for doing so is to establish school effectiveness persistence while taking into account the temporal dependence that typically exists in school performance from one year to the next. We propose two methods of incorporating temporal dependence in value-added models. In the first we model the random school effects that are commonly present in value-added models with an auto-regressive process. In the second approach, we incorporate dependence in value-added estimators by modeling the performance of one cohort based on the previous cohort's performance. An identification analysis allows us to make explicit the meaning of the corresponding value-added indicators: based on these meanings, we show that each model is useful for monitoring specific aspects of school persistence. Furthermore, we carefully detail how value-added can be estimated over time. We show through simulations that ignoring temporal dependence when it exists results in diminished efficiency in value-added estimation while incorporating it results in improved estimation (even when temporal dependence is weak). Finally, we illustrate the methodology by considering two cohorts from Chile's national standardized test in mathematics.

Key words: School value persistence, Value-added models, Temporal dependence.

1. Introduction

Value-added models are frequently utilized to evaluate the contributions of educational institutions or stakeholders to the educational process. In certain instances, these models have directly influenced education policies (Sass et al., 2012; Koedel et al., 2015; Kyriakides et al., 2018; Liu & Loeb, 2019; Hanushek, 2020). While there are several criticisms regarding their use (EPI Briefing Paper, 2010; Scherrer, 2011; Ehlert et al., 2014; Amrein-Beardsley & Holloway, 2019) , these critiques are primarily directed at the contexts in which they are applied, rather than their intrinsic

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11336-024-09979-0.

The manuscript was handled by the ARCS Editor Dr. Nidhi Kohl.

Correspondence should be made to Ernesto San Martín, Faculty of Mathematics, Pontificia Universidad Católica de Chile, Vicuna Mackenna 4860, Macul, Santiago, Chile. Email: esanmar@uc.cl

© 2024 The Author(s), under exclusive licence to The Psychometric Society

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

value in advancing educational research (McCaffey et al., 2004; Reynolds et al., 2014) . Nonetheless, value-added models appear to be valuable for monitoring the effectiveness of schools when different measures at both the school and student levels are taken over time.

Broadly speaking, two overarching perspectives regarding value-added model building exist in the school effectiveness literature. The first one considers an invariant group of people subjected to multiple measures over a time period. In this perspective, the school effect is constant over the time, capturing thus the effect of the school after considering the full process; this type of approach is developed through the so-called growth models (Potthoff & Roy, 1964; Strenio et al., 1983; Fitzmaurice et al., 2004; Guldemond & Bosker, 2009; Bianconcini & Cagnone, 2012) . In the second perspective, the composition of the group of individuals changes over time: each group is measured twice (pre- and post-test), allowing the identification of the school effect for each period; in this perspective, student achievement (which is often a standardized test score) is regressed onto previous attainment scores (i.e., standardized test result at the beginning of the value-added period).

This paper focuses on the second perspective because we aim to investigate how the performance of a previous cohort influences school effectiveness as the school accepts a new cohort. For instance, in our case study, the first cohort consists of students who were in the 4th grade in 2012, took the pre-test, and then took the post-test in 2014 as 6th graders. Subsequently, the second cohort, comprising students who took the pre-test in the 4th grade during the 2014 school year, took the post-test in 2016 as 6th graders. For similarly structured data, refer to Fig. 1 in Papay (2011, Figure 1).

When estimating a particular institution's value-added across multiple cohorts of students, there is much interest to determine the extent to which a school's effectiveness persists over time. One approach of modeling such a persistence is by considering school and/or teacher effects as "the effects cumulate over time" (Briggs & Weeks, 2011, p. 620). The underlying idea is that effectiveness varies over time, and school and/or teacher effects represent school or teacher impacts within each academic year (Sanders & Horn, 1994; Ballou et al., 2004) . At the model specification level, the test score of a student at time t is determined by covariates, particularly the test score at time $t - 1$, and a linear combination of school and/or teacher effects from t to the initial data collection time (McCaffey et al., 2004; Lockwood et al., 2007; Rothstein, 2010; Kinsler, 2012) . It's worth noting that this approach typically necessitates cohort scores available for at least two time periods (Vanwynsberghe et al., 2017; Tymms et al., 2018) .

An alternative approach to analyzing the persistence of school effectiveness involves assessing school value-added indicators over time. It's reasonable to assume that school effectiveness would generally exhibit stability, barring abrupt changes in faculty, resources, and leadership. From a statistical perspective, this suggests that value-added estimators would demonstrate temporal dependence. Consequently, proposing a model independently for each cohort of students might, at first glance, overlook the correlation among these estimators and result in a loss of efficiency when estimating the persistency of school effectiveness. Unfortunately, this approach is currently common practice.

Efforts to address this temporal dependence in value-added models often involve two-step procedures based on correlating value-added estimates post-model fit with each cohort being modeled separately (Gray et al., 2001; Thomas et al., 2007; Bellei et al., 2016) . However, this approach is suboptimal, as highlighted by Leckie (2018), who recently explored this method and demonstrated persistent biases. As mentioned by Leckie (2018), a much preferable approach would be to incorporate temporal dependence coherently in a statistical model by jointly modeling cohorts. This can be achieved in various ways, with Leckie (2018) proposing one such approach.

Our goal here is therefore to further develop time-dependent value-added models. We specify a time-dependent value-added model to answer the following question: is it possible that the current effectiveness of a school as it takes in a new cohort is jointly influenced by both its previous

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

effectiveness and some previous information from the former cohort? To answer this question, an identification analysis is necessary. The analysis we carry out shows that the parameters characterizing both the dependence of the previous school effectiveness and the previous cohort are identified. By using the model-free definition of value-added (Manzi et al., 2014), we derive the corresponding school value-added indicators. Afterward, we establish an interpretation of school effectiveness persistence for the case of two cohorts, which consists of decomposing the school value-added for cohort 2 into two components: the first one corresponds to the expectation of the school value-added for cohort 2 conditionally on the school value-added for cohort 1, whereas the second component corresponds to the school value-added for cohort 2 minus the latter conditional expectation. This type of decomposition, typically used in other fields of psychometrics (e.g., Classical Test Theory Zimmerman, 1975), Factor Analysis (Lord & Novick, 1968, Chapter 24) and School Effectiveness Manzi et al. (2014)), has the advantage that the first component corresponds to the explanation of cohort 2's school value-added by the school value-added for cohort 1, while the second component corresponds to everything in value-added for cohort 2 that cannot be explained by value-added for cohort 1. Thus, the persistence of school efficiency corresponds to an additive combination of both the school value-added for cohort 1 and the information coming from cohort 1. Moreover, we prove that the first additive component is related to a model that is nested in our general formulation, which we denote by Model 1: it is characterized by assuming that the school effects are correlated over time as in ARIMA-type models. Similarly, we prove that the second additive component is related to a separate model nested in our general formulation, which we denote by Model 2: it is characterized by assuming that the current school effect is influenced by the post-tests from previous cohorts as a kind of "information shock".

The rest of this paper is organized as follows. The time-dependent value-added models and the structural interpretation of the value-added indicators are derived and discussed in Sect. 2. Computation and model fitting, both based on Bayesian procedures, are briefly explained in Sect. 3. A simulation study to explore the impact of ignoring temporal dependence on value-added estimates is conducted in Sect. 4. Section 5 details a case study using data of the Chilean educational system. Conclusions and future work are gathered in Sect. 6.

2. Time-Dependent Value-Added Model

In this section we describe our approach to incorporate temporal dependence in value-added models. To begin we introduce notation that will be used throughout the article. Let Y_{tij} denote the jth measurement coming from the ith school for cohort t where $j = 1, \ldots, n_{ti}, i = 1, \ldots, I$, and $t = 1, \ldots T$ (in our application $T = 2$). Further, let $\boldsymbol{Y}_{ti} = (Y_{ti1}, \ldots, Y_{tin_{ti}})'$ be a $n_{ti} \times 1$ vector of response values for cohort t of the ith school. Let $\boldsymbol{X}_{tij}$ be a $p_t \times 1$ vector of covariates measured from the jth student at the ith school for cohort t and $\boldsymbol{X}_{ti} = (\boldsymbol{X}_{ti1}, \ldots, \boldsymbol{X}_{tin_{ti}})'$ denote the $n_{ti} \times p_t$ "stacked matrix" of all covariate vectors measured from the ith school for cohort t. Note that this matrix does not include a column vector of ones. When it is necessary to refer to the pth covariate, we will use $X_{tij,p}$. We will use α_{ti} to denote the ith latent school effect for cohort t respectively. Finally, $\boldsymbol{\beta}_t$ will denote a $p_t \times 1$ vector of regression coefficients for cohort t; the remaining parameters will be made explicit in the sequential specification below; for the time being, the set of all parameters for cohort t (including $\boldsymbol{\beta}_t$) will be denoted by $\boldsymbol{\psi}_t$.

2.1. Sequential Model Specification

When more than one cohort is available, the school effect α_{ti} can in principle be influenced by two types of temporal factors: one that is unobserved and corresponds to the school effect of the previous cohort, namely α_{t-1i}; and one that is observed corresponding to (a function of the)

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

post-tests from the previous cohort, namely $\boldsymbol{Y}_{t-1i}$. Thus, temporal dependence in value-added models is not only based on dependence between a school's current cohort performance and its previous one (which is captured by α_{t-1i}), but also includes the impact that the information shock contained in $\boldsymbol{Y}_{t-1i}$ has on current school performance.

As a result, a temporal dependent value-added model should be sequentially specified. More specifically, denoting $\boldsymbol{Y}_{1,i}^{t} = \{\boldsymbol{Y}_{1i}, \boldsymbol{Y}_{2i}, \ldots, \boldsymbol{Y}_{ti}\}$ (with similar notation for $\boldsymbol{X}_{1,i}^{t}$, $\alpha_{1,i}^{t}$ and $\boldsymbol{\psi}_1^t$) as the collection of response values (and covariates, latent school effects and parameters, respectively) for the ith school from time period one to time period t, the joint distribution generating $\{(\boldsymbol{Y}_{1,i}^{t}, \boldsymbol{X}_{1,i}^{t}, \alpha_{1,i}^{t}, \psi_1^t) : t = 1, \ldots, T\}$ for each school i is sequentially decomposed as follows:

$$\boldsymbol{Y}_{ti} \perp\!\!\!\perp \boldsymbol{Y}_{1,i}^{t-1}, \boldsymbol{X}_{1,i}^{T}, \alpha_{1,i}^{t}, \boldsymbol{\psi}_1^{T} \mid \boldsymbol{X}_{ti}, \alpha_{ti}, \boldsymbol{\beta}_t, \sigma_t^2 \quad t \geq 2; \tag{2.1}$$

$$(\boldsymbol{Y}_{ti} \mid \boldsymbol{X}_{ti}, \alpha_{ti}, \boldsymbol{\beta}_t, \sigma_t^2) \sim \mathcal{N}(\boldsymbol{X}_{ti}\boldsymbol{\beta}_t + \alpha_{ti}\boldsymbol{\iota}_{n_{ti}}, \sigma_t^2 \boldsymbol{I}_{n_{ti}}), \quad t \geq 2; \tag{2.2}$$

$$\alpha_{ti} \perp\!\!\!\perp \boldsymbol{Y}_{1,i}^{t-1}, \boldsymbol{X}_{1,i}^{T}, \alpha_{1,i}^{t-1}, \boldsymbol{\psi}_1^{T} \mid \boldsymbol{Y}_{t-1i}, \alpha_{t-1i}, \phi_{0t}, \phi_{1t}, \gamma_t, \tau_t^2 \quad t \geq 2; \tag{2.3}$$

$$(\alpha_{ti} \mid \boldsymbol{Y}_{t-1i}, \alpha_{t-1i}, \phi_{0t}, \phi_{1t}, \gamma_t, \tau_t^2) \sim \mathcal{N}(\phi_{0t} + \phi_{1t}\alpha_{t-1i} + \gamma_t \overline{Y}_{t-1i}, \tau_t^2(1 - \phi_{1t}^2)), \quad t \geq 2; \tag{2.4}$$

$$\boldsymbol{Y}_{1i} \perp\!\!\!\perp \boldsymbol{X}_{2,i}^{T}, \boldsymbol{\psi}_1^{T} \mid \boldsymbol{X}_{1i}, \alpha_{1i}, \boldsymbol{\beta}_1, \sigma_1^2; \tag{2.5}$$

$$(\boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \alpha_{1i}, \boldsymbol{\beta}_1, \sigma_1^2) \sim \mathcal{N}(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1 + \alpha_{1i}\boldsymbol{\iota}_{n_{1i}}, \sigma_1^2 \boldsymbol{I}_{n_{1i}}); \tag{2.6}$$

$$\alpha_{1i} \perp\!\!\!\perp \boldsymbol{X}_{1,i}^{T}, \boldsymbol{\psi}_1^{T} \mid \phi_{01}, \tau_1^2; \tag{2.7}$$

$$(\alpha_{1i} \mid \phi_{01}, \tau_1^2) \sim \mathcal{N}(\phi_{01}, \tau_1^2); \tag{2.8}$$

$$\boldsymbol{X}_{1,i}^{T} \perp\!\!\!\perp \boldsymbol{\psi}_1^{T}; \tag{2.9}$$

$$\boldsymbol{X}_{1,i}^{T} \text{ is left unspecified;} \tag{2.10}$$

$$\boldsymbol{\psi}_1^{T} \sim \pi_{\boldsymbol{\psi}}, \tag{2.11}$$

where the parameter $\boldsymbol{\psi}_t \doteq (\boldsymbol{\beta}_t, \sigma_t^2, \tau_t^2, \phi_{0t}, \gamma_t, \phi_{1t}) \in \mathbb{R}^{p_t} \times \mathbb{R}_+^2 \times \mathbb{R}^2 \times [-1, 1]$ for $t \geq 2$ and $\boldsymbol{\psi}_1 \doteq (\boldsymbol{\beta}_1, \sigma_1^2, \tau_1^2, \phi_{01}) \in \mathbb{R}^{p_1} \times \mathbb{R}_+^2 \times \mathbb{R}$. Here, $\boldsymbol{\iota}_n$ is a vector of ones that is of length n and $\overline{Y}_{t-1i} = \frac{1}{n_{t-1i}} \sum_{j=1}^{n_{t-1i}} Y_{t-1ij}$. The symbol $\doteq$ means "defined as". Note that in (2.1), (2.3) and (2.5), we consider $\boldsymbol{X}_{1,i}^{T}$ rather than $\boldsymbol{X}_{1,i}^{t}$ because both $\boldsymbol{Y}_{ti}$ and α_{ti} are related to the covariates at time t only, namely $\boldsymbol{X}_{ti}$; similarly, for the parameters. To understand the logic behind this sequential specification, readers are encouraged to consult Section C of the Supplementary Material, where such a decomposition is explained for the case of $T = 2$ cohorts.

A few more detailed comments regarding the sequential specification of our time-dependent value-added model are warranted.

1. For each cohort t, condition (2.1) implies that $\boldsymbol{Y}_{ti}$ is stochastically determined by the covariates $\boldsymbol{X}_{ti}$ and the random school effect α_{ti}; the parameters characterizing this conditional distribution are (β_t, σ_t^2). Condition (2.2) not only defines a specific functional relationship of the conditional expectation $E(Y_{tij} \mid \boldsymbol{X}_{ti}, \alpha_{ti})$, but also makes explicit that, conditionally on $(\boldsymbol{X}_{ti}, \alpha_{ti}, \boldsymbol{\beta}_t, \sigma_t^2)$, the Y_{tij}'s are mutually independent. This condition, known as the *Axiom of Local Independence*, defines the school effect in the sense that it explains the heterogeneity that is present in the Y_{tij}'s and that is not explained by the covariates $\boldsymbol{X}_{tij}$; for details, see Manzi et al. (2014) and Page et al. (2017). Similar comments can be made for conditions (2.5) and (2.6).
2. Conditions (2.3) and (2.4) specify the temporal dependence of the proposed value-added model: the school effect α_{ti} that impacts the t-th cohort's performance is determined by both the school effect α_{t-1i} and the information shock contained in the post-tests of

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

cohort $t-1$. This conditional model is parameterized by $(\phi_{0t}, \phi_{1t}, \gamma_t, \tau_t^2)$. Conditions (2.7) and (2.8) specify the initial condition at the school effect level; this model is parameterized by (ϕ_{01}, τ_1^2).

3. Condition (2.10) means that the covariates $\boldsymbol{X}_{1,i}^T$ are exogenous with respect to the school effect α_{1i}. This exogeneity explains why the covariates are left unspecified as stated in condition (2.10). For details on exogeneity, see Engle et al. (1983) and Mouchart and Oulhaj (2006).

Remark 1. An important and well-known approach to modeling value-added with multiple test measures over time is provided by growth models. Growth models with latent variables have been considered to measure achievement, offering a pathway to model value-added across multiple measures (e.g., Bianconcini & Cagnone, 2012). In this remark, we aim to contrast this approach with the model studied in the current paper. Firstly, in growth models, an invariant group of people is subject to multiple measures over a time period. By contrast, in our approach, the composition of the group of individuals changes over time. Each group is measured twice (pre- and post-test), allowing the identification of the school effect for each period. Another difference lies in the interpretation of the random effect (or latent variable, in the parlance of growth models). Growth models consider the random effect to be invariant over time, meaning that it captures the value of the school after observing the full measures over the time period. In our modeling strategy, conversely, the latent random effect is dynamic, meaning that it changes over time as it captures the school effect over each period (1, 2], (2, 3], etc. □

2.2. Likelihood Function

The sequential specification (2.1)–(2.11) corresponds to a Bayesian decomposition of the joint distribution of $(\boldsymbol{Y}_{1,i}^T, \boldsymbol{X}_{1,i}^T, \alpha_{1,i}^T, \boldsymbol{\psi}_1^T)$ for each school i across varying numbers of cohorts, denoted by T. The question concerns the criterion for selecting the likelihood function or statistical model, which is characterized by generating the observations alone. In other words, are the school effects $\alpha_{i,1}^T$ treated as parameters of the likelihood function, or is the likelihood derived after integrating them out? This inquiry is closely tied to the two overarching perspectives in value-added model construction: one that treats the school effect as a random effect, while the other regards it as a fixed effect (see, among many others, Aitkin and Longford (1986), Tekwe et al. (2004)).

In the fixed-effect perspective, the school intercept is included among the parameters of the likelihood function, whereas in the random-effect perspective, the likelihood function is derived after integrating out the school effect, making the school effect not a parameter of the likelihood function. The decision between the fixed-effect or random-effect approach should be based on the induced statistical model, that is, on our "understanding of the way in which the data are supposed to, or did in fact, originate" (Fisher, 1973, p.8). In the context of school effectiveness, it's crucial to conceptualize the school effect by considering its impact on observable variables. In the fixed-effect perspective, the school effect contributes the *same amount* to the student achievement regressed onto previous attainment scores (and potentially other factors). Thus, a school in this framework is seen as an entity adding a constant effect to each student's predicted achievement, without relating achievement scores among students. On the other hand, in a random-effect approach, the school effect is defined through the Axiom of Local Independence, which implies that the school effect explains the heterogeneity that is present in the students achievement and that is not explained by the previous attaintment scores (and possibly other factors). Consequently, under this approach, a school is perceived as an entity that introduces heterogeneity in students' achievements. This criterion is sometimes utilized in econometrics to differentiate between fixed-effect and random-effect models (Hsiao, 2014). Additionally, when considering the type of data generating process being modeled, empirical comparisons between models attempting to characterize different phenomena—such as contrasting the corresponding estimates of value-added

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

indicators from both approaches (Longford, 2012; Clarke et al., 2015) —might not necessarily aid in selecting the appropriate model. For details on this criterion, the reader is referred to the Supplementary Material, Appendix B.

Following Manzi et al. (2014) and Page et al. (2017), we adhere to the random-effect paradigm because of the underlying conception of school. Consequently, the likelihood function is derived after integrating out the school effects. The subsequent result derives the explicit joint distribution of the observations $\boldsymbol{Y}_{1,i}^T$ given $(\boldsymbol{X}_{1i}^T, \boldsymbol{\psi}_1^T)$:

Theorem 1. *Given the sequential model (2.1)–(2.11), the joint distribution of* $\boldsymbol{Y}_{1,i}^T$ *given* $(\boldsymbol{X}_{1i}^T, \boldsymbol{\psi}_1^T)$ *is a normal distribution for every school* i. *For* $T = 2$ *cohorts, its expectation is given by*

$$\begin{pmatrix} \eta_{02}\boldsymbol{\iota}_{n_{2i}} + \boldsymbol{X}_{2i}\boldsymbol{\beta}_2 + \gamma_2\overline{\boldsymbol{X}}_{1i}\boldsymbol{\beta}_1\boldsymbol{\iota}_{\boldsymbol{n_{2i}}} \\ \phi_{01}\boldsymbol{\iota}_{n_{1i}} + \boldsymbol{X}_{1i}\boldsymbol{\beta}_1 \end{pmatrix} \tag{2.12}$$

where $\overline{\boldsymbol{X}}_{1i} = \frac{1}{n_{1i}}\boldsymbol{\iota}'_{n_{1i}}\boldsymbol{X}_{1i}$ *is a* $1 \times p_1$ *vector of empirical means at the school level so that* $\overline{\boldsymbol{X}}_{1i}\boldsymbol{\beta}_1$ *is a scalar, and* $\eta_{02} \doteq \{\phi_{02} + \phi_{01}(\phi_{12} + \gamma_2)\}$; *and its variance-covariance matrix is given by*

$$\begin{pmatrix} \omega_{2i}\,\boldsymbol{J}_{n_{2i}} + \sigma_2^2\boldsymbol{I}_{n_{2i}} & \delta_{12i}\,\boldsymbol{\iota}_{n_{2i}}\boldsymbol{\iota}'_{n_{1i}} \\ & \tau_1^2\boldsymbol{J}_{n_{1i}} + \sigma_1^2\boldsymbol{I}_{n_{1i}} \end{pmatrix}, \tag{2.13}$$

where

$$\omega_{2i} \doteq \tau_1^2\,(\phi_{12} + \gamma_2)^2 + \frac{\gamma_2^2\sigma_1^2}{n_{1i}} + \tau_2^2(1 - \phi_{12}^2), \qquad \delta_{12i} \doteq \phi_{12}\tau_1^2 + \gamma_2\left(\tau_1^2 + \frac{\sigma_1^2}{n_{1i}}\right) \tag{2.14}$$

and $\boldsymbol{J}_n \doteq \boldsymbol{\iota}_n\boldsymbol{\iota}'_n$.

For $T > 2$ *cohorts, the expectation is given by*

$$E(\boldsymbol{Y}_{1,i}^t \mid \boldsymbol{X}_{1i}^T, \boldsymbol{\psi}_1^T) = \boldsymbol{X}_{ti}\boldsymbol{\beta}_t + E(\alpha_{ti} \mid \boldsymbol{X}_{1i}^T, \boldsymbol{\psi}_1^T) \quad \text{for } 3 \le t \le T, \tag{2.15}$$

where, for $3 \le t \le T$,

$$\begin{aligned} E(\alpha_{ti} \mid \boldsymbol{X}_{1i}^T, \boldsymbol{\psi}_1^T) = \phi_{0t} &+ \sum_{\ell=2}^{t}\prod_{k=\ell}^{t}(\phi_{1k} + \gamma_k)\phi_{0,\ell-1} \\ &+ \gamma_t\overline{\boldsymbol{X}}_{t-1,i}\boldsymbol{\beta}_{t-1} + \sum_{\ell=2}^{t}\prod_{k=\ell}^{t}(\phi_{1k} + \gamma_k)\gamma_{\ell-1}\overline{\boldsymbol{X}}_{\ell-2,i}\boldsymbol{\beta}_{\ell-2} \end{aligned} \tag{2.16}$$

the variances are given by

$$V(\boldsymbol{Y}_{ti} \mid \boldsymbol{X}_{1i}^T, \boldsymbol{\psi}_1^T) = V(\alpha_{ti} \mid \boldsymbol{X}_{1i}^t, \boldsymbol{\psi}_1^T)\iota_{n_{ti}}\iota_{n_{ti}}^t;$$

and the covariances are given for every $1 \leqslant s < t \leqslant T$ *by*

$$cov(\boldsymbol{Y}_{si}, \boldsymbol{Y}_{ti} \mid \boldsymbol{X}_{1i}^T, \boldsymbol{\psi}_1^T) = V(\alpha_{si} \mid \boldsymbol{X}_{1i}^t, \boldsymbol{\psi}_1^T)\iota_{n_{ti}}\iota_{n_{si}}^t\prod_{k=0}^{t-s-1}(\phi_{1,t-k} + \gamma_{t-k}).$$

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

where

$$V(\alpha_{ti} \mid \boldsymbol{X}_{1i}^{t}, \boldsymbol{\psi}_{1}^{T}) = V(\alpha_{1i} \mid \boldsymbol{X}_{1i}^{t}, \boldsymbol{\psi}_{1}^{T}) \prod_{k=0}^{t-1} A_k + \sum_{k=0}^{t-2} B_k \prod_{l=-1}^{k-1} A_l$$

with $A_k = (\phi_{1,t-k} + \gamma_{t-k})^2$ *and* $B_k = n_{t-k-1,i}^{-1} \gamma_{t-k}^2 \sigma_{t-k-1}^2 + \tau_{t-k}^2 (1 - \phi_{1,t-k}^2)$ *for every* $0 \leqslant k \leqslant t-1$ *and* $A_{-1} = 1$.

The proof of this result is to be found in the Supplementary Material: Section C provides the details for $T = 2$ cohorts; Section D provides the details for $T \geq 3$ cohorts.

Let us conclude by noting that in the Bayesian specification, if the distributions of random effects are interpreted as prior distributions, then the Bayesian model corresponds to the equivalent of the classical fixed-effect model. On the other hand, the Bayesian equivalent of the classical random-effect model entails using the probability distribution obtained after integrating the random effects as the likelihood function (or statistical model), which is the approach we have chosen. For further discussion on the importance of explicitly defining the likelihood function within a Bayesian framework, please refer to the Supplementary Material, Section A.

2.3. *Parameter Identification for Two Cohorts*

According to the Likelihood Principle, for a given model all the information the data contains about the model parameters is given by the likelihood function (Lindley, 1983). We argue that such information is related to the *identified parameters* only; for a discussion, see Supplementary Material, Section A. Thus, in the context of the temporally dynamic, cohort-varying value-added model, parameter identifiability should be analyzed with respect to the conditional distribution of $\boldsymbol{Y}_{1,i}^{T}$ given $(\boldsymbol{X}_{1i}^{T}, \boldsymbol{\psi}_{1}^{T})$. It's worth emphasizing that although both model specification and parameter estimation are entirely Bayesian, the identifiability of parameters should also be established beforehand. Furthermore, it should be noted that the prior distribution on $\boldsymbol{\psi}_{1}^{T}$ has no impact on the identification analysis, except through events where the prior probability equals 0 or 1. For detailed insights, please see Supplementary Material, Section A. This elucidates why, in the sequential specification, the prior distribution $\boldsymbol{\psi}_{1}^{T}$ in (2.11) remains unspecified.

Following the identification strategy outlined in San Martín et al. (2011), San Martín et al. (2013), San Martín et al. (2015), and Fariña et al. (2019), we demonstrate the identification of $(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)$ for the case of $T = 2$ cohorts through the following arguments, which remain analogous for more than two cohorts:

1. From the conditional expectation of $\boldsymbol{Y}_{1i}$, ϕ_{01} and $\boldsymbol{\beta}_1$ are identifiable provided that $\boldsymbol{\iota}_{n_{1i}} \notin \mathcal{R}(\boldsymbol{X}_{1i})$, which holds by construction. Here $\mathcal{R}(\boldsymbol{A})$ represents the space generated by the columns of the matrix $\boldsymbol{A}$.
2. From the conditional expectation of $\boldsymbol{Y}_{2i}$, $\eta_{02} + \gamma_2 \overline{\boldsymbol{X}}_{1i} \boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are identifiable provided that $\boldsymbol{\iota}_{n_{2i}} \notin \mathcal{R}(\boldsymbol{X}_{2i})$, which holds by construction.
3. From the variance of $\boldsymbol{Y}_{1i}$, τ_1^2 and σ_1^2 are identified. Similarly, from the variance of $\boldsymbol{Y}_{2i}$, σ_2^2 and ω_{2i} are identified for all school i.
4. From the covariance between $\boldsymbol{Y}_{1i}$ and $\boldsymbol{Y}_{2i}$, δ_{12i} are identified for all school i. Now, if there exist at least two schools i and k such that, for the first cohort, the total number of students is different, then

$$\delta_{12i} - \delta_{12k} = \gamma_2 \sigma_1^2 \left(\frac{1}{n_{1i}} - \frac{1}{n_{2k}} \right),$$

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

from which it follows that γ_2 is identified. Furthermore, using the definition of δ_{12i}, it follows that ϕ_{12} is identified; and using the definition and identifiability of ω_{2i}, it follows that τ_2^2 is identified.

5. Using the identifiability of γ_2, $\boldsymbol{\beta}_1$ and $\eta_{02} + \gamma_2 \overline{\boldsymbol{X}}_{1i}\boldsymbol{\beta}_1$, it follows that η_{02} is identifiedx'. Finally, from the definition of η_{02}, it follows that ϕ_{02} is identified. □

We summarize the previous arguments in the following proposition:

Proposition 1. *In the statistical model parameterized by* (2.12) *and* (2.13)*, the parameters* $(\boldsymbol{\psi}_1, \boldsymbol{\psi}_2)$ *are identified provided that* $\boldsymbol{\iota}_{n_{1i}} \notin \mathcal{R}(\boldsymbol{X}_{1i})$, $\boldsymbol{\iota}_{n_{2i}} \notin \mathcal{R}(\boldsymbol{X}_{2i})$ *and there exist at least two schools* i *and* k *such that, for the first cohort, the total number of students is different.*

In the remainder of this paper, the discussion is focused on $T = 2$ cohorts due to the nature of our case study.

2.4. Nested Value-Added Models

The sequential specification (2.1)–(2.11) for $T = 2$ cohorts (referred to as the "Full Model" or also as *Model 3*) aims to address whether a school's effectiveness upon receiving a new cohort is jointly influenced by both the effectiveness of the school for the previous cohort and some observed information from that cohort. The identification analysis assures us that such a scenario is empirically plausible. The crux of the underlying model lies in the temporal dependency of the latent school effect α_{2i} for cohort 2 on both the latent school effect α_{1i} for cohort 1 and the mean of the cohort 1 post-tests Y_{1ij}'s, as captured by the conditional distribution (2.4). At the statistical model level, the parameters of this conditional distribution, alongside σ_1^2 and τ_1^2, delineate both the within- and between-cohorts dependencies among the post-test scores. As a matter of fact,

$$cov(Y_{2ir}, Y_{2is} \mid \boldsymbol{X}_{1,i}^2, \boldsymbol{\psi}_1^2) = \tau_1^2(\phi_{12} + \gamma_2)^2 + \frac{\gamma_2^2\sigma_1^2}{n_{1i}} + \tau_2^2(1 - \phi_{12}^2), \quad r \neq s; \tag{2.17}$$

$$cov(Y_{1ij}, Y_{2ik} \mid \boldsymbol{X}_{1,i}^2, \boldsymbol{\psi}_1^2) = \tau_1^2(\phi_{12} + \gamma_2) + \frac{\gamma_2\sigma_1^2}{n_{1i}}, \quad j \neq k. \tag{2.18}$$

It can be noticed that the within-cohort covariance (2.17) is positive for all $\gamma_2 \in \mathbb{R}$ and $\phi_{12} \in [-1, 1]$, and hence the corresponding correlation; this is a standard fact in value-added models. Even more interesting is that the post-tests scores of the two cohorts are correlated; its sign basically depends on both the sign of $\phi_{12} + \gamma_2$ (which in turn determines the sign of the correlation between the school effects α_{1i} and α_{2i}) and the sign of γ_2:

1. If the information shock parameter γ_2 is such that $|\gamma_2| \geq 1$, for all schools the sign of (2.18) as well as of the correlation between α_{1i} and α_{2i} corresponds to the sign of γ_2.
2. If the information shock parameter γ_2 is such that $|\gamma_2| < 1$, then we distinguish the following cases:
   (a) If $\gamma_2 \in (-1, 0)$ and $\phi_{12} + \gamma_2 < 0$ then, for all schools, the sign of (2.18) is negative.
   (b) If $\gamma_2 \in (-1, 0)$ and $\phi_{12} + \gamma_2 > 0$ then, for a school i, the sign of (2.18) is positive if $\tau_1^2 > \sigma_1^2/n_{1i}$; and it is negative if $\tau_1^2 < \sigma_1^2/n_{1i}$.
   (c) If $\gamma_2 \in (0, 1)$ and $\phi_{12} + \gamma_2 > 0$ then, for all schools, the sign of (2.18) is positive.
   (d) If $\gamma_2 \in (0, 1)$ and $\phi_{12} + \gamma_2 < 0$ then, for a school i, the sign of (2.18) is positive if $\tau_1^2 < \sigma_1^2/n_{1i}$; and it is negative if $\tau_1^2 > \sigma_1^2/n_{1i}$.

   Note also that, for $|\gamma_2| < 1$, the sign of the correlation between α_{1i} and α_{2i} corresponds, for all schools, to the sign of $\phi_{12} + \gamma_2$.

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

These considerations motivate considering special cases of Model 3 that are based on particular values for the parameters ϕ_{12} and γ_2. This results in three models that are nested in Model 3:

1. Model 0 is obtained by setting $(\phi_{12}, \gamma_2) = (0, 0)$. In this case, the within-cohort covariance (2.17) is equal to τ_2^2, whereas both the between-cohort covariance (2.18) and covariance between α_{1i} and α_{2i} are equal to 0. In other words, all temporal dependence between cohorts breaks and therefore this model reflects the situation in which a school that does not learn from its past, nor does it intend to affect its future. Under condition $(\phi_{12}, \gamma_2) = (0, 0)$, it follows that

$$\alpha_{2i} \perp\!\!\!\perp \boldsymbol{X}_{1,i}^2, \boldsymbol{Y}_{1i}, \alpha_{1i}, \boldsymbol{\psi}_1^2 \mid \phi_{02}, \tau_2^2. \tag{2.19}$$

   Consequently, Model 0 can be specified hierarchically as

   (a) $(\boldsymbol{Y}_{2i} \mid \boldsymbol{X}_{2i}, \alpha_{2i}, \boldsymbol{\psi}_1^2) \sim \mathcal{N}(\boldsymbol{X}_{2i}\boldsymbol{\beta}_2 + \alpha_{2i}\boldsymbol{\iota}_{n_{2i}}, \sigma_2^2 \boldsymbol{I}_{n_{2i}})$;
   (b) $(\alpha_{2i} \mid \boldsymbol{\psi}_1^2) \sim \mathcal{N}(\phi_{02}, \tau_2^2)$;
   (c) $(\boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \alpha_{1i}, , \boldsymbol{\psi}_1^2) \sim \mathcal{N}(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1 + \alpha_{1i}\boldsymbol{\iota}_{n_{1i}}, \sigma_1^2 \boldsymbol{I}_{n_{1i}})$;
   (d) $(\alpha_{1i} \mid \boldsymbol{\psi}_1^2) \sim \mathcal{N}(\phi_{01}, \tau_1^2)$.

2. Model 1 is obtained by setting $\gamma_2 = 0$. In this case, the within-cohort covariance (2.17) is equal to $\tau_1^2\phi_{12}^2 + \tau_2^2(1-\phi_{12}^2)$, which is always positive for all $\phi_{12} \in [-1, 1]$. The between-cohort covariance (2.18) is equal to $\phi_{12}\tau_1^2$, which is equivalent to the covariance between α_{2i} and α_{1i}: if the school effect for cohort 2 is negatively (resp., positively) correlated to the school effect for cohort 1, then the between-cohort correlation is negative (resp. positive). In other words, the relationship that the school effect has with its past school effect is reflected (at the sign level) in the observed relationship between the post-test scores between the two cohorts. Under condition $\gamma_2 = 0$, it follows that

$$\alpha_{2i} \perp\!\!\!\perp \boldsymbol{X}_{1,i}^2, \boldsymbol{Y}_{1i}, \boldsymbol{\psi}_1^2 \mid \alpha_{1i}, \phi_{02}, \phi_{12}, \tau_2^2. \tag{2.20}$$

   Consequently, Model 1 can be specified hierarchically as

   (a) $(\boldsymbol{Y}_{2i} \mid \boldsymbol{X}_{2i}, \alpha_{2i}, \boldsymbol{\psi}_1^2) \sim \mathcal{N}(\boldsymbol{X}_{2i}\boldsymbol{\beta}_2 + \alpha_{2i}\boldsymbol{\iota}_{n_{2i}}, \sigma_2^2 \boldsymbol{I}_{n_{2i}})$;
   (b) $(\alpha_{2i} \mid \alpha_{1i}, , \boldsymbol{\psi}_1^2) \sim \mathcal{N}(\phi_{02} + \phi_{12}\alpha_{1i}, \tau_2^2(1 - \phi_{12}^2))$;
   (c) $(\boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \alpha_{1i}, , \boldsymbol{\psi}_1^2) \sim \mathcal{N}(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1 + \alpha_{1i}\boldsymbol{\iota}_{n_{1i}}, \sigma_1^2 \boldsymbol{I}_{n_{1i}})$;
   (d) $(\alpha_{1i} \mid, \boldsymbol{\psi}_1^2) \sim \mathcal{N}(\phi_{01}, \tau_1^2)$.

   As it can be recognized, this structure corresponds to an ARIMA-type model.

3. Model 2 is obtained by setting $\phi_{12} = 0$. In this case, the within-cohort covariance (2.17) is equal to $\tau_1^2\gamma_2^2 + \gamma_2^2\sigma_1^2/n_{1i} + \tau_2^2$, which is always positive for all $\gamma_2 \in \mathbb{R}$. The between-cohort covariance (2.18) is equal to $\gamma_2(\tau_1^2 + \sigma_1^2/n_{1i})$: its sign depends on the sign of γ_2. Furthermore, if $\gamma_2 > 0$ (resp., $\gamma_2 < 0$), the between-cohort covariance is larger (resp., smaller) than the covariance between α_{2i} and α_{1i}. In other words, the shock of information captured by γ_2 impacts both the correlation between the school effects and the between-cohort correlation: these dependency relationships provide an idea of what the model means by "shock of information". Under condition $\phi_{12} = 0$, it follows that

$$\alpha_{2i} \perp\!\!\!\perp \boldsymbol{X}_{1,i}^2, \alpha_{1i}, \boldsymbol{\psi}_1^2 \mid \boldsymbol{Y}_{1i}, \phi_{02}, \gamma_2, \tau_2^2. \tag{2.21}$$

   Consequently, Model 2 can be specified hierarchically as

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

(a) $(\boldsymbol{Y}_{2i} \mid \boldsymbol{X}_{2i}, \alpha_{2i}, \boldsymbol{\psi}_1^2) \sim \mathcal{N}(\boldsymbol{X}_{2i}\boldsymbol{\beta}_2 + \alpha_{2i}\boldsymbol{\iota}_{n_{2i}}, \sigma_2^2 \boldsymbol{I}_{n_{2i}})$;
(b) $(\alpha_{2i} \mid \boldsymbol{Y}_{1i}, \boldsymbol{\psi}_1^2) \sim \mathcal{N}(\phi_{02} + \gamma_2 \bar{Y}_{1i\bullet}, \tau_2^2)$;
(c) $(\boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \alpha_{1i}, \boldsymbol{\psi}_1^2) \sim \mathcal{N}(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1 + \alpha_{1i}\boldsymbol{\iota}_{n_{1i}}, \sigma_1^2 \boldsymbol{I}_{n_{1i}})$;
(d) $(\alpha_{1i} \mid \boldsymbol{\psi}_1^2) \sim \mathcal{N}(\phi_{01}, \tau_1^2)$.

Since

$$(2.19) \Longrightarrow (2.20) \Longrightarrow (2.3) \text{ with } T = 2$$
$$(2.19) \Longrightarrow (2.21) \Longrightarrow (2.3) \text{ with } T = 2,$$

Model 0 is nested in Model 1, which in turn is nested in Model 3; and Model 0 is nested in Model 2, which in turn is nested in Model 3.

2.5. *Value-Added Definition*

The concept of school value-added refers to the difference between the expected grade of a student at a given school and the expected grade of the same student at an *average* school. It is, in other terms, the gain (or loss) in the expected score at a specific school compared to a baseline established by an average score.

A model-free definition of school value added was introduced in Manzi et al. (2014). In the notation of the present paper, it is given by

$$VA_{ti}(\boldsymbol{X}_{ti}) \doteq \frac{1}{n_{ti}} \sum_{j=1}^{n_{ti}} \left[E(Y_{tij} \mid \boldsymbol{X}_{tij}, \alpha_i) - E(Y_{tij} \mid \boldsymbol{X}_{tij}) \right]. \tag{2.22}$$

It may be necessary to clarify what "model-free definition" means in this context. This definition involves conditional expectations that do not refer to any specific model, although later in this paper, we shall compute them within the context of the precise model presented from equations (2.1) to (2.11). Indeed, conditional expectations are related to the statistical distribution of data and, like expectations, can be defined and estimated without referring to any model. In probability theory, conditional expectations can be defined either through orthogonal projection of data (Neveu, 1972; Florens et al., 2007) or, more broadly, by the Radon–Nikodym Theorem (Kolmogorov, 1950; Billingsley, 1968). We view the aforementioned definition as advantageous because it provides a statistical interpretation of the concept of school value-added that is not reliant on any specific model. Therefore, it can be applied to any particular situation where the psychometrician *assumes* a model (or compares them).

Definition (2.22) offers a characterization of the average or reference school that is entirely determined by the vector of covariates $\boldsymbol{X}_{ti}$. If the covariates change, it impacts not only the conceptualization of the average school but also the interpretation of school effectiveness and the value-added indicators themselves. Consequently, this definition implies that school effectiveness should not be viewed as a universally meaningful concept (i.e., a school being effective or ineffective in the same manner under all circumstances). Instead, it is a contextually idiosyncratic concept that should not be reified. The relevant context for this analysis is defined by the covariates included in the model. Therefore, their selection should be closely tied to the policy context requiring a value-added analysis, as well as to the social context within which an educational system operates.

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

Using (2.22), we derive the following value-added indicator for Model 3:

$$\begin{aligned} VA_{1i}(\boldsymbol{X}_{1i}) &= \alpha_{1i} - \phi_{01}, \\ VA_{2i}(\boldsymbol{X}_{2i}) &= \alpha_{2i} - \big[\phi_{02} + \phi_{01}(\phi_{12} + \gamma_2)\big] - \frac{\gamma_2}{n_{2i}} \boldsymbol{\beta}_1' \sum_{j=1}^{n_{2i}} E(\overline{\boldsymbol{X}}_{1i}' \mid \boldsymbol{X}_{2ij}). \end{aligned} \tag{2.23}$$

For further details, see "Appendix A.1". The value-added indicators for Models 0, 1, and 2 are obtained by setting $\phi_{12} = \gamma_2 = 0$, $\gamma_2 = 0$, and $\phi_{12} = 0$, respectively.

Note that $VA_{2i}(\boldsymbol{X}_{2i})$ represents the school effect α_{2i} (centered at 0 by $\phi_{02} + \phi_{01}(\phi_{12} + \gamma_2)$), adjusted by an additive term dependent on $(n_{2i})^{-1} \sum_{j=1}^{n_{2i}} E(\overline{\boldsymbol{X}}_{1i}' \mid \boldsymbol{X}_{2ij})$. Assuming this regression to be linear, it can be demonstrated that

$$\frac{1}{n_{2i}} \sum_{j=1}^{n_{2i}} E(\overline{\boldsymbol{X}}_{1i}' \mid \boldsymbol{X}_{2ij}) = \begin{pmatrix} b_{10} & b_{11} & 0 & \cdots & 0 \\ b_{20} & 0 & b_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 \\ b_{p0} & 0 & 0 & \cdots & b_{pp} \end{pmatrix} \begin{pmatrix} 1 \\ \overline{\boldsymbol{X}}_{2i} \end{pmatrix},$$

where $p \doteq p_1 = p_2$; in this form, the regression parameters are identifiable. For more details, see "Appendix A.2".

Remark 2. For $T \geq 3$ cohorts, the value-added $VA_{Ti}(\boldsymbol{X}_{Ti})$ is given by

$$\begin{aligned} VA_{Ti}(\boldsymbol{X}_{Ti}) &= \alpha_{Ti} - E(\alpha_{Ti} \mid \boldsymbol{X}_{Ti}, \boldsymbol{\psi}_1^T) \\ &= \alpha_{Ti} - E[E(\alpha_{Ti} \mid \boldsymbol{X}_{1i}^T, \boldsymbol{\psi}_1^T) \mid \boldsymbol{X}_{Ti}, \boldsymbol{\psi}_1^T]. \end{aligned}$$

Using (2.16), we conclude that

$$\begin{aligned} VA_{Ti}(\boldsymbol{X}_{Ti}) = \alpha_{Ti} - \phi_{0t} - \sum_{\ell=2}^{t} \prod_{k=\ell}^{t} (\phi_{1k} + \gamma_k)\phi_{0,\ell-1} - \gamma_t E\left(\overline{\boldsymbol{X}}_{t-1,i}' \mid \boldsymbol{X}_{Ti}\right) \boldsymbol{\beta}_{t-1} - \\ \sum_{\ell=2}^{t} \prod_{k=\ell}^{t} (\phi_{1k} + \gamma_k)\gamma_{\ell-1} E\left(\overline{\boldsymbol{X}}_{\ell-2,i}' \mid \boldsymbol{X}_{Ti}\right) \boldsymbol{\beta}_{\ell-2}. \end{aligned}$$

□

2.6. *Structural Interpretation of the Persistence of School Effectiveness*

Time-dependent value-added models are intended to model the persistence of school effectiveness. Following Gray et al. (1996, 1999), the persistence of school effectiveness is described through trajectories of school value-added. This meaning of persistence critically depends on the time-dependent value-added model that is used. Studies devised to describe the persistence of school effectiveness have appeared in the literature, but are (implicitly) based on Model 0 (see, e.g., Gray et al., 2001, Thomas et al., 2007, Bellei et al., 2016). These approaches are limited because they assume that a school's current effectiveness is not affected by what the school did in the past. It seems reasonable to expect that a school's past performance would be useful in determining the future effectiveness of the school. Our approach that is based on Model 3, including its particular cases Model 1 and Model 2, overcomes this limitation.

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

Under the Model 0, Model 1, Model 2 or Model 3, the school valued-added for cohort 1 coincide with their respective school effect α_{1i} centered at 0 by ϕ_{01}. To understand the extent to which the school value-added for cohort 1 explains the school value-added for cohort 2, we decompose the latter into two components: the first component captures the explanation of the second value-added by the first one, whereas the second corresponds to everything of the second value-added that is not explained by the first; that is,

$$VA_{2i}(\boldsymbol{X}_{2i}) = E\left[VA_{2i}(\boldsymbol{X}_{2i}) \mid VA_{1i}(\boldsymbol{X}_{1i}), \boldsymbol{X}_{1,i}^2, \boldsymbol{\psi}_1^2\right] + \left\{VA_{2i}(\boldsymbol{X}_{2i}) - E\left[VA_{2i}(\boldsymbol{X}_{2i}) \mid VA_{1i}(\boldsymbol{X}_{1i}), \boldsymbol{X}_{1,i}^2, \boldsymbol{\psi}_1^2\right]\right\}, \quad (2.24)$$

where

$$E\left[VA_{2i}(\boldsymbol{X}_{2i}) \mid VA_{1i}(\boldsymbol{X}_{1i}), \boldsymbol{X}_{1,i}^2, \boldsymbol{\psi}_1^2\right] = (\phi_{12} + \gamma_2)\, VA_{1i}(\boldsymbol{X}_{1i}) + \frac{\gamma_2}{n_{2i}} \sum_{j=1}^{n_{2i}} \left[\overline{\boldsymbol{X}}_{1i} - E(\overline{\boldsymbol{X}}_{1i} \mid \boldsymbol{X}_{2ij})\right] \boldsymbol{\beta}_1; \quad (2.25)$$

for a proof, see Supplementary Material, Section E. It can also be verified that

$$Var\left[E(VA_{2i}(\boldsymbol{X}_{2i}) \mid VA_{1i}(\boldsymbol{X}_{1i}), \boldsymbol{X}_{1,i}^2, \boldsymbol{\psi}_1^2) \mid \boldsymbol{X}_{1,i}^2, \boldsymbol{\psi}_1^2\right] = \tau_1^2(\phi_{12} + \gamma_2)^2, \quad (2.26)$$

and

$$Var\left[VA_{2i}(\boldsymbol{X}_{2i}) \mid \boldsymbol{X}_{1,i}^2, \boldsymbol{\psi}_1^2\right] = \omega_{2i}. \quad (2.27)$$

Since by construction both terms at the right hand of decomposition (2.24) are uncorrelated, it follows that the variance of the second term at the right hand (typically called *measurement error*) is given by

$$Var\left[VA_{2i}(\boldsymbol{X}_{2i}) - E\left(VA_{2i}(\boldsymbol{X}_{2i}) \mid VA_{1i}(\boldsymbol{X}_{1i}), \boldsymbol{X}_{1,i}^2, \boldsymbol{\psi}_1^2\right) \mid \boldsymbol{X}_{1,i}^2, \boldsymbol{\psi}_1^2\right] = \frac{\gamma_2^2 \sigma_1^2}{n_{1i}} + \tau_2^2(1 - \phi_{12}^2); \quad (2.28)$$

for a proof, see Supplementary Material, Section E.

To facilitate the interpretation of (2.25), consider the case of only one covariate, namely the pre-test. In this case, $p = 1$ and therefore

$$\frac{1}{n_{2i}} \sum_{j=1}^{n_{2i}} E(\overline{X}_{1i} \mid X_{2ij}) = E(\overline{X}_{1i} \mid \overline{X}_{2i})$$

so that the average pre-test of the first cohort is regressed on the average pre-test of the second cohort. Then:

1. In the explanation of $VA_{2i}(\boldsymbol{X}_{2i})$ by $(VA_{1i}(\boldsymbol{X}_{1i}),\ \boldsymbol{X}_{1,i}^2)$, there are two components:

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

(a) The first component depends on $(\phi_{12}+\gamma_2)\,VA_{1i}(\boldsymbol{X}_{1i})$: the parameter $(\phi_{12}+\gamma_2)$ determines the sign of the correlation between α_{1i} and α_{2i}: if such a correlation is positive (resp., negative), $VA_{1i}(\boldsymbol{X}_{1i})$ is amplified (resp., contracted) as an explanatory factor of $VA_{2i}(\boldsymbol{X}_{2i})$.

(b) The second component corresponds to the residual of the regression of the average pre-test of the first cohort on the average pre-test of the second cohort: it is actually a regression that inverts variables temporal order. Recall that the school has already treated the first cohort. Now, if the average pre-test of the second cohort is uncorrelated with the average pre-test of the first cohort, then the initial information provided by the second cohort is in every respect different from the initial information provided by the first cohort. The school is facing a new initial condition and, consequently, the residual is the bigger one. Similarly, if the average previous exam of the second cohort predicts the average previous exam of the first cohort, then the school has initial information similar from when the first cohort was treated and therefore the residual is the smaller one. This residual, pre-multiplying by γ_2, is the second explanatory factor. Note that this interpretation remains valid for the case of $T > 2$ cohorts, particularly regarding the role played by the parameters γ_t's and the regressions of the form $E\left(\overline{\boldsymbol{X}}'_{t,i} \mid \boldsymbol{X}_{Ti}\right)$.

2. How much of the variance of $VA_{2i}(\boldsymbol{X}_{2i})$ does $E\left[VA_{2i}(\boldsymbol{X}_{2i}) \mid VA_{1i}(\boldsymbol{X}_{1i}), \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1\right]$ explain? This question can be addressed by computing the so-called *reliability*, which in this case is given by

$$\frac{\tau_1^2(\phi_{12}+\gamma_2)^2}{\omega_{2i}}.$$

Note that it is always strictly less than 1, which means that $E\left[VA_{2i}(\boldsymbol{X}_{2i}) \mid VA_{1i}(\boldsymbol{X}_{1i}), \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1\right]$ does not exhaust the entire explanation of $VA_{2i}(\boldsymbol{X}_{2i})$: in the persistence of school effectiveness there are always new aspects that $VA_{1i}(\boldsymbol{X}_{1i})$ cannot predict.

Along with this general interpretation of school persistence, it is instructive to show how they are simplified in Models 1 and 2.

Under Model 1, $\gamma_2 = 0$ and therefore

$$E\left[VA_{2i}(\boldsymbol{X}_{2i}) \mid VA_{1i}(\boldsymbol{X}_{1i}), \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1\right] = \phi_{12}\,VA_{1i}(\boldsymbol{X}_{1i}), \tag{2.29}$$

and the reliability becomes equal to $\tau_1^2\phi_{12}^2/[\tau_1^2\phi_{12}^2+\tau_2^2(1-\phi_{12}^2)]$. As a result, the persistence of school value-added corresponds to a uniform reduction or shrinkage of $VA_{1i}(\boldsymbol{X}_{1i})$. Thus, an empirical analysis of the persistence of school effectiveness under Model 1 means analyzing to what extent a school that takes on new cohorts maintains the effectiveness achieved when accommodating the first cohort, leacing additional aspects included in $VA_{2i}(\boldsymbol{X}_{2i})$ unexplained by the past of the school. What is clear is that Model 1 is very different from Model 3.

Under Model 2, $\phi_{12} = 0$ and therefore

$$E\left[VA_{2i}(\boldsymbol{X}_{2i}) \mid VA_{1i}(\boldsymbol{X}_{1i}), \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1\right] = \gamma_2\,VA_{1i}(\boldsymbol{X}_{1i}) + \frac{\gamma_2}{n_{2i}}\sum_{j=1}^{n_{2i}}\left[\overline{\boldsymbol{X}}_{1i} - E(\overline{\boldsymbol{X}}_{1i} \mid \boldsymbol{X}_{2ij})\right]\boldsymbol{\beta}_1, \tag{2.30}$$

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

and the reliability becomes equal to

$$\frac{\tau_1^2\gamma_2^2}{\tau_1^2\gamma_2^2 + \frac{\gamma_2^2\sigma_1^2}{n_{1i}} + \tau_1^2}$$

which is still strictly smaller than 1. As a result, when the sign of $\gamma_2 + \phi_{12}$ in Model 3 is determined by γ_2, the structural interpretation of the persistence of school effectiveness in Model 2 is practically the same as Model 3. Taking into account such an interpretation, school effectiveness corresponds to what a school does/adds with/to a new cohort after discounting what the school learned to do/add while accommodating the previous cohort.

3. Computation and Model Fitting

We adopt a Bayesian approach and as a result prior distributions for all parameters for Models 0–3 need to be specified. For parameters common to each model we employ the following regularly used conjugate priors $\tau_t^2 \sim IG(1, 1)$, $\phi_{0t} \sim \mathcal{N}(0, 10^2)$, $\beta_t \sim \mathcal{N}(0, 10^2)$, and $\sigma_{ti}^2 \sim IG(1, 1)$ for $t = 1, 2$ and $i = 1, \ldots, m$. Here $IG(a, b)$ denotes an inverse gamma distribution with shape a and scale b. For Model 1 and 3 we use $\phi_{12} \sim UN(-1, 1)$ where UN denotes a Uniform distribution and for Model 2 and 3, $\gamma_2 \sim \mathcal{N}(0, 10^2)$. Fitting Models 0–3 and estimating value-added across time as described in Sect. 2.5 requires sampling from joint posterior distribution of all model parameters. To carry this out we developed a straightforward Markov Chain Monte Carlo (MCMC) algorithm that uses a Gibbs sampler to update all parameters on a one-by-one basis save ϕ_{12}. To update this parameter within the MCMC sampler a random walk metropolis step with a gaussian candidate proposal is used. All computer codes needed to fit models and estimate each school's value-added are provide in the `R` package `modernVA` (Page, 2020).

4. Simulation Study

4.1. Design of the Simulation Study

The objective of the simulation study is to explore the impact that ignoring temporal dependence may have on value-added estimates. The experiment consisted of using Models 0, 1, 2, and 3 as data generating mechanisms and then for each synthetic data, using Models 0, 1, 2, and 3 to estimate value-added across time for each institution. In more detail, we set $I = 250$ and $n_i = 25$ for $i = 1, \ldots, I$ and using Model 0, 1, 2, or 3 generated post-test scores (i.e., Y_1 and Y_2) by setting $\beta_1 = 0.6$, $\beta_2 = 0.75$, $\sigma_1^2 = \sigma_2^2 = 25$ and $\tau_1^2 = \tau_2^2 = 100$. The pre-test scores (i.e., X_1 and X_2) for both time periods were generated independently using $\mathcal{N}(0, 200)$. The intercept values at each time point were $\phi_{01} = \phi_{02} = 10$. Synthetic data sets based on Models 1, 2, and 3 were generated by considering $\phi_{12} \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$ and $\gamma_2 \in \{0.1, 0.25, 0.5, 1.0, 2.0\}$.

For each data generating scenario 1000 datasets were created and to each the 4 models were fit. For each model the value-added at each time period was estimated for each school using posterior means of the estimators provided in (2.23) for $\phi_{12} = \gamma_2 = 0$ (value-added for Model 0), for $\gamma_2 = 0$ (value-added for Model 1), for $\phi_{12} = 0$ (value-added for Model 2) and for $\phi_{12} \neq$ and $\gamma_2 \neq 0$ (value-added for Model 3). Producing credible intervals for value-added estimates is straightforward once draws from the corresponding posterior distributions for each estimator are collected. To compare the methods we average the mean squared error (MSE), coverage, and 95% credible interval widths across the 250 school's value-added estimates for each time point.

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

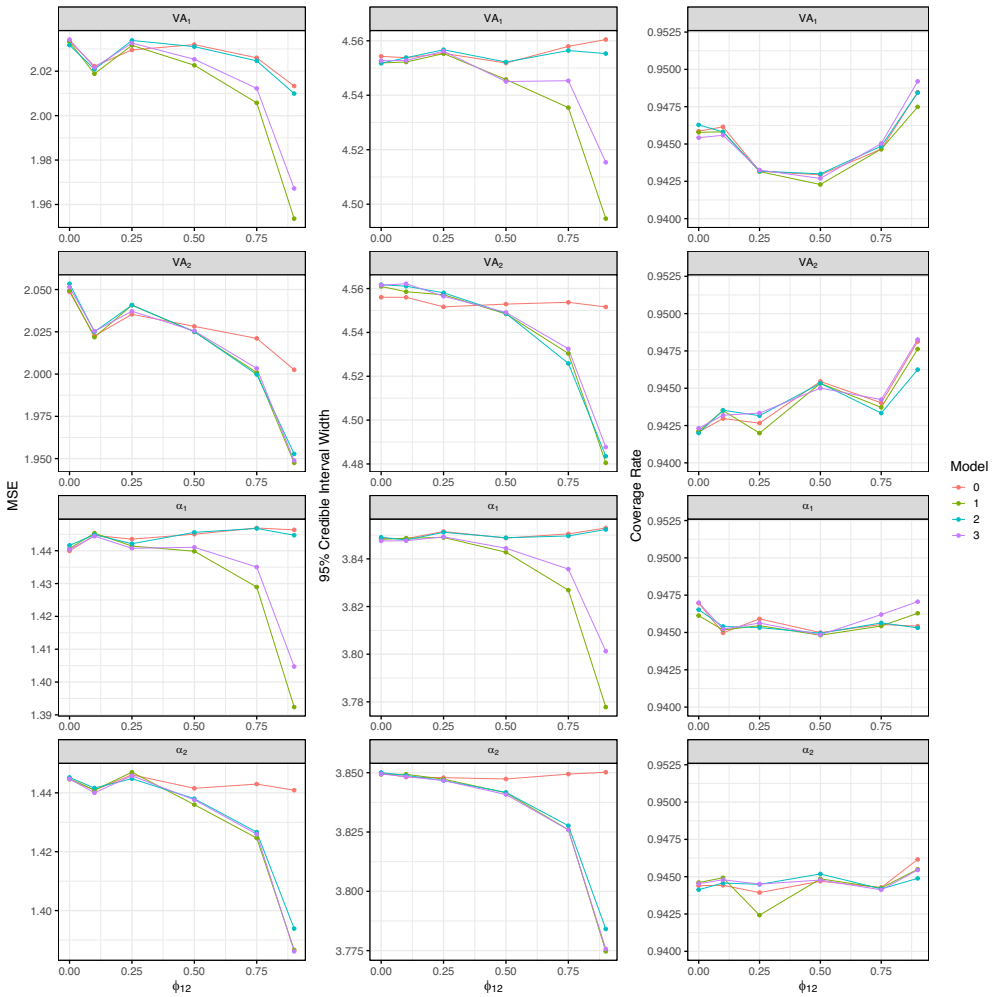

FIGURE 1.
Results from the simulation study when using Model 1 as a data generating mechanism. The MSE, interval widths, and coverage values are averages across 1000 generated sets and the 250 schools.

Coverage was estimated by calculating the proportion of the 95% credible intervals that contained the true value-added values (i.e., the values calculated using (2.23) for $\phi_{12} = \gamma_2 = 0$ (value-added for Model 0), for $\gamma_2 = 0$ (value-added for Model 1), for $\phi_{12} = 0$ (value-added for Model 2) and for $\phi_{12} \neq$ and $\gamma_2 \neq 0$ (value-added for Model 3)). Results associated with using Model 1 as a data generating mechanism are provided in Fig. 1, those for Model 2 are displayed in Fig. 2 and those for Model 3 in Fig. 4. In the Figs. 1 and 2, the first row corresponds to results associated with $VA_1(\boldsymbol{X}_1)$, the second row corresponds to $VA_2(\boldsymbol{X}_2)$, the third to α_1 and the fourth to α_2. The first column displays MSE values associated with the posterior mean estimator of VA_1, VA_2, α_1, and α_2 averaged over the $I = 250$ schools. The second column in both figures displays the 95% credible interval width averaged over the 250 schools, and the third column corresponds to the coverage property of the 95% credible intervals.

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

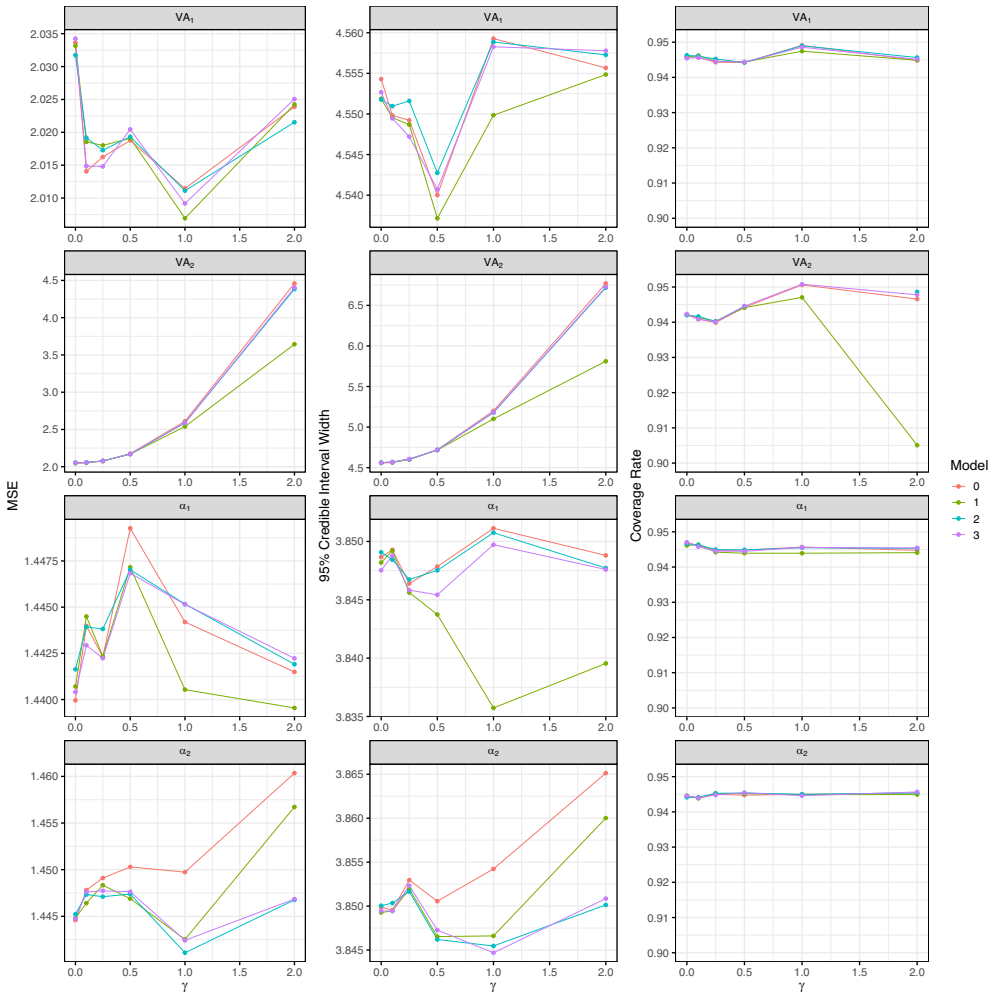

FIGURE 2.
Results from the simulation study when using Model 2 as a data generating mechanism. The MSE, interval widths, and coverage values are averages across 1000 generated sets and the 250 schools.

4.2. Conclusions of the Simulation Study

Focusing on Fig. 1, it appears that Model 1 performs best in estimating value-added for both time periods but much more so in the first time period. That is, the value-added estimates under Model 1 have the smallest MSE and the shortest credible interval widths while maintaining the same coverage rates as the other models. Even though Model 1 outperforms Model 3, Model 3's performance is vastly superior to that of model 0 and 2, which is to be expected. In addition, it appears that even if dependence between cohorts doesn't exist (i.e., $\phi_{12} = 0$), little is lost by using Model 1 (or 3) and in fact the benefits of using Model 1 (or 3) manifest themselves for relatively small values of ϕ_{12} (e.g., $\phi_{12} \approx 0.35$). Finally, it seems that incorporating any type of temporal dependence in a value-added model (even if misspecified) provides benefit as even Model 2 outperforms Model 0 at estimating VA.

The same trends seen in Fig. 1 also appear in Fig. 2 (which displays results when Model 2 was used to generate data) although not as stark. It does seem that Model 2 over all out performs

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

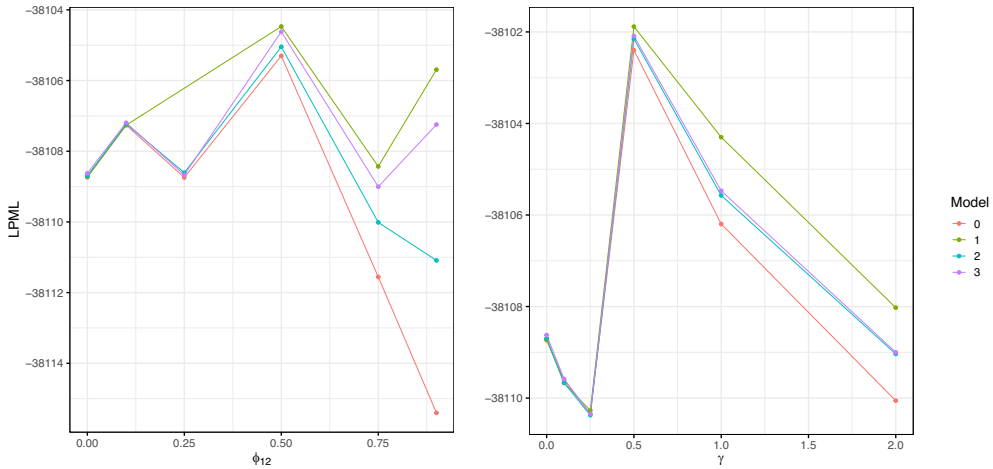

FIGURE 3.
Model fits measured using log pseudo marginal likelihood (LPML) for models 0, 1, and 2.

Model 0 at estimating value-added, but differences are more apparent in the estimation of the individual school random effects. Interestingly Model 1 performs better than Model 2 for VA_1, while for VA_2 the smaller MSE comes at a decrease in coverage for Model 1. Model 2 and 3 are similar in most scenariose. As before, even though the temporal dependence in Model 1 and 3 is misspecified in this scenario, there are clear benefits to include temporal dependence in some way as both Model's 1 and 3 outperforms Model 0 at estimating value-added particularly for the second cohort.

Next we provide Fig. 3 which displays the log pseudo marginal likelihood (LPML) values which can be used to evaluate model fit (see, Christensen et al. 2011, Chapter 4.9.2). Larger LPML values indicate a better fit. Here we see that even for weak temporal dependence incorporating the dependence in the value-added model results in better model fit. Interestingly, Model 1 tends to fit the data better than Model 2 even when Model 2 was used to generate data.

Lastly, Fig. 1 contains results when Model 3 is the true data generating mechanism. Here we only focus on performance for estimating VA_2 and trends are similar for other parameters and this allows us to be more concise in our description. For this figure, each column corresponds to the γ_2 value used to generate data while the tick marks on the x-axis correspond to the ϕ_{12} values. The first three rows correspond to the MSE, coverage and interval width associated with VA_2. The last row corresponds to the LPML model fit metric. Notice that as ϕ_{12} and γ_2 both increase (i.e., there is more temporal dependence), that Model 3 outperforms the other models in terms of LPML. Model 1 has the smallest MSE values for γ_2 but at the cost of a substantial decrease in coverage. Model 3's MSE and interval widths are the smaller than those of Model 0 and quite comparable to Model 1 (though slightly smaller).

The take home message from Figs. 1, 2, 3 and 4 is that incorporating temporal dependence in the model greatly improves value-added estimates when temporal dependence exists, and does so at a minimal cost when temporal dependence between cohorts is absent. Since the meaning of the value-added indicators in Models 1, 2, and 3 are very different in terms of school persistence (see the discussions in Sect. 2.6), model selection should be motivated by the intended use of the VA estimates.

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

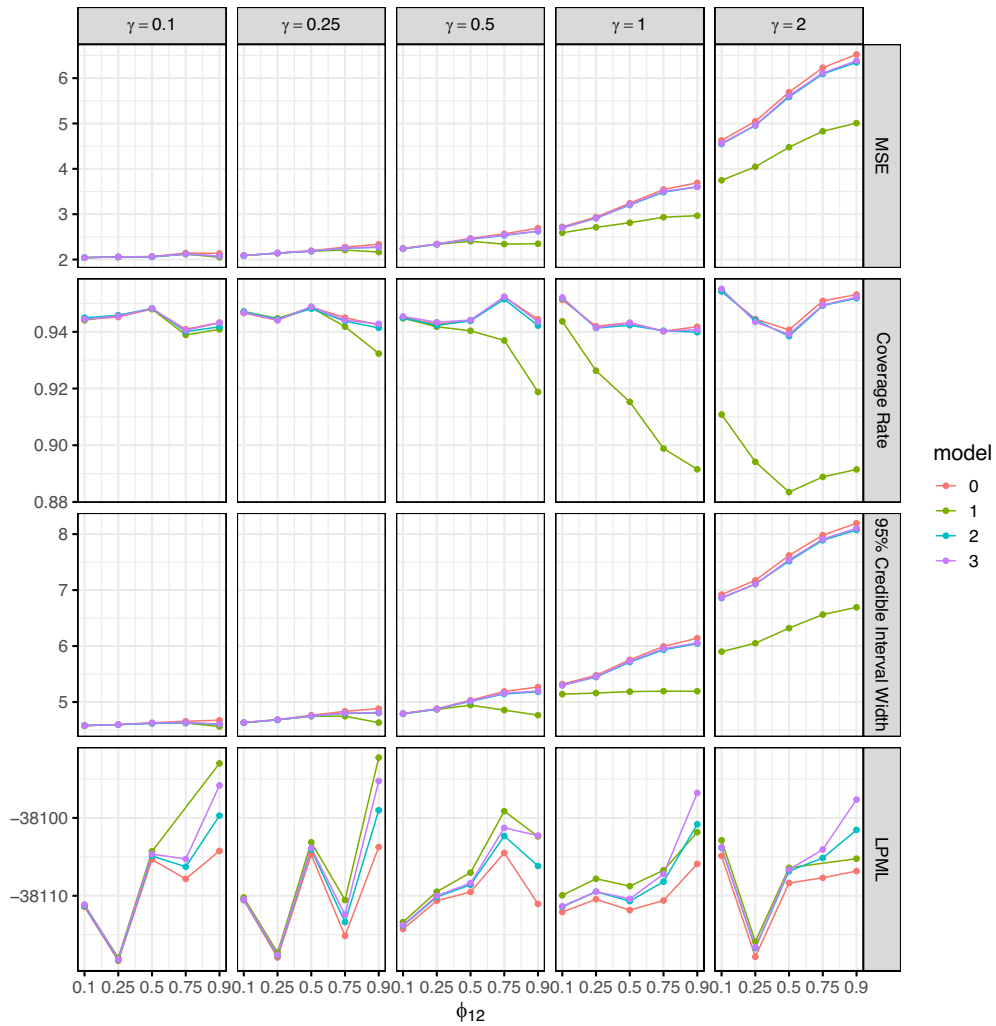

FIGURE 4.
Model fits using Model 3 as a data generating mechanism. The first three rows correspond to the MSE, interval width, and coverage of VA_{2i}. The last corresponds to the LPML. Columns indicate which γ_2 value was used to generate data and the x-axis tick marks indicate the same thing for ϕ_{12}.

5. Analysis of SIMCE Data

5.1. The SIMCE Test

Chile administers a yearly large-scale standardized test called SIMCE (Sistema de Medición de la Calidad de la Educación, Measurement System of Quality of Education). The main subjects of this test are Language and Mathematics. The SIMCE test was created at the end of the 1980's and coincided with the privatization of education which introduced issues such as competence among schools, private and public providers, vouchers to fund schools, universal school choice, for-profit schools, and co-payment (from 1993 to 2015). In this context, the SIMCE test was an instrument to aid parents in school choice decision-making, and to provide information necessary for schools to undertake data-based decision-making that would enhance school improvement

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

TABLE 1.
Model fit metrics for models 0, 1, and 2. For LPML, larger value indicates better fit, while for WAIC, smaller value indicates better fit.

Model	LPML	WAIC
0	− 1093957	2185061
1	− 1093837	2184869
2	− 1093844	2184913
3	− 1093853	2184900

efforts; for more details, see Meckes and Carrasco (2010), Manzi and Preiss (2013), Manzi et al. (2014) and Page et al. (2017).

5.2. *Data to be Used in the Value-Added Analysis*

We explore the extent of persistence in school effectiveness using two cohorts from schools to which the SIMCE test was administered. Here we only consider the results from the Mathematics part of the exam. The first cohort took the SIMCE exam as 4th graders in 2012 and then again as 6th graders in 2014 and the second cohort took the SIMCE exam as 4th graders during the 2014 school year and again during the 2016 school year as 6th graders. Thus X_{ij1} denotes the 4th grade SIMCE score of the ith student at the the jth school in 2012 and X_{ij1} the 4th grade SIMCE math score in 2014. Additionally, Y_{ij1} denotes the 6th grade math SIMCE score in 2014 and Y_{ij2} that of 2016. In total, the data set includes 2804 schools with the number of students per school ranging from 6 to 236 in cohort 1 and 6 to 258 in cohort 2. To these data, we fit Models 0, 1, 2, and 3 by retaining 1,000 MCMC iterates after discarding the first 10,000 as burnin and thinning by 200 (i.e., 210,000 total MCMC iterates were collected). Thinning was used, despite the potential loss of efficiency, to produce (essentially) independent samples from the posterior. The `tdVA` function in the `modernVA` R-package (Page, 2020) was used to fit all models and it took approximately 180 s for each.

5.3. *Results*

Before exploring school persistence in the four models, in Table 1 we provide the LPML and WAIC model fit metrics for each model. Again, larger LPML values indicate a better fit while smaller WAIC values indicate the same thing. It appears that both LPML and WAIC favor Models 1–3 over Model 0 with Model 1 fitting best based on both LPML and WAIC. The comparison between Model 0 and 1, and between Model 0 and 2 along with Model 1 and 3 and Models 2 and 3 seem adequate with respect to the nested structure of the models. Next in Table 2 we provide the posterior means and 95% credible intervals of β_1 and β_2 under each of the three models. Note that Models 2 and 3 produce very similar estimates of β_1 and β_2 while those for Model 0 and 1 differ slightly. Even so, it appears that differences (in terms of magnitude) between the four models in estimating β_1 and β_2 are minor.

5.3.1. School Persistence Under Model 1 As discussed in Sect. 2.6 persistence under Model 1 is based on ϕ_{12}. Since $|\phi_{12}| < 1$, persistence under this model corresponds to a uniform reduction in $VA_{1i}(\boldsymbol{X}_{1i})$ based on the magnitude of ϕ_{12}. For the SIMCE data the posterior mean of ϕ_{12} turned out to be 0.6 with a 95% credible interval of (0.57, 0.63). Based on the simulation study, the magnitude of the estimated value of ϕ_{12} is large enough to conclude that there is moderate to strong school effectiveness persistence among the schools of Chile based on these two cohorts.

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

TABLE 2.
Posterior summaries of β_1 and β_2 for each of the three models detailed in this paper.

	β_1		β_2	
Model	Mean	95% CI	Mean	95% CI
0	0.709	(0.705, 0.713)	0.722	(0.718, 0.726)
1	0.705	(0.701, 0.709)	0.717	(0.714, 0.722)
2	0.709	(0.705, 0.713)	0.713	(0.710, 0.718)
3	0.709	(0.705, 0.713)	0.714	(0.711, 0.718)

This results in a slight reduction in the credible interval widths for $VA_{1i}(\boldsymbol{X}_{1i})$ and $VA_{2i}(\boldsymbol{X}_{2i})$ compared to Model 0. In fact, the average credible interval width (across the 2804 schools) for $VA_{1i}(\boldsymbol{X}_{1i})$ under Model 0 is 19.1 compared to 18.6 under Model 1 and 18.9 for $VA_{2i}(\boldsymbol{X}_{2i})$ under Model 0 compared to 18.2 under Model 1.

As noted in (2.29), ϕ_{12} corresponds to the slope when regressing VA_2 onto VA_1 without an intercept. To verify this, for schools that have an average pre-test score for cohort 1 between 261 and 269 and an average pre-test score for cohort 2 between 253 and 261 (this resulted in 85 schools) we fit a least squares regression of the estimated VA_2 onto the estimated VA_1 without an intercept. The slope of this regression line turned out to be 0.603 which is very close to the posterior mean of ϕ_{12}. This emphasizes the fact that the VA_2 is in general smaller than VA_1 and, therefore, we expect to observe that schools taking on the new second cohort to not necessarily maintain the effectiveness achieved for the first cohort.

As a matter of fact, to further explore the school effectiveness persistence through reduction or shrinkage of $VA_{2i}(\boldsymbol{X}_{2i})$ based on ϕ_{12}, we provide Table 3. The table illustrates the "stability" of value-added estimates for cohort 1 and cohort 2 under Model 1 by presenting the percentage of schools (from the same 85 schools identified previously) according to the quartile in which they were located based on cohort 1 and cohort 2's value-added estimates. Note that since value-added is a metric that makes a comparison to a "reference" school, comparing estimated value-added percentiles across time is more reasonable than comparing the value-added estimates themselves. Thus, the values in Table 3 correspond to the percent of schools that belong to a particular combination of value-added quartiles between the two cohorts. For example, the entry at the upper left corner shows that 13% of the 85 schools had value-added estimates for cohort 1 and cohort 2 that belonged to the first quartile and the cell to its right shows that 5% of the 85 schools had a value-added estimate for cohort 1 that belonged to the second quartile, while that of cohort 2 belonged the first quartile. Other entries in the table can be interpreted similarly. Thus, higher values on the diagonal would indicate that the difference in value-added between the two cohorts is small. Notice that under Model 1 it appears that most of the differences between the two cohort's value-added can be found among the schools that are in the second and third quartiles. The majority of schools whose school effectiveness for cohort 1 is strong (or weak) also have strong (or weak) effectiveness for cohort 2. In fact, very few schools go from high value-added to low (or visa-versa) under Model 1.

To highlight further the bearing that ϕ_{12} has on school effectiveness persistence under Model 1 we provide as a contrast Table 4, which displays the same information as Table 3 but for Model 0. Contrasting the results in these two tables it is possible to see that differences between cohort 1 and cohort 2's value-added estimates under Model 0 are greater than that under Model 1. This can be seen first in terms of values on the diagonal, which are smaller in Table 4, 36% of schools, relative to the 42% in Table 3. Similarly, under Model 0 almost a quarter of the schools (24%) exhibited a change in the results of 2 or even 3 quartiles, as opposed the results under Model 1,

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

TABLE 3.
The percent of schools in each value-added quartile based on cohort 1 and cohort 2 for **Model 1**.

		Value-added quartile for cohort 1				
		[min, Q_1] (%)	(Q_1, Q_2] (%)	(Q_2, Q_3] (%)	(Q_3, max] (%)	Total (%)
Value-added quartile for cohort 2	[min, Q_1]	13	5	8	0	26
	(Q_1, Q_2]	7	11	6	1	25
	(Q_2, Q_3]	3	6	5	11	25
	(Q_3, max]	2	3	6	13	24
	Total	25	25	25	25	100

The schools included in this table are those that have average pre-test scores in cohort 1 between 261 and 269 and for cohort 2 between 253 and 261 (this resulted in 85 schools). Thus, the schools included in the table have similar student abilities in both cohorts.

TABLE 4.
The percent of schools in each value-added quartile based on cohort 1 and cohort 2 for **Model 0**.

		Value-added quartile for cohort 1				
		[min, Q_1] (%)	(Q_1, Q_2] (%)	(Q_2, Q_3] (%)	(Q_3, max] (%)	Total (%)
Value-added quartile for cohort 2	[min, Q_1]	10	4	8	4	26
	(Q_1, Q_2]	9	8	6	1	24
	(Q_2, Q_3]	2	8	6	8	24
	(Q_3, max]	4	5	5	12	26
	Total	25	25	25	25	100

The schools included in this table are those that have average pre-test scores in cohort 1 between 261 and 269 and for cohort 2 between 253 and 261 (this resulted in 85 schools). Thus, the schools included in the table have similar student abilities in both cohorts.

where only 17% of the schools presented such large differences. Though this contrast indicates that Model 1 produces more stable results than Model 0, it is important to keep in mind that the stability of the results in absolute terms is low overall, which can be summarized overall through the use of Cohen's Kappa, which corresponds to.24 in Table 3 and.38 in Table 4.

The improvement of stability in the results is worth highlighting as a relevant by-product of the use of Model 1 in this applied setting. Under Model 0, 4% of schools would receive results for cohort 1 indicating that they are in the highest quartile only to then receive a report for cohort 2 indicating they are in the lowest quartile, a scenario which is effectively eliminated under Model 1. The stability of results across time in value-added models is particularly relevant in applied settings, as presenting highly variable can appear to school officials as hard to interpret or simply as noise, a result that can undermine the credibility of the system and diminish the overall usefulness of the results.

Summarizing, though any value-added results will involve some degree of instability in the relative positions of schools across time, the expectation is that the variation is attributable to a real effect from schools related changes. However, the stability of value-added results will be inevitably affected by multiple sources of error, including the reliability of the test scores, the intraclass correlation of results at the school level, and the uncertainty associated with the value-added estimates themselves. In this case, the lack of stability between cohort1 and cohort 2's value-added estimates under Model 0 could be due either to failing to account for temporal dependence between the school effects, or to some school changes between periods 1 and 2: this ambiguity is avoided by using Model 1.

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

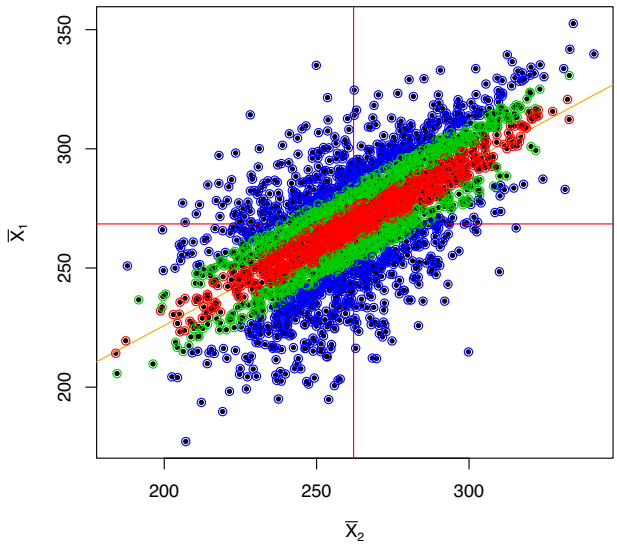

FIGURE 5.
Scatter plot of $\bar{X}_{2i}$ and $\bar{X}_{1i}$. The strong linear dependence between the pre-test of the two cohorts indicates that prior information about the two cohorts is similar.

5.3.2. School Persistence Under Model 2 We now explore school persistence under Model 2. As mentioned in Sect. 2.6 school persistence based on Model 2 focuses on to what extent a school is able to do "new things" with a new cohort. If the information about the cohort is similar (i.e. pre-test of both cohorts are similar), then cohorts have similar abilities and teaching strategies devised for cohort 1 should be useful for cohort 2. The extent to which a school is able to do "new things" is reflected in γ_2. The posterior mean of γ_2 turned out to be 0.35 with 95% credible interval (0.34, 0.37). Thus based on Model 2, persistence in school effectiveness depends on the way in which the corresponding school effect α_{2i} is corrected by $n_{2i}^{-1}\,\beta_1\gamma_2\,E(\bar{X}_{1i} \mid \bar{X}_{2i})$, where $\widehat{\beta}_1 \times \widehat{\gamma}_2 \approx 0.25$.

To quantify the amount of prior information provided to schools for each cohort by way of a pre-test we provide Fig. 5. The figure contains a scatterplot of $\bar{X}_{2i}$ and $\bar{X}_{1i}$. The correlation between the pairs $(\bar{X}_{1i}, \bar{X}_{2i})$ turns out to be 0.68 indicating quite strong correlation between pre-test of cohort 1 and cohort 2. To further explore persistence under Model 2, we group schools based on the similarity between $\bar{X}_{1i}$ and $\bar{X}_{2i}$. This is done by forming a group of schools whose residual value (r_i) from the regression line found in Fig. 5 is $|r_i| < 1$ (red group in Fig. 5), $1 \leq |r_i| < 2$ (green group in Fig. 5), and $|r_i| > 2$ (blue group in Fig. 5). For each group, tables similar to those in Sect. 5.3.1 are provided based on schools that had significantly different value-added estimates between the two cohorts (significance was establish if 95% credible intervals for VA_1 and VA_2 failed to intersect). First note that there are only a few schools that fall within the middle quartiles. This is a result of only considering schools with significantly different value-added estimates between cohorts. Notice further that as the information provided schools becomes less correlated (i.e., points farther from the regression line), there is a higher percentage of schools that improve their quartile group: 40% for red schools, 42% for green schools and 49% for blue schools. Moreover, for red schools 14% do not change of quartile group, for green schools 13% do not change, and for blue schools 13% do not change. This indicates that the less similar the pre-test cohorts are, the better the persistence of the schools. This suggests that the schools in the sample are able to improve their results by doing "new things" for the new cohort (Tables 5, 6, and 7).

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

TABLE 5.
The percent of schools in each value-added quartile based on cohort 1 and cohort 2 based on Model 2's value-added estimates.

		Value-added for cohort 1				
		$[\min, Q_1]$ (%)	$(Q_1, Q_2]$ (%)	$(Q_2, Q_3]$ (%)	$(Q_3, \max]$ (%)	Total (%)
Value-added	$[\min, Q_1]$	2	5	9	6	22
for cohort 2	$(Q_1, Q_2]$	6	0	2	9	17
	$(Q_2, Q_3]$	12	2	0	14	28
	$(Q_3, \max]$	5	10	6	12	33
	Total	25	17	17	41	100

Schools included are those that correspond to the red points in Fig. 5 and that had significantly different value-added estimates in cohort 1 relatively to cohort 2 based on the criteria that the posterior 95% credible intervals didn't intersect. The total number of schools is 210.

TABLE 6.
The percent of schools in each value-added quartile based on cohort 1 and cohort 2 based on model 2's value-added estimates.

		Value-added for cohort 1				
		$[\min, Q_1]$ (%)	$(Q_1, Q_2]$ (%)	$(Q_2, Q_3]$ (%)	$(Q_3, \max]$ (%)	Total (%)
Value-added	$[\min, Q_1]$	3	3	7	7	20
for cohort 2	$(Q_1, Q_2]$	7	0	2	13	22
	$(Q_2, Q_3]$	12	0	0	13	25
	$(Q_3, \max]$	4	10	9	10	33
	Total	26	13	18	43	100

Schools included correspond to the green points in Fig. 5 and that had significantly different value-added estimates in cohort 1 relatively to cohort 2 based on the criteria that the posterior 95% credible intervals didn't intersect. The total number of schools is 208.

TABLE 7.
The percent of schools in each value-added quartile based on cohort 1 and cohort 2 for model 2 based on model 2's value-added estimates.

		Value-added for cohort 1				
		$[\min, Q_1]$ (%)	$(Q_1, Q_2]$ (%)	$(Q_2, Q_3]$ (%)	$(Q_3, \max]$ (%)	Total (%)
Value-added	$[\min, Q_1]$	3	7	9	7	26
for cohort 2	$(Q_1, Q_2]$	9	0	3	7	19
	$(Q_2, Q_3]$	14	2	0	11	27
	$(Q_3, \max]$	7	10	7	4	28
	Total	33	19	19	29	100

Schools included correspond to the blue points in Fig. 5 and that had significantly different value-added estimates in cohort 1 relatively to cohort 2 based on the criteria that the posterior 95 credible intervals didn't intersect. The total number of schools is 229.

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

6. Conclusions

Value-added models are a plausible tool for monitoring the effectiveness of a school across time. From a policy perspective, it is important to identify schools whose effectiveness has changed dramatically (either improvement or deterioration) and those that maintain their effectiveness. However, the modeling challenge is being able to disentangle the instability of value-added indicators due to internal and/or external changes affecting the effectiveness of schools from the instability due to the specification of value-added models. To meet this challenge, we formulate time dependent value-added models, which are basically characterized by specifying the school effect related to cohort t as a function of the past performance of the school. By doing so, we intend to eliminate one source of instability that is contained in a value-added model where the school effects are mutually independent across time.

More specifically, we have proposed a value-added model that incorporates temporal dependence from two different perspectives, namely a dependence in the school random effects (Model 1) and a "shock" based on the post-test performance from the previous cohort (Model 2). The identification analysis indicated that both models are nested into Model 3, which in turn incorporates both temporal dependencies. The value-added indicators induced by Models 1 and 2 have different statistical interpretations and, consequently, are useful for different policy purposes. As a matter of fact, Model 1 assumes that the school effects are correlated over time as in ARIMA-type models, whereas Model 2 assumes that the current school effect is influenced by the post-tests from previous cohorts as a kind of "information shock". An empirical analysis of the persistence of school effectiveness under Model 1 means analyzing to what extent a school that takes on new cohorts maintains the effectiveness achieved when accommodating the first cohort. Such an effectiveness may be viewed as a base line: the farther a cohort is from the first, the harder it is to maintain the effectiveness achieved with that first cohort. An empirical analysis of the persistence of school effectiveness under Model 2, means analyzing to what extent a school that deals with a new cohort is capable of doing "new things" with it. The focus of the empirical analysis is on the future performance after taking into account the "shocks" of information. In spite of that, Model 3 incorporates both time dependencies, and allows us to see that the persistence of school effectiveness corresponds to an additive combination of both the school value-added for cohort 1 and the information coming from cohort 1: the first additive component is related to the ARIMA-type model, whereas the second additive component is related to the "information shock" model. It is important to point out that if the parameter of shock γ_2 is large enough (in absolute value), then the persistence of school effectiveness under Model 2 and 3 are quite similar.

In order to show that this modeling strategy leads to control one source of instability, we have also discussed the effects of Models 1, 2 and 3 in contrast with the traditional value-added Model 0 in terms of the stability of the school results in different cohorts. The results of the applied example in this study show that the use of models that include temporal dependence improves the consistency of the school results when contrasted with the single cohort model. This result is potentially relevant for the use of value-added models in applied settings, as high instability of the estimates can be hard to interpret, or worse, can be perceived as random by schools officials, potentially affecting the credibility of the overall value-added system. Thus, it seems that a reasonable approach to employing our method is to first fit Model 3 (i.e., the Full Model) and then carry out hypothesis tests (our use Bayes Factors) to determine if would be appropriate to use either Model 1 or 2 (i.e., assume that $\phi_{12} = 0$ or $\gamma_2 = 0$). What we provide in this paper is an all encompassing modeling strategy regardless of which model a particular data set favors. The interpretation of value-added and persistence in school effectiveness for all models has been studied and all are based on theoretically sound arguments.

A related, and relevant, characteristic of the proposed models is the "shrinkage" effect associated with the use of random effects. We consider this as another valuable characteristic of the

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

proposed models, as the effect will tend to moderate results where there are fewer cases (i.e., less evidence) to draw inferences regarding the school value added effect. This more conservative estimate will not necessarily disadvantage schools that may have fewer students in a particular cohort, as in systems that have incentives to both punish and reward with lower VA estimates (and particularly in those that just include incentives to punish schools), the conservative estimates will tend to shield these schools from suffering penalties due to extreme negative results. Though this effect will also preclude them from potentially being rewarded due to extreme positive results, we contend that this is a reasonable trade-off; particularly when these effects will play out in the face of less empirical evidence than the one available for other schools. This moderating effect is particularly relevant as part of the larger issue of inclusion or removal of schools with few observations as part of a national system, particularly given that the number of cases with common responses can vary year to year. Though one approach is to remove these schools from the analysis, we consider that shrinkage effects constitute a feature that opens the possibility of maintaining them as part of a larger system even though in some years they would have been excluded based on an arbitrary cut point for sample size was used.

From a modeling perspective, we emphasize that Model 3 (along with its nested models) includes parameters that can be considered as characterizations of an educational system, namely $\phi_{12} + \gamma_2$, γ_2 and ϕ_{12}: the first one characterizes, for all schools, the sign of the correlation between school effect for cohort 1 and school effect for cohort 2; this correlation is essentially characterized by ϕ_{12}. The first and second parameters characterize, for all schools, the sign of the between-cohort correlation. As next steps in developing this modeling approach we propose specifying parameters like ϕ_{12} and γ_2 at sub-groups levels to identify population heterogeneity of these effects.

Last, but not least, we highlight the fact that the proposed models can be fit using the `R` package `modernVA` (Page, 2020).

Acknowledgments

The second, third and fourth authors were partially supported by the Millennium Nucleus on Intergenerational Mobility MOVI. Part of this research was developed in the context of a scientific consultancy for the National Agency for Quality Education of the Government of Chile. The perspective developed in this paper does not necessarily represent those of the Agency. The authors thank two anonymous referees for their comments and questions, which helped improve the paper.

Data availibility The data used in this paper will be included in the supplementary material.

Declarations

Conflict of interest The authors report no Conflict of interest.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

A. Appendix

A.1. Value-added in the Full Model

Let us provide the derivations for Model 2 only: for the first cohort, we have

$$\begin{aligned} VA_{1i}(\boldsymbol{X}_{1i}) &\doteq \frac{1}{n_{1i}} \sum_{j=1}^{n_{1i}} \left[E(Y_{1ij} \mid \boldsymbol{X}_{1ij}, \alpha_{1i}) - E(Y_{1ij} \mid \boldsymbol{X}_{1ij}) \right] \\ &= \frac{1}{n_{1i}} \sum_{j=1}^{n_{1i}} \left\{ \boldsymbol{X}'_{1ij}\boldsymbol{\beta}_1 + \alpha_{1i} - E[E(Y_{1ij} \mid \boldsymbol{X}_{2ij}, \boldsymbol{X}_{1ij}) \mid \boldsymbol{X}_{1ij}] \right\} \\ &= \frac{1}{n_{1i}} \sum_{j=1}^{n_{1i}} \left\{ \boldsymbol{X}'_{1ij}\boldsymbol{\beta}_1 + \alpha_{1i} - E[\phi_{01} + \boldsymbol{X}'_{1ij}\boldsymbol{\beta}_1 \mid \boldsymbol{X}_{1ij}] \right\} \\ &= \alpha_{1i} - \phi_{01}. \end{aligned}$$

Similarly, for cohort 2,

$$\begin{aligned} VA_{2i}(\boldsymbol{X}_{2i}) &\doteq \frac{1}{n_{2i}} \sum_{j=1}^{n_{2i}} \left[E(Y_{2ij} \mid \boldsymbol{X}_{2ij}, \alpha_{2i}) - E(Y_{2ij} \mid \boldsymbol{X}_{2ij}) \right] \\ &= \frac{1}{n_{2i}} \sum_{j=1}^{n_{2i}} \left\{ \boldsymbol{X}'_{2ij}\boldsymbol{\beta}_2 + \alpha_{2i} - E[E(Y_{2ij} \mid \boldsymbol{X}_{2ij}, \boldsymbol{X}_{1ij}) \mid \boldsymbol{X}_{2ij}] \right\} \\ &= \frac{1}{n_{2i}} \sum_{j=1}^{n_{2i}} \left\{ \boldsymbol{X}'_{2ij}\boldsymbol{\beta}_2 + \alpha_{2i} - E[\phi_{02} + \phi_{01}(\phi_{12} + \gamma_2) + \boldsymbol{X}'_{2ij}\boldsymbol{\beta}_2 + \gamma_2 \overline{\boldsymbol{X}}_{1i}\boldsymbol{\beta}_1 \mid \boldsymbol{X}_{2ij}] \right\} \\ &= \alpha_{2i} - [\phi_{02} + \phi_{01}(\phi_{12} + \gamma_2)] - \frac{\gamma_2}{n_{2i}} \boldsymbol{\beta}'_1 \sum_{j=1}^{n_{2i}} E(\overline{\boldsymbol{X}}'_{1i} \mid \boldsymbol{X}_{2ij}). \end{aligned}$$

A.2. Correction factor for the school value-added under the Full Model

As it was discussed in Sect. 2.5, the school value-added for cohort 2 under the Full Model is equal to the centered school effect $\alpha_{2i} - (\phi_{02} + \phi_{01}\gamma_2)$ plus a correction factor given by

$$\frac{\gamma_2}{n_{2i}} \boldsymbol{\beta}'_1 \sum_{j=1}^{n_{2i}} E(\overline{\boldsymbol{X}}'_{1i} \mid \boldsymbol{X}_{2ij}), \quad \text{where } (\gamma_2, \boldsymbol{\beta}_1) \in \mathbb{R} \times \mathbb{R}^{p_1}.$$

Then

$$E(\overline{\boldsymbol{X}}'_{1i} \mid \boldsymbol{X}_{2ij}) = \begin{pmatrix} E(\overline{X}_{1i,1} \mid \boldsymbol{X}_{2ij}) \\ E(\overline{X}_{1i,2} \mid \boldsymbol{X}_{2ij}) \\ \vdots \\ E(\overline{X}_{1i,p_1} \mid \boldsymbol{X}_{2ij}) \end{pmatrix}$$

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

Now, for the k-th covariate, let us assume that $E(\overline{X}_{1i,k} \mid \boldsymbol{X}_{2ij}) = b_{k0} + \boldsymbol{b}_k' \boldsymbol{X}_{2ij}$, where $\boldsymbol{b}_k$ is a $p_2 \times 1$ vector and $b_{k0} \in \mathbb{R}$. One possibility is to consider $b_k = d_k\, e_k$ where $d_k \in \mathbb{R}$ and e_k is the k-th vector of the canonical basis of $\mathbb{R}^{p_2}$. Therefore,

$$E(\overline{\boldsymbol{X}}_{1i}' \mid \boldsymbol{X}_{2ij}) = \begin{pmatrix} b_{10} & b_{11} & b_{12} & \cdots & b_{1p_2} \\ b_{20} & b_{21} & b_{22} & \cdots & b_{2p_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{p_10} & b_{p_11} & b_{p_12} & \cdots & b_{p_1p_2} \end{pmatrix} \begin{pmatrix} 1 \\ X_{2ij,1} \\ X_{2ij,2} \\ \vdots \\ X_{2ij,p_2} \end{pmatrix} \doteq \boldsymbol{B}\boldsymbol{Z}_{2ij},$$

where $\boldsymbol{B}$ is a $p_1 \times (p_2 + 1)$ matrix and $\boldsymbol{Z}_{2ij}$ is a $(p_2 + 1) \times 1$ vector.
We stack the conditional expectations by considering the n_{2i} students:

$$\begin{pmatrix} E(\overline{\boldsymbol{X}}_{1i}' \mid \boldsymbol{X}_{2i1}) \\ E(\overline{\boldsymbol{X}}_{1i}' \mid \boldsymbol{X}_{2i2}) \\ \vdots \\ E(\overline{\boldsymbol{X}}_{1i}' \mid \boldsymbol{X}_{2in_{2i}}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{B}\boldsymbol{Z}_{2i1} \\ \boldsymbol{B}\boldsymbol{Z}_{2i2} \\ \vdots \\ \boldsymbol{B}\boldsymbol{Z}_{2in_{2i}} \end{pmatrix} = \left(\boldsymbol{I}_{n_{2i}} \otimes \boldsymbol{B}\right) \begin{pmatrix} \boldsymbol{Z}_{2i1} & & & \\ & \boldsymbol{Z}_{2i2} & & \\ & & \ddots & \\ & & & \boldsymbol{Z}_{2in_{2i}} \end{pmatrix}$$

Then

$$\begin{aligned} \sum_{j=1}^{n_{2i}} E(\overline{\boldsymbol{X}}_{1i}' \mid \boldsymbol{X}_{2ij}) &= (\boldsymbol{\iota}_{n_{2i}}' \otimes \boldsymbol{I}_{p_1}) \begin{pmatrix} E(\overline{\boldsymbol{X}}_{1i}' \mid \boldsymbol{X}_{2i1}) \\ E(\overline{\boldsymbol{X}}_{1i}' \mid \boldsymbol{X}_{2i2}) \\ \vdots \\ E(\overline{\boldsymbol{X}}_{1i}' \mid \boldsymbol{X}_{2in_{2i}}) \end{pmatrix} \\ &= (\boldsymbol{\iota}_{n_{2i}}' \otimes \boldsymbol{I}_{p_1}) \left(\boldsymbol{I}_{n_{2i}} \otimes \boldsymbol{B}\right) \begin{pmatrix} \boldsymbol{Z}_{2i1} & & & \\ & \boldsymbol{Z}_{2i2} & & \\ & & \ddots & \\ & & & \boldsymbol{Z}_{2in_{2i}} \end{pmatrix} \\ &= (\boldsymbol{\iota}_{n_{2i}}' \otimes \boldsymbol{B}) \begin{pmatrix} \boldsymbol{Z}_{2i1} & & & \\ & \boldsymbol{Z}_{2i2} & & \\ & & \ddots & \\ & & & \boldsymbol{Z}_{2in_{2i}} \end{pmatrix} \\ &= \sum_{j=1}^{n_{2i}} \boldsymbol{B}\boldsymbol{Z}_{2ij} = \boldsymbol{B} \sum_{j=1}^{n_{2i}} \boldsymbol{Z}_{2ij} \end{aligned}$$

Therefore,

$$\frac{1}{n_{2i}} \sum_{j=1}^{n_{2i}} E(\overline{\boldsymbol{X}}_{1i}' \mid \boldsymbol{X}_{2ij}) = \boldsymbol{B}\, \frac{1}{n_{2i}} \sum_{j=1}^{n_{2i}} \boldsymbol{Z}_{2ij} = \boldsymbol{B}\, \overline{\boldsymbol{Z}}_{2i} = \boldsymbol{B} \begin{pmatrix} 1 \\ \overline{X}_{2i,1} \\ \vdots \\ \overline{X}_{2i,p_2} \end{pmatrix}.$$

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

Now, a linear regression is based on the decomposition

$$\overline{\boldsymbol{X}}'_{1i} = E(\overline{\boldsymbol{X}}'_{1i} \mid \boldsymbol{X}_{2ij}) + \left(\overline{\boldsymbol{X}}'_{1i} - E(\overline{\boldsymbol{X}}'_{1i} \mid \boldsymbol{X}_{2ij})\right) = E(\overline{\boldsymbol{X}}'_{1i} \mid \boldsymbol{X}_{2ij}) + u_i.$$

Then, after averaging on j, we have

$$\underbrace{\overline{\boldsymbol{X}}'_{1i}}_{p_1 \times 1} = \underbrace{\boldsymbol{B}}_{p_1 \times (p_2+1)} \underbrace{\overline{\boldsymbol{Z}}_{2i}}_{(p_2+1)\times 1} + u_i.$$

Therefore, assuming that there are I schools, we have

$$\begin{pmatrix} \overline{\boldsymbol{X}}'_{11} \\ \vdots \\ \overline{\boldsymbol{X}}'_{1I} \end{pmatrix} = (\boldsymbol{I}_I \otimes \boldsymbol{B}) \begin{pmatrix} \overline{\boldsymbol{Z}}'_{21} \\ \vdots \\ \overline{\boldsymbol{Z}}'_{2I} \end{pmatrix} + \boldsymbol{u}.$$

References

Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society: Series A (General), 149*(1), 1–26.

Amrein-Beardsley, A., & Holloway, J. (2019). Value-added models for teacher evaluation and accountability: Commonsense assumptions. *Educational Policy, 33*(3), 516–542.

Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics, 29*(1), 37–65.

Bellei, C., Vanni, X., Valenzuela, J. P., & Contreras, D. (2016). School improvement trajectories: An empirical typology. *School Effectiveness and School Improvement, 27*, 275–292.

Bianconcini, S., & Cagnone, S. (2012). A general multivariate latent growth model with applications to student achievement. *Journal of Educational and Behavioral Statistics, 37*(2), 339–364.

Billingsley, P. (1968). *Convergence of probability measures*. Wiley.

Briggs, D. C., & Weeks, J. P. (2011). The persistence of school-level value-added. *Journal of Educational and Behavioral Statistics, 36*, 616–637.

Christensen, R., Johnson, W., Branscum, A., & Hanson, T. (2011). Bayesian ideas and data analysis: An introduction for scientists and statisticians. Taylor & Francis. https://books.google.com/books?id=qPERhCbePNcC

Clarke, P., Crawford, C., Steele, F., & Vignoles, A. (2015). Revisiting fixed-and random-effects models: Some considerations for policy-relevant education research. *Education Economics, 23*(3), 259–277.

Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M. J. (2014). The sensitivity of value-added estimates to specification adjustments: Evidence from school- and teacher-level models in missouri. *Statistics and Public Policy, 1*(1), 19–27. https://doi.org/10.1080/2330443X.2013.856152

Engle, R. E., Hendry, D. F., & Richard, J. F. (1983). Exogeneity. *Econometrica, 51*, 277–304.

EPI Briefing Paper. (2010). Problems with the use of student test scores to evaluate teachers. Economic Policy Institute.

Fariña, P., González, J., & San Martín, E. (2019). The use of an identifiability-based strategy for the interpretation of parameters in the 1PL-G and Rasch models. *Psychometrika, 84*, 511–528.

Fisher, R. A. (1973). *Statistical methods for research workers*. Hafner Publishhing.

Fitzmaurice, G., Laird, N., & Ware, J. (2004). *Applied longitudinal analysis*. Wiley.

Florens, J.-P., Marimoutou, V., & Péguin-Feissolle, A. (2007). *Econometric modeling and inference*. Cambridge University Press.

Gray, J., Goldstein, H., & Jesson, D. (1996). Changes and improvements in schools' effectiveness: Trends over five years. *Research Papers in Education, 11*, 35–51.

Gray, J., Goldstein, H., & Thomas, S. (2001). Predicting the future: The role of past performance in determining trends in institutional effectiveness at A level. *British Educational Research Journal, 27*, 391–405.

Gray, J., Hopkins, D., Reynolds, D., Wilcox, B., Farrell, S., & Jesson, D. (1999). *Improving schools: Performance and potential*. Open University Press.

Guldemond, H., & Bosker, R. J. (2009). School effects on students' progress: A dynamic perspective. *School Effectiveness and School Improvement, 20*(2), 255–268.

Hanushek, E. A. (2020). Education production functions. In S. Bradley & C. Green (Eds.), *The economics of education* (pp. 161–170). Elsevier.

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

Hsiao, C. (2014). *Analysis of panel data*. Cambridge University Press.
Kinsler, J. (2012). Beyond levels and growth: Estimating teacher value-added and its persistence. *Journal of Human Resources, 47*(3), 722–753.
Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review, 47*, 180–195.
Kolmogorov, A. N. (1950). *Foundations of the theory of probability*. Chelsea Publishing Company.
Kyriakides, L., Georgiou, M. P., Creemers, B. P., Panayiotou, A., & Reynolds, D. (2018). The impact of national educational policies on student achievement: A European study. *School Effectiveness and School Improvement, 29*(2), 171–203.
Leckie, G. (2018). Avoiding bias when estimating the consistency and stability of value-added school effects. *Journal of Educational and Behavioral Statistics, 43*, 440–468.
Lindley, D. V. (1983). Bayesian statistics: A review. In: *CBMS-NSf regional conference series in applied mathematics*, Philadelphia.
Liu, J., & Loeb, S. (2019). Engaging teachers: Measuring the impact of teachers on student attendance in secondary school. *Journal of Human Resources*, pp. 1216–8430R3.
Lockwood, J., McCaffrey, D. F., Mariano, L. T., & Setodji, C. (2007). Bayesian methods for scalable multivariate value-added assessment. *Journal of Educational and Behavioral Statistics, 32*, 125–150.
Longford, N. T. (2012). A revision of school effectiveness analysis. *Journal of Educational and Behavioral Statistics, 37*(1), 157–179.
Lord, F. M., & Novick, M. (1968). *Statistical theories of mental test scores*. Addison Wesley.
Manzi, J., & Preiss, D. (2013). Educational Assessment and Educational Achievement in South America. In J. Hattie & E. M. Anderman (Eds.), International guide to student achievement (p. chapter 9). Taylor and Friends.
Manzi, J., San Martín, E., & Van Bellegem, S. (2014). School system evaluation by value added analysis under endogeneity. *Psychometrika, 79*(1), 130–153.
McCaffey, D. F., Lockwood, J., Koretz, D. M., Louis, T. A., & Hamilton, L. S. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics, 29*, 67–101.
Meckes, L., & Carrasco, R. (2010). Two decades of simce: An overview of the national assessment system in Chile. *Assessment in Education: Principles, Policy and Practice, 17*, 233–248.
Mouchart, M., & Oulhaj, A. (2006). The role of exogenous randomness in the identification of conditional models. *Metron - International Journal of Statistics*, LXIV:253–271.
Neveu, J. (1972). *Martingales 'a temps discret*. Paris: Masson et CIE.
Page, G. L. (2020). modernva: An implementation of two modern education-based value-added models [Computer software manual]. https://CRAN.R-project.org/package=modernVA (R-package version 0.1.1)
Page, G. L., San Martín, E., Orellana, J., & González, J. (2017). Exploring complete school effectiveness via quantile value-added. *Journal of the Royal Statistical Society Series A, 180*, 315–340.
Papay, J. P. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal, 48*(1), 163–193.
Potthoff, R. F., & Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika, 51*(3–4), 313–326.
Reynolds, D., Sammons, P., Fraine, B. D., Damme, J. V., Townsend, T., Teddlie, C., & Stringfield, S. (2014). Educational effectiveness research (EER): A state-of-the-art review. *School Effectiveness and School Improvement, 25*, 197–230.
Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *The Quarterly Journal of Economics, 125*(1), 175–214.
Sanders, W. L., & Horn, S. P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in education, 8*(3), 299–311.
San Martín, E., González, J., & Tuerlinckx, F. (2015). On the unidentifiability of the fixed-effects 3PL model. *Psychometrika, 80*, 450–467.
San Martín, E., Jara, A., Rolin, J.-M., & Mouchart, M. (2011). On the Bayesian nonparametric generalization of IRT-type models. *Psychometrika, 76*(3), 385–409.
San Martín, E., Rolin, J.-M., & Castro, L. M. (2013). Identification of the 1PL model with guessing parameter: Parametric and semi-parametric results. *Psychometrika, 78*, 341–379.
Sass, T. R., Hannaway, J., Xu, Z., Figlio, D. N., & Feng, L. (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of urban Economics, 72*(2–3), 104–122.
Scherrer, J. (2011). Measuring teaching using value-added modeling: The imperfect panacea. *NASSP Bulletin, 95*(2), 122–140.
Strenio, J. F., Weisberg, H. I., & Bryk, A. S. (1983). Empirical bayes estimation of individual growth-curve parameters and their relationship to covariates. *Biometrics*, pp. 71–86.
Tekwe, C. D., Carter, R. L., Ma, C.-X., Algina, J., Lucas, M. E., Roth, J., & Resnick, M. B. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics, 29*(1), 11–36.
Thomas, S., Peng, W. J., & Gray, J. (2007). Modelling patterns of improvement over time: Value added trends in English secondary school performance across ten cohorts. *Oxford Review of Education, 33*, 261–295.
Tymms, P., Merrell, C., & Bailey, K. (2018). The long-term impact of effective teaching. *School Effectiveness and School Improvement, 29*(2), 242-261.

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.

Vanwynsberghe, G., Vanlaar, G., Van Damme, J., & De Fraine, B. (2017). Long-term effects of primary schools on educational positions of students 2 and 4 years after the start of secondary education. *School Effectiveness and School Improvement, 28*(2), 167–190.

Zimmerman, D. W. (1975). Probability spaces, hilbert spaces, and the axioms of test theory. *Psychometrika, 40*(3), 395–412.

Manuscript Received: 26 APR 2023
Accepted: 10 MAY 2024
Published Online Date: 22 JUN 2024

Downloaded from https://www.cambridge.org/core. 09 Jan 2025 at 05:50:16, subject to the Cambridge Core terms of use.