Cohort Varying, Temporally Dynamic, Value-Added Models

Supplementary Material

A Theory of Bayesian Identification

It is known that a Bayesian model is a combination of both a likelihood function and a prior probability distribution. However, in order to grasp the relevance of identifiability in a Bayesian framework, it seems useful to recall the general construction of a Bayesian model.

A.1 Construction of a Bayesian model

The starting point for building a Bayesian model is the *statistical model*. A statistical model involves three components:

1. The space of observations, commonly called *sample space*, which is represented by a measurable space $(S, \mathcal{S})$, where $\mathcal{S}$ is the σ-field of subsets of S: $\mathcal{S}$ contains the events of interest.
2. A family of probability distributions P^a defined on the sample space $(S, \mathcal{S})$, where a is a parameter. These probabilities are called *sampling probabilities* in order to emphasize that a statistical model specifies *the observations*.
3. The *parameter space* A.

Thus, a statistical model is a family of sampling distributions indexed by a parameter, and it is compactly written as

$$\mathcal{E}_s = \{(S, \mathcal{S}), P^a : a \in A\};$$

the subscript s emphasizes the fact that a statistical model deals with the sampling process generating the observations; for details, see Fisher (1922), Cox and Hinkley (1974), Basu (1975), Raoult (1975), Barra (1981), Florens, Mouchart, and Rolin (1990) and McCullagh (2002). In the context of the statistical model $\mathcal{E}_s$, the parameter a is said to be s-identified if the mapping $a \longmapsto P^a$ is injective

Remark A.1 The statistical model $\mathcal{E}_c$ not only includes the so-called *parametric models*, but also *non-parametric* and *semi-parametric* models. As a matter of fact, if the parameter space A is (a subset of a) vector space of finite dimension, then the statistical model is called *parametric statistical model*; if A

is (a subset of a) vector space of infinite dimension, then the statistical model is called *non-parametric statistical model*; finally, if A is (a subset of a) cartesian product of a finite dimensional vector space and an infinite dimensional vector space, then the statistical model is called *semi-parametric statistical model*.

Once the statistical model $\mathcal{E}_c$ is given, it is necessary to endow the sampling probabilities with a measurability structure at the parameter space level. More specifically, for each observable event $W \in \mathcal{S}$, it is introduced a σ-field $\mathcal{A}$ of subset of the parameter space A such that $P^{\bullet}(W) : A \longrightarrow [0,1]$ is $\mathcal{A}$-measurable, that is,

$$[P^{\bullet}(W)]^{-1}\left(\mathcal{B}_{[0,1]}\right) \subset \mathcal{A} \quad \text{for } W \in \mathcal{S}, \tag{A.1}$$

where $\mathcal{B}_{[0,1]}$ denotes the Borel σ-field of $[0,1]$ and $[P^{\bullet}(W)]^{-1}(\bullet)$ denotes the pre-image of $P^{\bullet}(W)$ as a function of the parameter. Thus, the sampling probabilities become *sampling transitions*, namely $P^{\bullet}(\bullet) : A \times \mathcal{S} \longrightarrow [0,1]$ such that

1. For each parameter $a \in A$, $P^{a}(\bullet)$ is a sampling probability defined on $(S, \mathcal{S})$.
2. For each observable event $W \in \mathcal{S}$, $P^{\bullet}(W)$ is $\mathcal{A}$-measurable.

Thus, a *classical statistical model* is obtained, which is compactly written as

$$\mathcal{E}_c = \{(S, \mathcal{S}), P^{a} : a \in (A, \mathcal{A})\}.$$

This structure deserves two fundamental comments:

1. If the sampling transitions $\{P^{a} : a \in (A, \mathcal{A})\}$ are represented through density probability functions $\{p^{a} : a \in (A, \mathcal{A})\}$ –which is valid provided that $\{P^{a} : a \in (A, \mathcal{A})\}$ are dominated by a σ-finite measure λ (as the Lebesgue measure)–, $\mathcal{E}_c$ is equivalently rewritten as

   $$\mathcal{E}_c = \{(S, \mathcal{S}), p^{a} : a \in (A, \mathcal{A})\}.$$

   The density probability function p^{a} is the *likelihood function*: it is actually an $\mathcal{A}$-measurable function from $A \times S$ into $[0,1]$. In most applications, it is only required that the likelihood function be continuous (hence, measurable) as a function of the parameter.

2. From (A.1) it follows that the smallest σ-field that makes the sampling probabilities measurable as a function of the parameter is given by

   $$\bigvee_{W \in \mathcal{S}} [P^{\bullet}(W)]^{-1}\left(\mathcal{B}_{[0,1]}\right), \tag{A.2}$$

   where, for $\mathcal{B}_1$ and $\mathcal{B}_2$ σ-fields, $\mathcal{B}_1 \vee \mathcal{B}_2$ denotes the smallest σ-field containing $\mathcal{B}_1 \cup \mathcal{B}_2$.

3. To ensure that the sampling probabilities become sampling transitions it is enough to choose the σ-field $\mathcal{A}$ such that it contains (A.2). However, this is precisely the origin of the identification problem: the modelling challenge is that the parameters are measurable functions of the sampling transitions, that is,

$$\mathcal{A} \subset \bigvee_{W \in \mathcal{S}} [P^{\bullet}(W)]^{-1} \left(\mathcal{B}_{[0,1]}\right). \tag{A.3}$$

In this way, the parameters can be considered as representing features of the population of interest. This statement become explicit if we introduce the theory of Bayesian identification, which in turn requires defining the concepts of sufficient parameter and minimal sufficient parameter. This will be done next.

On the parameter space $(A, \mathcal{A})$, we introduce a probability measure μ, typically called *prior* distribution. It follows that there exists a *unique* probability measure Π defined on the product space $(A \times S, \mathcal{A} \vee \mathcal{S})$ such that

$$\Pi(E \times Y) = \int_E P^a(Y)\mu(da) \qquad E \times Y \in \mathcal{A} \times \mathcal{S}.$$

For details and references, see Florens et al. (1990, Chapter 1). By construction, P^a becomes a regular version of the restriction to $\mathcal{S}$ of the conditional probability Π given $\mathcal{A}$ and, therefore, it is denoted as $P^{\mathcal{A}}$. The prior distribution μ corresponds to the marginal probability $\Pi(\bullet \times S)$ on $(A, \mathcal{A})$. The marginal probability on $(S, \mathcal{S})$ given by $P(W) = \Pi(A \times W)$ for $W \in \mathcal{S}$ is called *predictive distribution*.

The unique probability measure Π can be decomposed into the predictive distribution P and a regular conditional probability given $\mathcal{S}$, represented by a transition denoted a $\mu^{\mathcal{S}}$: this is the so-called *posterior transition*.

A.2 Conditional independence

In order to develop the theory of Bayesian identification, we need to define the measurable completion of a σ-field as well as the concept of conditional independence. Let $(M, \mathcal{M}, P)$ be a probability space. Let $\mathcal{M}_0 = \{\emptyset, M\}$ be the trivial σ-field. The *completed trivial σ-field* $\overline{\mathcal{M}}_0$ is defined as

$$\overline{\mathcal{M}}_0 = \{E \in \mathcal{M} : P(E)^2 = P(E)\}.$$

Note that this completed σ-filed is still a sub-σ-field of $\mathcal{M}$ because we complete it through measurable sets of probability 0 or 1 (also called *null sets*). We also denote, for $\mathcal{N}$ a sub-σ-field of $\mathcal{M}$,

$$\overline{\mathcal{N}} = \mathcal{N} \vee \overline{\mathcal{M}}_0.$$

This is the smallest σ-field containing $\mathcal{N}$ and the measurable null sets and it is usually called the *completed sub-σ field* $\overline{\mathcal{N}}$.

Let $\mathcal{N}$ be a sub-σ-field of $\mathcal{M}$. We denote by $[\mathcal{N}]^+$ the set of positive $\mathcal{N}$-measurable random variables. Let $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ be sub-σ-fields of $\mathcal{M}$. $\mathcal{M}_1$ is independent of $\mathcal{M}_2$ conditionally on $\mathcal{M}_3$, which we write as $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 \mid \mathcal{M}_3$, if and only if one of the following equivalent conditions hold:

1. $E(m_1 m_2 \mid \mathcal{M}_3) = E(m_1 \mid \mathcal{M}_3)\, E(m_2 \mid \mathcal{M}_3)$ for all $m_1 \in [\mathcal{M}]_1^+$ and $m_2 \in [\mathcal{M}]_2^+$.
2. $E(m_1 \mid \mathcal{M}_2 \vee \mathcal{M}_3) = E(m_1 \mid \mathcal{M}_3)$ for all $m_1 \in [\mathcal{M}]_1^+$.

For a proof, see Florens et al. (1990, Theorem 2.2.1).

The first condition mimics the definition of mutual independence of σ-fields. However, it shows that the conditional independence $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 \mid \mathcal{M}_3$ is symmetric in $\mathcal{M}_1$ and $\mathcal{M}_2$. The second condition provides a semantic meaning of it, namely $\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 \mid \mathcal{M}_3$ means that the process generating m_1 conditionally on $\mathcal{M}_2 \vee \mathcal{M}_3$ only depends on $\mathcal{M}_3$. As it is well known, this is quite useful when for specifying a statistical model. But it is not only a semantic interpretation, but also corresponds to a characterization of the conditional independence in terms of measurability:

$$\mathcal{M}_1 \perp\!\!\!\perp \mathcal{M}_2 \mid \mathcal{M}_3 \iff E(m_1 \mid \mathcal{M}_2 \vee \mathcal{M}_3) \in [\overline{\mathcal{M}}_3]^+ \quad \forall\, m_1 \in [\mathcal{M}_1]^+; \tag{A.4}$$

that is, the conditional expectation $E(m_1 \mid \mathcal{M}_2 \vee \mathcal{M}_3)$ is a.s. a measurable fucntion of $\mathcal{M}_3$. Note that the *a.s.-condition* involves null sets: this will be critical for defining Bayesian identification. For a proof of the previous equivalence, see Florens et al. (1990, Theorem 2.2.6).

A.3 Sufficient parameters

The advantage of this general construction is that all sub-σ-field $\mathcal{T}$ of $\mathcal{S}$ corresponds to a *statistics*, whereas all sub-σ-field $\mathcal{B}$ of $\mathcal{A}$ corresponds to a *parameter of interest*. If all these σ-fields are generated by random variables, the Lemma of Dynkin-Doob (Rao, 2005, Chapter 2.1, Proposition 3) ensures this interpret.

Let $\mathcal{B} \subset \mathcal{A}$ be a parameter of interest. $\mathcal{B}$ is a *sufficient parameter for* $\mathcal{S}$ if and only if

$$\mathcal{S} \perp\!\!\!\perp \mathcal{A} \mid \mathcal{B}. \tag{A.5}$$

Using the semantic interpretation of a conditional independence, and noticing that $\mathcal{B} \subset \mathcal{A} \iff \mathcal{A} = \mathcal{A} \vee \mathcal{B}$, it holds that

1. $E(f \mid \mathcal{A}) = E(f \mid \mathcal{B})$ for all $f \in [\mathcal{S}]^+$. That is, the sampling process is fully characterized by the parameter $\mathcal{B}$, being $\mathcal{A}$ non-informative; or, equivalently,
2. $E[b \mid \mathcal{A} \vee \mathcal{S}] = E[b \mid \mathcal{A}]$ for all $b \in [\mathcal{B}]^+$. That is, conditionally on $\mathcal{A}$, the prior distribution on $\mathcal{B}$ is not updated by the data $\mathcal{S}$.

Once a sufficient parameter have been made explicit, it is possible to find multiples sufficient parameters. As a matter of fact, using the properties of conditional independence, it follows that

$$\mathcal{C} \subset \mathcal{A} \quad \text{and} \quad \text{Condition (A.5)} \Longrightarrow \mathcal{S} \perp\!\!\!\perp \mathcal{A} \mid \mathcal{B} \vee \mathcal{C},$$

which means that $\mathcal{B} \vee \mathcal{C}$ is also a sufficient parameter for $\mathcal{S}$.

A.4 Minimal sufficient parameter

This last implication raises the question of how to construct a *minimum sufficient parameter*, that is, one that conditionally to $\mathcal{A}$ is not updated by the observations and that if there is another sufficient parameter, the former is a function of the latter. In order to answer this question, let $\Im = \{\mathcal{B} \subset \mathcal{A} \,:\, \mathcal{S} \perp\!\!\!\perp \mathcal{A} \mid \mathcal{B}\}$ be the class of all sufficient parameters for $\mathcal{S}$. Since $\mathcal{A} \in \Im$, it follows that $\Im \neq \emptyset$ and therefore it is possible to choose two sufficient parameters $\mathcal{B}_1$ and $\mathcal{B}_2 \in \Im$ for $\mathcal{S}$. Using the characterization (A.4), it follows that

$$\begin{aligned} \mathcal{S} \perp\!\!\!\perp \mathcal{A} \mid \mathcal{B}_1 &\iff E(s \mid \mathcal{A}) \in [\overline{\mathcal{B}}_1]^+ \quad \text{for all } s \in [\mathcal{S}]^+, \\ \mathcal{S} \perp\!\!\!\perp \mathcal{A} \mid \mathcal{B}_2 &\iff E(s \mid \mathcal{A}) \in [\overline{\mathcal{B}}_2]^+ \quad \text{for all } s \in [\mathcal{S}]^+, \end{aligned}$$

from which it follows that

$$E(s \mid \mathcal{A}) \in [\overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2]^+ \quad \text{for all } s \in [\mathcal{S}]^+ \iff \mathcal{S} \perp\!\!\!\perp \mathcal{A} \mid \overline{\mathcal{B}}_1 \cap \overline{\mathcal{B}}_2,$$

where $\overline{\mathcal{B}}_j = \mathcal{B}_j \vee \{E \in \mathcal{A} \,:\, \mu(E)^2 = \mu(E)\}$ where μ is the prior probability distribution defined as $\mu(E) = \Pi(E \times S)$. Consequently, the minimal sufficient parameter $\mathcal{B}_{\text{min}} \in \Im$ *always* exists and it is given by

$$\mathcal{B}_{\text{min}} = \bigcap_{\mathcal{B} \in \Im} \overline{\mathcal{B}}. \tag{A.6}$$

It is possible to show that $\overline{\mathcal{B}_{\text{min}}} = \mathcal{B}_{\text{min}}$. This equality implies that the minimum sufficient parameter contains *all the events of priori probability 0 or 1*; these events are sometimes interpreted as dogmatic beliefs that, in the case of a Bayesian dominated model, are not updated by observations; for a proof, see Mouchart (1976) and San Martín (2018).

The minimal sufficient parameter $\mathcal{B}_{\text{min}}$ can be expressed in an operational terms. As a matter of fact, since $\mathcal{B}_{\text{min}}$ is a sufficient parameter, if follows from (A.4) that

$$\mathcal{S} \perp\!\!\!\perp \mathcal{A} \mid \mathcal{B}_{\text{min}} \iff E(s \mid \mathcal{A}) \in [\mathcal{B}_{\text{min}}]^+ \quad \text{for all } s \in [\mathcal{S}]^+.$$

But the σ-field generated by every version of the sampling expectations $\sigma\{E(s \mid \mathcal{A}) \,:\, s \in [\mathcal{S}]^+\}$ is the smallest sub-σ-field of $\mathcal{A}$ which makes measurable those sampling expectations; this follows by the definition of a σ-field generated by a random variable. Consequently,

$$\sigma\{E(s \mid \mathcal{A}) \,:\, s \in [\mathcal{S}]^+\} \subset \mathcal{B}_{\text{min}}.$$

On the other hand, $\sigma\{E(s \mid \mathcal{A}) \,:\, s \in [\mathcal{S}]^+\} \subset \mathcal{A}$. Therefore, by the definition of conditional independence, it follows that

$$\mathcal{S} \perp\!\!\!\perp \mathcal{A} \mid \sigma\{E(s \mid \mathcal{A}) \,:\, s \in [\mathcal{S}]^+\},$$

that is, $\sigma\{E(s \mid \mathcal{A}) \,:\, s \in [\mathcal{S}]^+\}$ is sufficient parameter for $\mathcal{S}$. Consequently,

$$\mathcal{B}_{\text{min}} \subset \sigma\{E(s \mid \mathcal{A}) \,:\, s \in [\mathcal{S}]^+\}.$$

Thus, the minimal sufficient parameter is given by $\sigma\{E(s \mid \mathcal{A}) \,:\, s \in [\mathcal{S}]^+\}$, which corresponds to the smallest sub-σ-field of $\mathcal{A}$ that makes measurable the sampling expectations.

A.5 Bayesian identification

Following Florens et al. (1990), a Bayesian model $(A \times S, \mathcal{A} \vee \mathcal{S}, \Pi)$ is said to be b-identified if $\mathcal{A}$ is the minimal sufficient parameter, which means that

$$\mathcal{A} \subset \sigma\{E(s \mid \mathcal{A}) \,:\, s \in [\mathcal{S}]^+\}. \tag{A.7}$$

This definition deserves the following comments:

1. This concept of b-identification is related to the concept of s-identification. More specifically, if $\mathcal{A}$ is a Blackwell space and $\mathcal{S}$ is a separable σ-field, then s-identification *always* implies b-identification for all prior distribution defined on $(A, \mathcal{A})$; for details and proofs, see Florens, Mouchart, and Rolin (1985) and Florens et al. (1990, Chapter 4.6.2). In the context of the models discussed in the main text, both the sample space and the parameter space are Euclidean spaces and, therefore, are separable and Blackwell; for details on Blackwell spaces, see Blackwell (1956).

2. Relationship (A.7) shows that a parameter $\mathcal{A}$ is b-identified if and only if it is a measurable function of the minima sufficient parameter; or, equivalently, the parameter $\mathcal{A}$ is b-identified if it can be written as a measurable function of sampling expectations.

3. We have pointed out above that the identification problem arises when a measurability structure is introduced on the sampling distributions as functions of the parameters. Specifically, the identification problem is related to the possibility or not of expressing the parameters as functions of the sampling transitions; see relation (A.3). b-identifiability explicitly is focused on this aspect as relation (A.7) shows. Moreover, it can be proved that

$$\bigvee_{W \in \mathcal{S}} [P^\bullet(W)]^{-1}\left(\mathcal{B}_{[0,1]}\right) \vee \{E \in \mathcal{A} \,:\, \mu(E)^2 = \mu(E)\} = \sigma\{E(s \mid \mathcal{A}) \,:\, s \in [\mathcal{S}]^+\}$$

   where μ is the prior probability distribution defined on $(A, \mathcal{A})$; for a proof, see Florens et al. (1990, Theorem 4.4.12).

4. The previous relation shows that the a priori distribution plays no role in identifiability except by means of events of prior probability 0 or 1; these probabilities can be interpreted as dogmatic restrictions. Consequently, it is not correct to state that a "good or correct prior elicitation" can overcome the identification problems.

5. Identifiability is rather related to the minimal sufficient parameter, which by definition captures all the information provided by the sampling process. In our view, this perfectly fits the Likelihood Principle: *For a given model the only information the data S contains about the parameter a is given by the likelihood function* (Lindley, 1983).

6. Moreover, the minimal sufficient parameter (or the b-identified parameter) is the only one which is updated by the data. As a matter of fact, let $\mathcal{C} \subset \mathcal{A}$ a parameter of interest, which is not necessarily a sufficient parameter. The sufficiency of $\mathcal{B}_{\min}$ implies that $\mathcal{S} \perp\!\!\!\perp \mathcal{C} \mid \mathcal{B}_{\min}$. Therefore,

$$E(c \mid \mathcal{S}) = E[E(c \mid \mathcal{B}_{\min} \vee \mathcal{S}) \mid \mathcal{S}] = E[E(c \mid \mathcal{B}_{\min}) \mid \mathcal{S}] \quad \text{for all } c \in [\mathcal{C}]^+,$$

where the first equality follows from the iterative property of the conditional expectation (Rao, 2005, Chapter 2.2, Proposition 1) and the second one follows from the sufficiency of $\mathcal{B}_{\text{min}}$. It can be seen that in a Bayesian model it is not possible to learn by the observations of any parameter except the identified one. Moreover, what is learned by observations of any other parameter is reduced to the learning of the identified parameter.

7. This theory of b-identification has been used to correct claims that it is possible to update unidentified parameters and that, therefore, the Bayesian framework can make inferences that are impossible in the sampling theory framework. Some examples are the following:

   (a) Poirier (1998), in a series of examples, illustrates how it is possible to update unidentified parameters. Two of these examples are discussed in San Martín, Rolin, and Castro (2013): a basic case in Poirier (1998, Section 3.1) and a hierarchical model in Poirier (1998, Section 3.2). In the first example, the reasoning error lies in modifying the sampling process by marginalizing a nuisance parameter. However, Bayesian inference begins by *fixing a sampling process*, which cannot be subjectively changed unless the research question is modified. The second example intends to show that it is possible to calculate the a posteriori distribution of an unidentified parameter and, consequently, to update it. However, San Martín et al. (2013) show that such a posteriori distribution is a function of the a posteriori distribution of the identified parameter. Let us mention that this example has been used to claim that the behavior of Markov chains in a Gibbs procedure is affected by the lack of identifiability; see, among many others, Kass, Carlin, Gelman, and Neal (1998), Carlin and Louis (2000), Eberly and Carlin (2000) and Xie and Carlin (2006). However, this perspective is erroneous: identifiability has nothing to do with the convergence of the Gibbs sampler; for a discussion, see San Martín et al. (2013, Section 5.2).

   (b) "Unidentifiability causes no real difficulty in the Bayesian framework", claimed Lindley (1983). In this line, Wechsler, Izbicki, and Esteves (2013) discuss a simple example where it is possible to compute the posterior distribution of an unidentified parameter and to make inferences. The example can be related to classification through a diagnostic test. However, using the characterization of minimal sufficient parameter when the parameter space is discrete (San Martın & González, 2010, Section 3), San Martín (2018) show that such a posterior distribution is the posterior distribution of the identified parameter. Consequently, unidentifiability causes *real difficulty* in the Bayesian framework also.

Let us finish this section by pointing out that part of this theory has been presented in the psychometric literature by San Martín, Jara, Rolin, and Mouchart (2011).

A.6 Consequences for the present manuscript

The main text deals with the specification of a cohort varying, temporally dynamic, value-added models. Such a specification is motivated by the theory above presented in at least two aspects:

1. The specification of the likelihood function or sampling distribution. As discussed above, a Bayesian model is constructed by fixing the sampling process, by introducing a measurement structure over the sampling distributions and finally by endowing the parameter space with a priori probability distribution. As it is discussed in the main text, value-added models are typically specified either in a fixed-effect framework or in a random -effect one. Under the first approach, the school effect is a parameter of the likelihood function. However, under the second approach, it corresponds to a unobserved random variable and consequently the likelihood function is obtained after integrating out the school effect. Motivated by the meaning of a school underlying random-effect paradigm, we specify a Bayesian model such that the likelihood function or sampling process corresponds to the distribution which is obtained after integrating out the school effect. The parameters of this distribution are accordingly endowed with the prior specification. At the estimation level, we work in the structural model generating jointly the observations and the random effects, conditionally on the covariates. However, at this level, the Bayesian inference is done on the identified parameters. Let us remark that if the distributions of school effects are viewed *as* priori distributions, the corresponding Bayesian specification and the subsequent inference is the counterpart of a fixed-effect model.

2. The identification analysis developed in the main text is essentially based on the Bayesian concept of identification. As a matter of fact, after inducing the likelihood function, the identifiability of the parameters of interest is obtained by proving that they are function of both the conditional expectation and conditional variance-covariance matrix of the likelihood. Note that all the information provided by the sampling process is a function of the two first moments because the likelihood function is a normal distribution.

3. Finally, the statistical meaning of the parameters of interest as well as of the value-added indicators is based on an identifiability analysis. This is in agreement with the Likelihood Principle which, according to pur discussion of Bayesian identifiability, fits the very nature of parameter identification.

B Choosing Between Fixed-Effects and Random-Effects Models: A Proposed Modelling Criterion

Value-added models can be approached from two perspectives: fixed effects and random effects. In this paper, we adopt the latter. The question arises: how should we decide between these two approaches? To address this query, let's delve into a standard value-added model that delineates the association between a n_i-dimensional vector $\boldsymbol{Y}_i$ of measurements from school i, incorporating both a school effect α_i and an $n_i \times p$ matrix $\boldsymbol{X}_i$ of covariates (excluding a column of 1's). The model is expressed as follows:

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \alpha_i \imath_{n_i} + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \sigma^2 I_{n_i}). \tag{B.1}$$

As widely recognized, this specification embodies a random-effects model if the school effect α_i is random. In such instances, (B.1) should be incorporated within a *marginal-conditional* framework, defined by:

$$
\begin{array}{rl}
\text{(i)} & (\boldsymbol{Y}_i \mid \boldsymbol{X}_i, \alpha_i) \sim \mathcal{N}(\boldsymbol{X}_i\boldsymbol{\beta} + \alpha_i \imath_{n_i}, \sigma^2 I_{n_i}); \\
\text{(ii)} & (\alpha_i \mid \boldsymbol{X}_i) \overset{\text{iid}}{\sim} \mathcal{N}(\phi_0, \tau^2), \ i = 1, \ldots, I; \\
\text{(iii)} & \boldsymbol{X}_i \text{ is left unspecified.}
\end{array}
\tag{B.2}
$$

This specification yields a statistical model that explicates the generation of observable data only:

$$
(\boldsymbol{Y}_i \mid \boldsymbol{X}_i) \sim \mathcal{N}(\phi_0 \imath_{n_1} + \boldsymbol{X}_i\boldsymbol{\beta}, \tau^2 J_{n_1} + \sigma^2 I_{n_1}), \tag{B.3}
$$

where $J_{n_1} = \imath_{n_1}\imath_{n_1}'$, I_{n_1} is the $n_1 \times n_1$ identity matrix, and $\imath_{n_1}$ is a n_1-dimensional vector of 1's.

It is well-established that this model implies correlation between the scores of two examinees:

$$
\text{cor}(Y_{ij}, Y_{ik} \mid \boldsymbol{X}_i) = \frac{\tau^2}{\tau^2 + \sigma^2}, \qquad j \neq k.
$$

This correlation is *induced* by the school effect due to:

$$
\begin{aligned}
\text{cov}(Y_{ij}, Y_{ik} \mid \boldsymbol{X}_i) &= \text{cov}[E(Y_{ij} \mid \alpha_i, \boldsymbol{X}_i),\, E(Y_{ik} \mid \alpha_i, \boldsymbol{X}_i) \mid \boldsymbol{X}_i] + E[\text{cov}(Y_{ij}, Y_{ik} \mid \alpha_i, \boldsymbol{X}_i) \mid \boldsymbol{X}_{\boldsymbol{i}}] \\
&= \text{Var}(\alpha_i \mid \boldsymbol{X}_i) \ + \ 0,
\end{aligned}
$$

where the component equal to 0 is implied by (B.2.i).

This equality leads to the *characterization* of the school effect: α_i explains the unexplained heterogeneity in the Y_{ij} scores that is not accounted for by the covariates $\boldsymbol{X}_i$. Consequently, under the random effects approach, schools are perceived as entities that introduce heterogeneity in students' achievements as measured by test scores Y_{ij} (after adjusting for observable covariates at both student and school levels). If we accept this characterization of a school, then the random-effects approach is preferable.

Conversely, if α_i in (B.1) is considered fixed, then the corresponding statistical model is:

$$
(\boldsymbol{Y}_i \mid \boldsymbol{X}_i) \sim \mathcal{N}(\boldsymbol{X}_i\boldsymbol{\beta} + \alpha_i \imath_{n_1}, \sigma^2 I_{n_i}). \tag{B.4}
$$

It is evident that the statistical model (B.4) diverges significantly from (B.3). In particular, in (B.4), the school effect only influences the predicted score of Y_{ij}, adding the *same* quantity to the predicted scores of all examinees. This implies that, under this approach, schools are characterized by a common feature: each school acts as an entity that has the same effect on student achievements by adding an additive factor, but it does not establish relationships between these scores. If we accept this characterization of a school, then we should opt for a fixed-effects approach.

C Joint distribution for the case of $T = 2$ cohorts

For $T = 2$, the following arguments prove relationships (2.10)-(2.12) of the main text. The modelling process is based on the following marginal-conditional decomposition, which is displayed taking into account the temporal order of the phenomenon:

$$\begin{aligned}
p(\boldsymbol{Y}_{2i}, \boldsymbol{Y}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{X}_{1i}, \alpha_{2i}, \alpha_{1i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1) &= p(\boldsymbol{Y}_{2i} \mid \boldsymbol{Y}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{X}_{1i}, \alpha_{2i}, \alpha_{1i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1) \times p(\alpha_{2i} \mid \boldsymbol{Y}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{X}_{1i}, \alpha_{1i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1) \times \\
&\quad p(\boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{2i}, \boldsymbol{X}_{1i}, \alpha_{1i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1) \times p(\alpha_{1i} \mid \boldsymbol{X}_{2i}, \boldsymbol{X}_{1i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1) \times \\
&\quad p(\boldsymbol{X}_{2i}, \boldsymbol{X}_{1i} \mid \boldsymbol{\psi}_2, \boldsymbol{\psi}_1) \times p(\boldsymbol{\psi}_2, \boldsymbol{\psi}_1) \\
&= p(\boldsymbol{Y}_{2i} \mid \boldsymbol{X}_{2i}, \alpha_{2i}, \boldsymbol{\beta}_2, \sigma_2^2) \times p(\alpha_{2i} \mid \boldsymbol{Y}_{1i}, \alpha_{1i}, \phi_{02}, \phi_{12}, \gamma_2, \tau_2^2) \times \\
&\quad p(\boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \alpha_{1i}, \boldsymbol{\beta}_1, \sigma_1^2) \times p(\alpha_{1i} \mid \phi_{01}, \tau_1^2) \times p(\boldsymbol{X}_{2i}, \boldsymbol{X}_{1i}) \times p(\boldsymbol{\psi}_2, \boldsymbol{\psi}_1).
\end{aligned}$$

The second equality allows us to make explicit the structural assumptions underlying the model:

$$\begin{array}{rl}
\text{(i)} & \boldsymbol{Y}_{2i} \perp\!\!\!\perp \boldsymbol{Y}_{1i}, \boldsymbol{X}_{1i}, \alpha_{1i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1 \mid \boldsymbol{X}_{2i}, \alpha_{2i}, \boldsymbol{\beta}_2, \sigma_2^2; \\
\text{(ii)} & \alpha_{2i} \perp\!\!\!\perp \boldsymbol{X}_{2i}, \boldsymbol{X}_{1i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1 \mid \boldsymbol{Y}_{1i}, \alpha_{1i}, \phi_{02}, \phi_{12}, \gamma_2, \tau_2^2; \\
\text{(iii)} & \boldsymbol{Y}_{1i} \perp\!\!\!\perp \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1 \mid \boldsymbol{X}_{1i}, \alpha_{1i}, \boldsymbol{\beta}_1, \sigma_1^2; \\
\text{(iv)} & \alpha_{1i} \perp\!\!\!\perp \boldsymbol{X}_{2i}, \boldsymbol{X}_{1i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1 \mid \phi_{01}, \tau_1^2; \\
\text{(iv)} & \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i} \perp\!\!\!\perp \boldsymbol{\psi}_2, \boldsymbol{\psi}_1.
\end{array} \tag{C.1}$$

The specification of the model is completed by the following assumptions:

$$\begin{array}{rl}
\text{(i)} & (\boldsymbol{Y}_{2i} \mid \boldsymbol{X}_{2i}, \alpha_{2i}, \boldsymbol{\beta}_2, \sigma_2^2) \sim \mathcal{N}(\boldsymbol{X}_{2i}\boldsymbol{\beta}_2 + \alpha_{2i}\imath_{n_{2i}}, \sigma_2^2 I_{n_{2i}}) \\
\text{(ii)} & (\alpha_{2i} \mid \boldsymbol{Y}_{1i}, \alpha_{1i}, \phi_{02}, \phi_{12}, \gamma_2, \tau_2^2) \sim \mathcal{N}(\phi_{02} + \phi_{12}\alpha_{1i} + \gamma_2\, \overline{Y}_{1i\bullet}, \tau_2^2(1-\phi_{12}^2)) \\
\text{(iii)} & (\boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \alpha_{1i}, \boldsymbol{\beta}_1, \sigma_1^2) \sim \mathcal{N}(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1 + \alpha_{1i}\imath_{n_{1i}}, \sigma_1^2 I_{n_{1i}}) \\
\text{(iv)} & (\alpha_{1i} \mid \phi_{01}, \tau_1^2) \sim \mathcal{N}(\phi_{01}, \tau_1^2) \\
\text{(v)} & \text{The distribution of } (\boldsymbol{X}_{2i}, \boldsymbol{X}_{1i}) \text{ remains unspecified} \\
\text{(vi)} & (\boldsymbol{\psi}_2, \boldsymbol{\psi}_1) \sim \pi_\psi.
\end{array} \tag{C.2}$$

Conditions (C.1) and (C.2) corresponds to conditions (2.1)-(2.9) of the main text for $T = 2$.

In what follows, we obtain the joint distribution of $(\boldsymbol{Y}_{2i}, \boldsymbol{Y}_{1i}, \alpha_{2i}, \alpha_{1i} \mid \boldsymbol{X}_{2i}, \boldsymbol{X}_{1i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1)$, which is normal. Each component of the mean is computed as follows:

- $E(\alpha_{1i} \mid \boldsymbol{X}_{2i}, \boldsymbol{X}_{1i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1) = E(\alpha_{1i} \mid \phi_{01}, \tau_1^2) = \phi_{01}$ by successively applying (C.1.iv) and (C.2.iv).

- Applying successively (C.1.iii), (C.2.iii), (C.1.iv) and (C.2.iv) we obtain

$$\begin{aligned}
E(\boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1) &= E[E(\boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \alpha_{1i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1] \\
&= E[E(\boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \alpha_{1i}, \boldsymbol{\beta}_1, \sigma_1^2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1] \\
&= E[\boldsymbol{X}_{1i}\boldsymbol{\beta}_1 + \alpha_{1i}\imath_{n_{1i}} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1] \\
&= \boldsymbol{X}_{1i}\boldsymbol{\beta}_1 + E[\alpha_{1i}\imath_{n_{1i}} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1] \\
&= \boldsymbol{X}_{1i}\boldsymbol{\beta}_1 + E[\alpha_{1i}\imath_{n_{1i}} \mid \phi_{01}, \tau_1^2] \\
&= \boldsymbol{X}_{1i}\boldsymbol{\beta}_1 + \phi_{01}\imath_{n_{1i}}.
\end{aligned}$$

- Applying successively (C.1.ii), (C.2.ii), (C.1.iv), (C.2,iv) and the previous result, we obtain

$$\begin{aligned}
E(\alpha_{2i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1) &= E[E(\alpha_{2i} \mid \boldsymbol{Y}_{1i}, \alpha_{1i}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1] \\
&= E[E(\alpha_{2i} \mid \boldsymbol{Y}_{1i}, \alpha_{1i}, \phi_{02}, \phi_{12}, \gamma_2, \tau_2^2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1] \\
&= E[\phi_{02} + \phi_{12}\alpha_{1i} + \gamma_2\,\overline{Y}_{1i\bullet} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1] \\
&= \phi_{02} + \phi_{12}\phi_{01} + \gamma_2 E(\overline{Y}_{1i\bullet} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1) \\
&= \phi_{02} + \phi_{12}\phi_{01} + \frac{\gamma_2}{n_{1i}}\imath'_{n_{1i}} E(\boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1) \\
&= \phi_{02} + \phi_{12}\phi_{01} + \frac{\gamma_2}{n_{1i}}\imath'_{n_{1i}}[\phi_{01}\imath_{n_{1i}} + \boldsymbol{X}_{1i}\boldsymbol{\beta}_1] \\
&= [\phi_{02} + (\phi_{12} + \gamma_2)\phi_{01}] + \gamma_2\overline{X}_{1i\bullet}\boldsymbol{\beta}_1,
\end{aligned}$$

  where $\overline{X}_{1i\bullet} = \frac{1}{n_{1i}}\imath'_{n_{1i}}\boldsymbol{X}_{1i}$ is a $1 \times p_1$ vector of means at the school level and, therefore, $\overline{X}_{1i\bullet}\boldsymbol{\beta}_1$ is a 1×1 matrix.

- Applying successively (C.1.i), (C.2.i) and the previous result, we obtain

$$\begin{aligned}
E(\boldsymbol{Y}_{2i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1) &= E[E(\boldsymbol{Y}_{2i} \mid \boldsymbol{Y}_{1i}, \alpha_{1i}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1] \\
&= E[E(\boldsymbol{Y}_{2i} \mid \boldsymbol{X}_{2i}, \alpha_{1i}, \boldsymbol{\beta}_2, \sigma_2^2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1] \\
&= E[\boldsymbol{X}_{2i}\boldsymbol{\beta}_2 + \alpha_{2i}\imath_{n_{2i}} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1] \\
&= \boldsymbol{X}_{2i}\boldsymbol{\beta}_2 + E(\alpha_{2i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1)\imath_{n_{2i}} \\
&= \boldsymbol{X}_{2i}\boldsymbol{\beta}_2 + [\phi_{02} + (\phi_{12} + \gamma_2)\phi_{01}]\imath_{n_{2i}} + \gamma_2\overline{X}_{1i\bullet}\boldsymbol{\beta}_1\imath_{n_{2i}}.
\end{aligned}$$

Consequently, the mean of $(\boldsymbol{Y}'_{2i}, \boldsymbol{Y}'_{1i}, \alpha_{2i}, \alpha_{1i} \mid \boldsymbol{X}_{2i}, \boldsymbol{X}_{1i}, \boldsymbol{\psi}_2, \boldsymbol{\psi}_1)'$ is given by

$$\begin{pmatrix}
\{\phi_{02} + \phi_{01}(\phi_{12} + \gamma_2)\}\imath_{n_{2i}} + \boldsymbol{X}_{2i}\boldsymbol{\beta}_2 + \gamma_2\overline{X}_{1i\bullet}\boldsymbol{\beta}_1\imath_{n_{2i}} \\
\phi_{01}\imath_{n_{1i}} + \boldsymbol{X}_{1i}\boldsymbol{\beta}_1 \\
\{\phi_{02} + \phi_{01}(\phi_{12} + \gamma_2)\} + \gamma_2\overline{X}_{1i\bullet}\boldsymbol{\beta}_1 \\
\phi_{01}
\end{pmatrix}.$$

Now, let compute each element of the variance-covariance matrix:

- $Var(\alpha_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) = \tau_1^2$ by (C.1.iv) and (C.2.iv).
- Applying successively (C.1.iii) and (C.2.iii), we obtain

$$\begin{aligned} Var(\boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) &= Var[E(\boldsymbol{Y}_{1i} \mid \alpha_{1i}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] + \\ &\quad + E[Var(\boldsymbol{Y}_{1i} \mid \alpha_{1i}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\ &= Var[E(\boldsymbol{Y}_{1i} \mid \alpha_{1i}, \boldsymbol{X}_{1i}, \boldsymbol{\beta}_1, \sigma_1^2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] + \\ &\quad + E[Var(\boldsymbol{Y}_{1i} \mid \alpha_{1i}, \boldsymbol{X}_{1i}, \boldsymbol{\beta}_1, \sigma_1^2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\ &= Var(\alpha_{1i}\imath_{n_{1i}} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \phi_{01}, \tau_1^2) + \sigma_1^2 I_{n_{1i}} \\ &= \tau_1^2\, J_{n_{1i}} + \sigma_1^2 I_{n_{1i}} \end{aligned}$$

  where $J_{n_{1i}} = \imath_{n_{1i}}\, \imath'_{n_{1i}}$.

- Applying successively (C.1.iii), (C.2.iii), (C.1.iv) and (C.2.iv) we obtain

$$\begin{aligned} cov(\boldsymbol{Y}_{1i}, \alpha_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) &= cov[E(\boldsymbol{Y}_{1i} \mid \alpha_{1i}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2), \boldsymbol{\alpha_{1i}} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\ &= cov[E(\boldsymbol{Y}_{1i} \mid \alpha_{1i}, \boldsymbol{X}_{1i}, \boldsymbol{\beta}_1, \sigma_1^2), \boldsymbol{\alpha_{1i}} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\ &= cov(\alpha_{1i}\imath_{n_{1i}}, \alpha_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \phi_{01}, \tau_1^2) \\ &= \tau_1^2 \imath_{n_{1i}}. \end{aligned}$$

- Applying successively (C.1.ii), (C.2.ii), (C.1.iii) and the previous results, we obtain

$$\begin{aligned} cov(\boldsymbol{Y}_{1i}, \alpha_{2i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) &= cov[\boldsymbol{Y}_{1i}, E(\alpha_{2i} \mid \boldsymbol{Y}_{1i}, \alpha_{1i}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\ &= cov[\boldsymbol{Y}_{1i}, E(\alpha_{2i} \mid \boldsymbol{Y}_{1i}, \alpha_{1i}, \phi_{02}, \phi_{12}, \gamma_2, \tau_2^2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\ &= cov(\boldsymbol{Y}_{1i}, \phi_{12}\alpha_{1i} + \gamma_2\, \overline{Y}_{1i\bullet} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\ &= \phi_{12} cov(\boldsymbol{Y}_{1i}, \alpha_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) + \\ &\quad \gamma_2 cov(\boldsymbol{Y}_{1i}, \overline{Y}_{1i\bullet} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \\ &= \phi_{12}\tau_1^2\imath_{n_{1i}} + \frac{\gamma_2}{n_{1i}} cov(\boldsymbol{Y}_{1i}, \boldsymbol{Y}'_{1i}\imath_{n_{1i}} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \\ &= \phi_{12}\tau_1^2\imath_{n_{1i}} + \frac{\gamma_2}{n_{1i}} V(\boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\beta}, \sigma_1^2)\imath_{n_{1i}} \\ &= \phi_{12}\tau_1^2\imath_{n_{1i}} + \frac{\gamma_2}{n_{1i}} \left[\tau_1^2\, J_{n_{1i}} + \sigma_1^2 I_{n_{1i}}\right] \imath_{n_{1i}} \\ &= \left[\phi_{12}\tau_1^2 + \gamma_2 \left(\tau_1^2 + \frac{\sigma_1^2}{n_{1i}}\right)\right] \imath_{n_{1i}}. \end{aligned}$$

- Applying successively (C.1.ii), (C.2.ii), (C.1.iv), (C.2.iv), (C.1.iii), (C.2.iii) and the previous results, we obtain

$$
\begin{aligned}
Var(\alpha_{2i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) &= Var[E(\alpha_{2i} \mid \boldsymbol{Y}_{1i}, \alpha_{1i}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] + \\
& \quad E[Var(\alpha_{2i} \mid \boldsymbol{Y}_{1i}, \alpha_{1i}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, , \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\
&= Var[E(\alpha_{2i} \mid \boldsymbol{Y}_{1i}, \alpha_{1i}, \phi_{02}, \phi_{12}, \gamma_2, \tau_2^2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}] + \\
& \quad E[Var(\alpha_{2i} \mid \boldsymbol{Y}_{1i}, \alpha_{1i}) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \phi_{02}, \phi_{12}, \gamma_2, \tau_2^2] \\
&= Var[\phi_{12}\alpha_{1i} + \gamma_2 \overline{Y}_{1i\bullet} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, , \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] + \tau_2^2(1 - \phi_{12}^2) \\
&= \phi_{12}^2\, \tau_1^2 + \gamma_2^2 Var\left[\frac{1}{n_{1i}} \imath'_{n_{1i}} \boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, , \boldsymbol{\psi}_1, \boldsymbol{\psi}_2\right] + \\
& \quad 2\phi_{12}\gamma_2\, cov\left[\alpha_{1i}, \frac{1}{n_{1i}} \boldsymbol{Y}'_{1i}\imath_{n_{1i}} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, , \boldsymbol{\psi}_1, \boldsymbol{\psi}_2\right] + \tau_2^2(1 - \phi_{12}^2) \\
&= \phi_{12}^2\, \tau_1^2 + \frac{\gamma_2^2}{n_{1i}^2} \imath'_{n_{1i}} Var(\boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\beta}_1, \sigma_1^2)\imath_{n_{1i}} + \\
& \quad 2\phi_{12} \frac{\gamma_2}{n_{1i}}\, cov(\alpha_{1i}, \boldsymbol{Y}'_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2)\imath_{n_{1i}} + \tau_2^2(1 - \phi_{12}^2) \\
&= \phi_{12}^2\, \tau_1^2 + \frac{\gamma_2^2}{n_{1i}^2} \imath'_{n_{1i}} \left[\tau_1^2\, J_{n_{1i}} + \sigma_1^2 I_{n_{1i}}\right] \imath_{n_{1i}} + \\
& \quad 2\phi_{12} \frac{\gamma_2}{n_{1i}} \tau_1^2 \imath'_{n_{1i}} \imath_{n_{1i}} + \tau_2^2(1 - \phi_{12}^2) \\
&= \phi_{12}^2\, \tau_1^2 + \gamma_2^2 \tau_1^2 + \gamma_2^2 \frac{\sigma_1^2}{n_{1i}} + 2\phi_{12}\gamma_2\tau_1^2 + \tau_2^2(1 - \phi_{12}^2) \\
&= \tau_1^2(\phi_{12} + \gamma_2)^2 + \gamma_2^2 \frac{\sigma_1^2}{n_{1i}} + \tau_2^2(1 - \phi_{12}^2)
\end{aligned}
$$

- Applying successively (C.1.ii), (C.2.ii) and the previous results, we obtain

$$
\begin{aligned}
cov(\alpha_{2i}, \alpha_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) &= cov[E(\alpha_{2i} \mid \boldsymbol{Y}_{1i}, \alpha_{1i}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2), \alpha_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\
&= cov[E(\alpha_{2i} \mid \boldsymbol{Y}_{1i}, \alpha_{1i}, \phi_{02}, \phi_{12}, \gamma_2, \tau_2^2), \alpha_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\
&= cov(\phi_{12}\alpha_{1i} + \gamma_2 \overline{Y}_{1i\bullet}, \alpha_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \\
&= \phi_{12}\tau_1^2 + \frac{\gamma_2}{n_{1i}} \imath'_{n_{1i}} cov(\boldsymbol{Y}_{1i}, \alpha_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \\
&= \phi_{12}\tau_1^2 + \frac{\gamma_2}{n_{1i}} \imath'_{n_{1i}} \imath_{n_{1i}} \tau_1^2 \\
&= \tau_1^2(\phi_{12} + \gamma_2)
\end{aligned}
$$

- Applying successively (C.1.i), (C.2.i) and the previous results, we obtain

$$
\begin{aligned}
cov(\boldsymbol{Y}_{2i}, \alpha_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) &= cov[E(\boldsymbol{Y}_{2i} \mid \alpha_{1i}, \alpha_{2i}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2), \alpha_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\
&= cov[E(\boldsymbol{Y}_{2i} \mid \alpha_{2i}, \boldsymbol{X}_{2i}, \boldsymbol{\beta}_2, \sigma_2^2), \alpha_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\
&= cov(\alpha_{2i}\imath_{n_{2i}}, \alpha_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \\
&= \tau_1^2(\phi_{12} + \gamma_2)\imath_{n_{2i}}
\end{aligned}
$$

- Applying successively (C.1.i), (C.2.i) and the previous results, we obtain

$$\begin{aligned} cov(\boldsymbol{Y}_{2i}, \alpha_{2i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) &= cov[E(\boldsymbol{Y}_{2i} \mid \alpha_{2i}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2), \alpha_{2i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\ &= cov[E(\boldsymbol{Y}_{2i} \mid \alpha_{2i}, \boldsymbol{X}_{2i}, \boldsymbol{\beta}_2, \sigma_2^2), \alpha_{2i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\ &= cov(\alpha_{2i}\imath_{n_{2i}}, \alpha_{2i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \\ &= \left[\tau_1^2(\phi_{12}+\gamma_2)^2 + \gamma_2^2\frac{\sigma_1^2}{n_{1i}} + \tau_2^2(1-\phi_{12}^2)\right] \imath_{n_{2i}}. \end{aligned}$$

- Applying successively (C.1.i), (C.2.i) and the previous results, we obtain

$$\begin{aligned} cov(\boldsymbol{Y}_{2i}, \boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) &= cov[E(\boldsymbol{Y}_{2i} \mid \boldsymbol{Y}_{1i}, \alpha_{2i}, \boldsymbol{X}_{2i}, \boldsymbol{X}_{1i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2), \boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\ &= cov[E(\boldsymbol{Y}_{2i} \mid \alpha_{2i}, \boldsymbol{X}_{2i}, \boldsymbol{\beta}_2, \sigma_2^2), \boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\ &= cov(\alpha_{2i}\imath_{n_{2i}}, \boldsymbol{Y}_{1i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \\ &= \left[\phi_{12}\tau_1^2 + \gamma_2\left(\tau_1^2 + \frac{\sigma_1^2}{n_{1i}}\right)\right] \imath_{n_{2i}}\imath_{n_{1i}}' \end{aligned}$$

- Applying successively (C.1.i), (C.2.i) and the previous results, we obtain

$$\begin{aligned} Var(\boldsymbol{Y}_{2i} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) &= Var[E(\boldsymbol{Y}_{2i} \mid \alpha_{2i}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] + \\ &\quad + E[Var(\boldsymbol{Y}_{2i} \mid \alpha_{2i}, \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\ &= Var[E(\boldsymbol{Y}_{2i} \mid \alpha_{2i}, \boldsymbol{X}_{2i}, \boldsymbol{\beta}_2, \sigma_2^2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] + \\ &\quad + E[Var(\boldsymbol{Y}_{2i} \mid \alpha_{2i}, \boldsymbol{X}_{2i}, \boldsymbol{\beta}_2, \sigma_2^2) \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2] \\ &= Var(\alpha_{2i}\imath_{n_{2i}} \mid \boldsymbol{X}_{1i}, \boldsymbol{X}_{2i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2) + \sigma_2^2 I_{n_{2i}} \\ &= \left[\tau_1^2(\phi_{12}+\gamma_2)^2 + \gamma_2^2\frac{\sigma_1^2}{n_{1i}} + \tau_2^2(1-\phi_{12}^2)\right] J_{n_{2i}} + \sigma_2^2 I_{n_{2i}} \end{aligned}$$

Consequently, the variance-covariance matrix of $(\boldsymbol{Y}_{2i}', \boldsymbol{Y}_{1i}', \alpha_{2i}, \alpha_{1i} \mid \boldsymbol{X}_{2i}, \boldsymbol{X}_{1i}, \boldsymbol{\psi}_1, \boldsymbol{\psi}_2)'$ is given by

$$\begin{pmatrix} \omega_{2i}\, J_{n_{2i}} + \sigma_2^2 I_{n_{2i}} & \delta_{12i}\, \imath_{n_{2i}}\imath_{n_{1i}}' & \omega_{2i}\imath_{n_{2i}} & (\phi_{12}+\gamma_2)\tau_1^2\imath_{n_{2i}} \\ & \tau_1^2 J_{n_{1i}} + \sigma_1^2 I_{n_{1i}} & \delta_{12}\ \imath_{n_{1i}} & \tau_1^2\imath_{n_{1i}} \\ & & \omega_{2i} & (\phi_{12}+\gamma_2)\tau_1^2 \\ & & & \tau_1^2 \end{pmatrix}, \tag{C.3}$$

where

$$\omega_{2i} \doteq \tau_1^2(\phi_{12}+\gamma_2)^2 + \frac{\gamma_2^2\sigma_1^2}{n_{1i}} + \tau_2^2(1-\phi_{12}^2), \qquad \delta_{12i} \doteq \phi_{12}\tau_1^2 + \gamma_2\left(\tau_1^2 + \frac{\sigma_1^2}{n_{1i}}\right). \tag{C.4}$$

D Joint distribution for the general case $T > 3$

The preceding derivations provide the key to obtaining recursive equations for deriving the full distribution of $\boldsymbol{Y}_{1,i}^t$ given $(\boldsymbol{X}_{1i}^t, \boldsymbol{\psi}_1^T)$, which is a normal one. Its expectations are given by

$$E(\boldsymbol{Y}_{ti} \mid \boldsymbol{X}_{1i}^T, \boldsymbol{\psi}_1^T) = E(E(\boldsymbol{Y}_{ti} \mid \boldsymbol{X}_{1i}^T, \alpha_{ti}, \boldsymbol{\psi}_1^T) \mid \boldsymbol{X}_{1i}^T, \boldsymbol{\psi}_1^T) = \boldsymbol{X}_{ti}\boldsymbol{\beta}_t + E(\alpha_{ti} \mid \boldsymbol{X}_{1i}^T, \boldsymbol{\psi}_1^T)\boldsymbol{\iota}_{n_{ti}}$$

for every $t = 1, \ldots, T$. The sequential model (2.1)–(2.11) enables us to derive a recursive formula for the conditional expectation of the random effect, which is

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

that holds true for every $t = 3, \ldots, T$. The expectations for $t = 1$ and $t = 2$ have been provided in the previous section. The derivation of the general formula for $t > 2$ follows using the techniques detailed in the previous section.

For every $1 \leqslant s < t \leqslant T$ and from the sequential model (2.1)–(2.11) with similar techniques we derive

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

Similarly we derive the recursive equations on the variance of random effects. Denote $a_t = V(\alpha_{ti} \mid \boldsymbol{X}_{1i}^t, \boldsymbol{\psi}_1^T)$ (for a fixed school i), then we can derive

$$a_t = (\phi_{1t} + \gamma_t)^2 a_{t-1} + n_{t-1,i}^{-1}\gamma_t^2\sigma_{t-1}^2 + \tau_t^2(1 - \phi_{1t}^2) .$$

for every $t > 2$ (a_1 and a_2 were derived in the Section C). An explicit solution is given by

$$V(\alpha_{ti} \mid \boldsymbol{X}_{1i}^t, \boldsymbol{\psi}_1^T) = V(\alpha_{1i} \mid \boldsymbol{X}_{1i}^t, \boldsymbol{\psi}_1^T) \prod_{k=0}^{t-1} A_{t-k} + \sum_{k=0}^{t-2} \prod_{l=-1}^{k-1} A_{t-l}B_{t-k}$$

where $A_k = (\phi_{1,t-k} + \gamma_{t-k})^2$ 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 $A_{-1} = 1$.

E Computation of (2.24)–(2.27)

In order to simplify the notation, let us denote $VA_{ti}(\boldsymbol{X}_{ti})$ as VA_{ti} for $t = 1, 2$. Then

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

because the σ-field generated by $\{VA_{1i}, \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1\}$ is equivalent to the σ-field generated by $\{\alpha_{1i}, \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1\}$. Now

$$\begin{aligned}
E(\alpha_{2i} \mid \alpha_{1i}, \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1) &= E\left[E(\alpha_{2i} \mid \alpha_{1i}, \boldsymbol{Y}_{1i}, \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1) \mid \alpha_{1i}, \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1\right] \\
&= E\left[\phi_{02} + \phi_{12}\alpha_{1i} + \gamma_2 \overline{Y}_{1i\bullet} \mid \alpha_{1i}, \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1\right] \\
&= \phi_{02} + \phi_{12}\alpha_{1i} + \frac{\gamma_2}{n_{1i}} \imath'_{n_{1i}} E(\boldsymbol{Y}_{1i} \mid \alpha_{1i}, \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1) \\
&= \phi_{02} + \phi_{12}\alpha_{1i} + \frac{\gamma_2}{n_{1i}} \imath'_{n_{1i}} (\boldsymbol{X}_{1i}\boldsymbol{\beta}_1 + \imath_{n_{1i}}\alpha_{1i}) \\
&= \phi_{02} + (\phi_{12} + \gamma_2)\alpha_{1i} + \gamma_2 \overline{\boldsymbol{X}}_{1i\bullet}\boldsymbol{\beta}_1.
\end{aligned}$$

It follows that

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

obtaining this (2.24); here we use the fact that

$$\boldsymbol{\beta}'_1 \overline{\boldsymbol{X}}'_{1i\bullet} = \left(\boldsymbol{\beta}'_1 \overline{\boldsymbol{X}}'_{1i\bullet}\right)' = \overline{\boldsymbol{X}}_{1i\bullet}\boldsymbol{\beta}_1$$

since $\boldsymbol{\beta}'_1 \overline{\boldsymbol{X}}'_{1i\bullet}$ is of dimension 1×1.

Equality (2.25) follows from the following arguments:

$$\begin{aligned}
Var\left[E(VA_{2i} \mid VA_{1i}, \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1) \mid \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1\right] &= Var\left[(\phi_{12} + \gamma_2)VA_{1i} \mid \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1\right] \\
&= Var\left[(\phi_{12} + \gamma_2)\alpha_{1i} \mid \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1\right] \\
&= \tau^2_1(\phi_{12} + \gamma_2)^2.
\end{aligned}$$

Similarly, (2.26) follows from the following argument:

$$\begin{aligned}
Var\left[VA_{2i} \mid \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1\right] &= Var\left[\alpha_{2i} \mid \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1\right] \\
&= \omega_{2i}.
\end{aligned}$$

Finally, (2.27) follows because $E\left[VA_{2i} \mid VA_{1i}, \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1\right]$ and $VA_{2i} - E\left[VA_{2i} \mid VA_{1i}, \boldsymbol{X}^2_{1,i}, \boldsymbol{\psi}^2_1\right]$ are always orthogonal: it is enough to recall that a conditional expectation is a orthogonal projection in the Hilbert space L^2, which underlies our approach.

References

Barra, J. R. (1981). *Mathematical Basis of Statistics*. Academic Press, New York, USA.

Basu, D. (1975). Statistical Information and Likelihood (with Discussion). *Sankhyā: The Indian Journal of Statistics*, *37*, 1–71.

Blackwell, D. (1956). On a class of probability spaces. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, *2*, 1-6.

Carlin, B., & Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis (2nd. Edition)*. Chapman & Hall/CRC, London.

Cox, D. R., & Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London, UK.

Eberly, L. E., & Carlin, B. P. (2000). Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Statistics in Medicine*, *19*(17-18), 2279–2294.

Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London, Series A*, *222*, 309–368.

Florens, J. P., Mouchart, M., & Rolin, J.-M. (1985). On two definitions of identification. *Statistics*, *16*, 213-218.

Florens, J. P., Mouchart, M., & Rolin, J.-M. (1990). *Elements of Bayesian Statistics*. New York: Marcel Dekker.

Kass, R. E., Carlin, B. P., Gelman, A., & Neal, R. M. (1998). Markov chain Monte Carlo in practice: a roundtable discussion. *The American Statistician*, *52*(2), 93–100.

Lindley, D. V. (1983). *Bayesian Statistics. A Review*. CBMS-NSF Regional Conference Series in Applied Mathematics, Philadelphia.

McCullagh, P. (2002). What is a statistical model? (with discussion). *The Annals of Statistics*, *30*, 1225-1310.

Mouchart, M. (1976). A note on Bayes theorem. *Statistica*, *36*, 349-357.

Poirier, D. J. (1998). Revising beliefs in nonidentified models. *Econometric theory*, *14*(4), 483–509.

Rao, M. M. (2005). *Conditional Measures and Applications* (Vol. 271). CRC Press.

Raoult, J. P. (1975). *Structures statistiques*. Presses Universitaires de France, Paris, France.

San Martín, E. (2018). Identifiability of structural characteristics: How relevant is it for the Bayesian approach? *Brazilian Journal of Probability and Statistics*, *32*(2), 346-373.

San Martın, E., & González, J. (2010). Bayesian identifiability: Contributions to an inconclusive debate. *Chilean Journal of Statistics*, *1*(2), 69–91.

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.

Wechsler, S., Izbicki, R., & Esteves, L. G. (2013). A Bayesian look at nonidentifiability: A simple example. *The American Statistician*, *67*(2), 90–93.

Xie, Y., & Carlin, B. P. (2006). Measures of Bayesian learning and identifiability in hierarchical models. *Journal of Statistical Planning and Inference*, *136*(10), 3458–3477.