Moore–Penrose Estimators of Age–Period–Cohort Effects: Their Interrelationship and Properties

Ethan Fosse,[a] Christopher Winship[b]

a) Princeton University; b) Harvard University

Abstract: The intrinsic estimator (IE) has become a widely used tool for the analysis of age–period–cohort (APC) data in sociology, demography, and other fields. However, it has been recently recognized that the IE is a subtype of a larger class of estimators based on the Moore–Penrose generalized inverse (MP estimators) and that different estimators can lead to radically divergent estimates of the true, unknown APC effects. To clarify the differences and similarities of MP estimators, we introduce a canonical form of the linear constraints imposed on the true temporal effects. Using this canonical form, we compare the IE to related MP estimators, examining the conditions under which they recover the true temporal effects, the impact of the size and sign of nonlinearities on the estimated linear effects, and their sensitivity to the number of age, period, and cohort categories. We show that two MP estimators, which we call the difference estimator (DE) and the orthogonal estimator (OE), impose constraints that are both less sensitive and easier to interpret than those of the IE. We conclude with practical guidelines for researchers interested in using MP estimators to estimate temporal effects.

Keywords: age–period–cohort analysis; intrinsic estimator; identification problem; Moore–Penrose generalized inverse

THE intrinsic estimator (IE) has become a popular technique for estimating age–period–cohort (APC) effects[1] across a wide range of fields. In recent years, researchers have used the IE to examine temporal trends in pornography use (Price et al. 2016), behavioral problems in adolescents (Keyes et al. 2017), heart disease mortality (Kramer, Valderrama, and Casper 2015), breast cancer mortality (Li, Yu, and Wang 2015), obesity in China (Fu and Land 2015), and social trust (Hu 2015), among other topics. Although popular, the IE is just one of a broader class of estimators based on the Moore–Penrose generalized inverse (MP estimators hereafter).[2] As recently recognized in the methodological literature (Land et al. 2016), MP estimators are defined by applying the Moore–Penrose generalized inverse to different design matrices, and the IE is that particular MP estimator that uses a design matrix of sum-to-zero effect (or deviation) coding.[3]

The fact that the IE is a subtype of a larger class of MP estimators raises a crucial question, especially in light of recent criticisms leveled against the IE (Luo 2013; O'Brien 2011, 2015): which MP estimator, if any, should researchers use when analyzing temporal trends? The desirable statistical properties typically used to justify the IE are of little guidance because these are shared by all MP estimators. That is, all MP estimators produce results that are estimable,[4] are unbiased given the linear constraint imposed by the estimator, and have minimum variance among estimators based on the same design matrix (Fu 2016; O'Brien 2015; Yang and Land 2013a). However, MP estimators are unalike in that they impose different linear

constraints on the true, unknown APC effects, thereby generating divergent sets of parameter estimates, as shown by two recent articles (Luo et al. 2016; Pelzer et al. 2014). Unfortunately, the *form* of the linear constraints also differs across MP estimators, making it exceedingly difficult, if not impossible, for researchers to compare the specific assumptions of these estimators, let alone ascertain whether or not such assumptions are reasonable in any given application. This is especially problematic because the typical goal of APC analysis is to recover the true, unknown temporal effects (Luo 2013: 1951–52).

To resolve this problem, we introduce a canonical form of the linear constraints imposed by the IE and related MP estimators on the true, unknown APC effects. Using this canonical form of the constraints, we compare various MP estimators, evaluating the conditions under which they recover the true temporal effects, the impact of the magnitude and direction of nonlinearities on their estimated linear effects, and their sensitivity to the number of age, period, and cohort categories. Importantly, we show that two MP estimators, which we call the orthogonal estimator (OE) and the difference estimator (DE), impose linear constraints that are both less sensitive and easier to interpret than those of the IE. However, we emphasize that there is, to our knowledge, no social, biological, or cultural theory that claims that the true, unknown linear effects should conform to the particular constraints imposed by MP estimators. In deriving the canonical constraints of MP estimators, we further prove that, mathematically, the APC identification problem is always restricted to the linear effects and that the solution line, which is a geometric representation of the identification problem, is always reducible to just three dimensions defined by the set of possible slopes. Accordingly, the IE and related MP estimators will, in general, produce the same set of nonlinear effects but different linear effects.

The rest of the article is organized as follows. First, we briefly review the APC identification problem and the Moore–Penrose generalized inverse. Second, we define the IE and discuss how all MP estimators are minimum-norm,[5] least-squares estimators. Third, we discuss the shared as well as divergent properties of MP estimators. Fourth, we illustrate how any APC design matrix can be separated into the linear and nonlinear components using a transformation matrix. We then prove that the APC identification problem is always restricted to the linear effects and that the solution line can always be simplified to three dimensions. Fifth, we introduce a canonical form of the linear constraints of the IE and related MP estimators, demonstrating that these estimators impose a linear constraint on the slopes but not the nonlinearities. Next, using the canonical form of the constraints, we compare the IE and related MP estimators mathematically, showing explicitly how the constraints can differ based on the size and direction of nonlinear effects, the choice of reference category, and the number of age, period, and cohort groups. Then, using simulated data, we evaluate the efficacy of several MP estimators, including the IE, in recovering the true temporal effects. Finally, we conclude with practical guidelines for applied researchers wishing to use the IE and related MP estimators to analyze APC data.

Generalized Inverses and the APC Identification Problem

To clarify the discussion that follows, suppose we have categorically coded age, period, and cohort data for a set of n respondents.[6] We let $i = 1, \ldots, I$ denote the unique age groups, $j = 1, \ldots, J$ the unique period groups, and $k = 1, \ldots, K$ the unique cohort groups with $k = j - i + I$ and $K = I + J - 1$.[7] We let n denote the number of respondents. The model we would like to run is

$$Y_{ijk} = \mu + \sum_{i=1}^{I}(\alpha_i)(\text{age}_i) + \sum_{j=1}^{J}(\pi_j)(\text{period}_j) + \sum_{k=1}^{K}(\gamma_k)(\text{cohort}_k) + \epsilon_{ijk}, \quad (1)$$

which we refer to as the classical APC (C-APC) model, also known as the multiple classification or accounting model. For simplicity, we can express the C-APC in matrix terms as

$$\mathbf{y} = \mathbf{Xb} + \epsilon, \quad (2)$$

where $\mathbf{y}$ is an $n \times 1$ outcome vector, which without loss of generality, we assume to be continuous; $\mathbf{X}$ is a design matrix of categorical age, period, and cohort variables with dimension $n \times p$; $\mathbf{b}$ is a $p \times 1$ parameter vector with elements corresponding to the age, period, and cohort groups; and ϵ is an $n \times 1$ vector of random errors.[8] If there were no linear dependence in $\mathbf{X}$, then we could obtain a unique least-squares solution

$$\mathbf{b}_{\text{OLS}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}, \quad (3)$$

where the superscripted -1 indicates the regular inverse. However, due to linear dependence, $\mathbf{X}$ is rank deficient one, and a regular inverse of $\mathbf{X}^T\mathbf{X}$ does not exist. Thus, we cannot estimate $\mathbf{b}_{\text{OLS}}$, and any particular least-squares solution requires an additional constraint.

Using what is known as a generalized inverse, it is possible to produce constrained estimates of the parameters $\mathbf{b}$ that are consistent with the data (e.g., see O'Brien 2015: 27–29). Unfortunately, there are an infinite number of generalized inverses, and in general, different inverses will produce different sets of estimates. Each of the constrained estimates for a particular design matrix lie on what is called the *solution line* of estimates in multidimensional space (O'Brien 2015: 27–28). The solution line is a geometric representation of the identification problem, reflecting the lack of a unique set of estimates. To construct the solution line, we let $\mathbf{b}^*$ denote any specific constrained least-squares solution to the least-squares normal equations. For any particular constraint, we can construct a generalized inverse of $\mathbf{X}^T\mathbf{X}$ to find a corresponding solution

$$\mathbf{b}^* = (\mathbf{X}^T\mathbf{X})^*\mathbf{X}^T\mathbf{y}, \quad (4)$$

where the superscript $*$ denotes the appropriate generalized inverse. The vector $\mathbf{b}^*$ is *a* least-squares solution to the normal equations such that $\mathbf{X}^T\mathbf{b}^* = \mathbf{X}^T\mathbf{y}$. We can then write[9]

$$\mathbf{b} = \mathbf{b}^* + s\mathbf{v}, \quad (5)$$

where s is an arbitrary scalar that can take on any real number and $\mathbf{v}$ is the eigenvector with the zero eigenvalue, or *null vector*, of a given design matrix. The null vector gives the form of the linear dependency for a particular design matrix and is unique up to a scalar (O'Brien 2015: 30–32,56–57). By varying s, we trace out the solution line for a given design matrix, resulting in an infinite number of constrained least-squares solutions.

Among the infinite number of generalized inverses for a particular design matrix, what is known as the Moore–Penrose generalized inverse has received significant attention due to its unique properties (Ben-Israel 2002). Using the Moore–Penrose generalized inverse gives us, parallel to the traditional ordinary least-squares (OLS) formula

$$\mathbf{b}_{\mathrm{MP}} = (\mathbf{X}^T\mathbf{X})^+\mathbf{X}^T\mathbf{y} = \mathbf{X}^+\mathbf{y}, \tag{6}$$

where the superscript $+$ denotes the Moore–Penrose generalized inverse and, again, $\mathbf{X}$ is a design matrix of categorical age, period, and cohort variables. Equation 6 underscores that the MP estimator is based on $(\mathbf{X}^T\mathbf{X})^+$ or, equivalently, $\mathbf{X}^+$. Formally, $\mathbf{X}^+$ is defined as a generalized inverse meeting four Moore–Penrose conditions:

1. General condition: $\mathbf{X}\mathbf{X}^+\mathbf{X} = \mathbf{X}$
2. Reflexive condition: $\mathbf{X}^+\mathbf{X}\mathbf{X}^+ = \mathbf{X}^+$
3. Normalized condition: $(\mathbf{X}\mathbf{X}^+)^T = \mathbf{X}\mathbf{X}^+$
4. Reverse normalized condition: $(\mathbf{X}^+\mathbf{X})^T = \mathbf{X}^+\mathbf{X}$.

These conditions specify that for any particular matrix $\mathbf{X}$, the Moore–Penrose generalized inverse always exists (i.e., it is well defined) and is unique (i.e., there is only one such generalized inverse that meets these conditions). From the four conditions above, we can express the solution in terms of the normal equations, and it can be shown that if $\mathbf{X}$ is of full rank, then $\mathbf{X}^+\mathbf{y} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$.[10] In the following section, we discuss how to interpret the MP solution in the case of APC data, which is rank deficient one.

Deriving and Defining MP Estimators

As mentioned previously, the application of the Moore–Penrose generalized inverse to different design matrices or coding schemes defines a broad class of constrained estimators, which we term MP estimators. As long as the design matrix represents the full range of possible APC categories, various coding schemes may be used, such as treatment coding with the first category omitted or sum-to-zero effect coding with the last category omitted.[11] The IE is defined as that particular MP estimator based on a design matrix of sum-to-zero effect (or deviation) contrasts (Land et al. 2016: 964). However, it is important to clarify that the phrase "intrinsic estimator" has been used in at least two different ways by the proponents of the IE (e.g., Fu 2000, 2008, 2016; Land et al. 2016; Yang and Land 2013a: 79). In particular, in his sole-authored work, Fu defines the IE as a general class of estimators derived from the Moore–Penrose generalized inverse without reference to the specific structure of the design matrix (Fu 2000; Fu 2008: 332–3; Fu 2016). In contrast, Land and colleagues

(Land et al. 2016: 964–71) state explicitly that "only the sum-to-zero [effect] coding is used to define and estimate the IE." Furthermore, they assert that because there are any number of design matrices that can produce different sets of estimates, there are "infinitely many possible pseudo-IE estimators." To eliminate this ambiguity, we will refer to Fu's broad class of estimators simply as MP estimators and the narrower set of MP estimators with sum-to-zero effect coding as the IE, as preferred by Land and colleagues. We now turn to how to derive and define MP estimators formally.

Deriving MP Estimators

There are many ways to derive the estimates of an MP estimator (e.g., see Yang and Land 2013a: 79–80). An especially intuitive approach, which has not received significant attention in the APC literature, is to find the Moore–Penrose generalized inverse of $\mathbf{X}^T\mathbf{X}$ using a decomposition technique. For a given design matrix $\mathbf{X}$, we can write

$$\mathbf{X}^T\mathbf{X} = V\Lambda V^T, \tag{7}$$

where $V\Lambda V^T$ is the spectral (or eigenvalue) decomposition of $\mathbf{X}^T\mathbf{X}$. The $p \times p$ diagonal matrix Λ consists of the eigenvalues of $\mathbf{X}^T\mathbf{X}$ in descending order $\lambda_1, \lambda_2, \ldots, \lambda_{r-1}$, with the rank of the matrix $r = p - 1$. It is straightforward to find the Moore–Penrose generalized inverse of $\mathbf{X}^T\mathbf{X}$ using this decomposition. First, we take the reciprocal of the nonzero eigenvalues along the diagonal, keeping the zero eigenvalues. This will give the generalized inverse of Λ, denoted as Λ^+. Second, we calculate $V\Lambda^+V^T = (\mathbf{X}^T\mathbf{X})^+$. Finally, we use $(\mathbf{X}^T\mathbf{X})^+$ to find the Moore–Penrose generalized inverse estimates $\mathbf{b}^+ = (\mathbf{X}^T\mathbf{X})^+\mathbf{X}^T\mathbf{y}$.

In addition to allowing for the estimation of MP estimators, the decomposition of $\mathbf{X}^T\mathbf{X}$ reveals two crucial features of the data. First, the number of nonzero eigenvalues in Λ gives the rank of the matrix $\mathbf{X}^T\mathbf{X}$. Because the columns of the data are linearly dependent, there will always be one zero eigenvalue, and thus the design matrix for any APC data set is always rank deficient one. Second, along with the zero eigenvalue in Λ, there will always be a corresponding eigenvector in Λ, which is the orthonormal basis for the null space of $\mathbf{X}^T\mathbf{X}$. This is the null vector, or the eigenvector with a zero eigenvalue, which encodes the linear dependency in the data.[12] In the case of APC data, different design matrices have different forms of linear dependencies and thus different null vectors. As we discuss later, converting the null vector of an MP estimator into a canonical form is crucial for understanding and comparing the assumptions it requires about the true, unknown temporal effects.

Defining MP Estimators

An MP estimator can be defined formally as the solution orthogonal (or perpendicular) to the null vector of a particular design matrix (O'Brien 2015: 30–32). To understand this definition, note that like any constrained estimator, we can use an MP estimator to construct a solution line for a given design matrix such that $\mathbf{b} = \mathbf{b}_{\text{MP}} + s\mathbf{v}$. Among the values of s, an MP estimator assumes $s = 0$ in the

equation $\mathbf{b} = \mathbf{b}_{\text{MP}} + s\mathbf{v}$ (Fu 2016: 183). Geometrically, an MP estimator is the least-squares solution corresponding to the point on the line closest to the origin in terms of Euclidean distance. This point coincides with the minimium (Euclidean) length of $\mathbf{b}_{\text{MP}} + s\mathbf{v}$, which is at its minimum when $s = 0$.[13] Equivalently, the vector $\mathbf{b}_{\text{MP}}$ is the projection of $\mathbf{b}$ on the nonnull space of $\mathbf{X}$, which is orthogonal to the null space (Yang and Land 2013a: 81)

$$\mathbf{b}_{\text{MP}} = (\mathbf{I} - \widehat{\mathbf{v}}\widehat{\mathbf{v}}^T)\mathbf{b}, \tag{8}$$

where $\mathbf{I}$ is a $p \times p$ identity matrix and $\widehat{\mathbf{v}}$ is the normalized null vector so that $\widehat{\mathbf{v}}^T\widehat{\mathbf{v}} = 1$.[14] That is, an MP estimator separates the true, unknown parameter $\mathbf{b}$ into two orthogonal components, the null vector $\mathbf{v}$ and the vector of MP estimates $\mathbf{b}_{\text{MP}}$. Because these two vectors are perpendicular to each other, $\mathbf{v}^T\mathbf{b}_{\text{MP}} = 0$ or, equivalently, $s = 0$ in the solution line defined by $\mathbf{b}_{\text{MP}} + s\mathbf{v}$.

Because an MP estimator finds that particular set of least-squares estimates on the solution line for which the (Euclidean) length is minimized, it is a *minimum-norm, least-squares* (MNLS) solution (Ben-Israel 2002: 109). Although it has not been explicitly stated as such in the APC literature, any MP estimator is fundamentally a two-stage minimization algorithm:

Stage 1:   Find those values of $\mathbf{b}$ that minimize $\|\mathbf{X}\mathbf{b} - \mathbf{y}\|_2$ (9)

Stage 2:   Minimize $\|\mathbf{b}\|_2$ among all solutions from Stage 1,   (10)

where the operator $\|\,.\,\|_2$ denotes the L_2 (or Euclidean) norm and $\mathbf{X}$ is a given design matrix of age, period, and cohort variables. In the first stage, the MP estimator finds the least-squares solution to $\mathbf{X}\mathbf{b} = \mathbf{y}$. For full rank data, there is a unique solution, and the algorithm ends. However, in the case of APC data, the design matrix $\mathbf{X}$ is rank deficient one, and thus there is no unique least-squares solution; rather, as described previously, there are many such solutions lying on a line in multidimensional space. In the second stage, the MP estimator selects a solution by applying the minimum-norm constraint. Among the least-squares estimates on the solution line, the MP estimator selects that particular set of estimates with the minimum (Euclidean) length or, equivalently, that is closest to the origin in terms of Euclidean distance.[15]

Properties of MP Estimators

We now turn to an examination of the shared as well as divergent properties of MP estimators, including the IE. We focus here on those shared properties that are considered desirable in the methodological literature (Fu 2000; Fu 2016; Fu and Hall 2006; Fu, Land, and Yang 2011; Yang et al. 2008; Yang, Fu, and Land 2004): first, an MP estimator has minimum sampling variance among all possible estimators based on its specific design matrix; second, it is an estimable function, meaning that it produces a unique set of estimates for the effects of age, period and cohort; finally, it is unbiased, meaning that the average of any estimates produced by an MP estimator over an infinite number of simple random samples will be equal to that estimator's values when it is applied to the full population data. However,

MP estimators diverge in a critically important way: because they are based on different design matrices, they will have different null vectors and in turn impose different linear constraints on the true, unknown temporal effects (Luo et al. 2016; Pelzer et al. 2014).

Shared Desirable Statistical Properties

There are several desirable statistical properties shared by all MP estimators. First, for any given design matrix, an MP estimator will be that estimator with the minimum sampling variance (Yang et al. 2004: 102–3,108; Yang et al. 2008: 1709; Yang and Land 2013a: 86,116–7). This is a function of the fact that any MP estimator will give that set of estimates that is shortest in Eucldean norm. Consequently, if we were to use another type of generalized inverse other than the Moore–Penrose for some fixed design matrix, the resulting estimates would have greater sampling variance.

Second, an MP estimator is always an estimable function (Fu et al. 2011: 456–8; Yang and Land 2013a: 84–85). To state that a function is estimable means that when applied to data, it produces a unique set of estimates. Intuitively, it means that it is possible for the data to tell us what the function equals. For example, if we have variables that are linearly dependent, then the standard OLS function is not estimable because it is based on a regular inverse that is not well defined (i.e., it does not exist) due to the linear dependence in the design matrix. In contrast, MP estimators are estimable because the Moore–Penrose generalized inverse is well defined (i.e., it exists) even when the variables are linearly dependent. Specifically, an MP estimator applies a particular mathematical constraint on the estimates that will only be satisfied by a single point on the solution line. As stated previously, the estimates will be that set under the Euclidean distance metric (or, equivalently, correspond with the minimum L_2 norm). The MP estimates are estimable in the very specific sense that for a particular design matrix and outcome, there is a set of points closest in Euclidean distance to the origin on the solution line.[16]

Finally, all MP estimators are unbiased, meaning that the average of any estimates produced by an MP estimator over an infinite number of simple random samples will equal that estimator's values when it is applied to the full population data (Yang et al. 2004: 101–2,107; Yang et al. 2008: 1709; Yang and Land 2013a: 86,115–6). A function is unbiased if, when it is calculated for an infinite number of random samples, its average is equal to its value when it is calculated on the population as a whole.[17] In the context of any specific MP estimator, this means that if we had an infinite number of samples and applied this estimator in each sample, then the average of the estimates across these samples would equal the estimates produced by using this same estimator on the whole population.

The Linear Constraint on the True Temporal Effects

The foregoing underscores that the desirable statistical properties typically used to justify the IE are not unique to it but shared by all MP estimators.[18] To emphasize, using various alternative design matrices with the Moore–Penrose generalized inverse will also produce estimates, like the IE, that are estimable, unbiased, and

have minimum variance relative to other estimators based on that design matrix. However, results from MP estimators differ, sometimes radically so, due to the structure of the design matrix.

To recognize why this is the case, it is crucial to understand that the null vector encodes the linear dependency among the columns of a design matrix. For any given design matrix, at least one of the columns of $\mathbf{X}$ can be rewritten as a linear combination of the other columns.[19] Formally, there is a nontrivial linear combination of the columns of $\mathbf{X}$ that results in a vector of zeros such that $\mathbf{Xv} = \mathbf{0}$, where $\mathbf{v}$ is the $p \times 1$ null vector and $\mathbf{0}$ is an $n \times 1$ vector of zeros. As mentioned previously, the null vector $\mathbf{v}$ represents the null space of $\mathbf{X}$ and is unique up to multiplication by an arbitrary scalar s. Accordingly, the equation $\mathbf{Xv} = \mathbf{0}$ generalizes to $\mathbf{X}s\mathbf{v} = \mathbf{0}$.

The main assumption of any MP estimator is that the true, unknown temporal effects conform to the linear dependency of the particular design matrix on which it is based. To show this, note that the linear constraint imposed by an MP estimator will yield the true parameter of APC effects $\mathbf{b}$ only if $s = 0$ in the equation $\mathbf{b} = \mathbf{b}_{\mathrm{MP}} + s\mathbf{v}$ (see Land et al. 2016: 966; Yang and Land 2013a: 82), so

$$
\begin{aligned}
\mathbf{b}_{\mathrm{MP}} &= \mathbf{b}, \quad \text{when } s = 0 \text{ and} \\
\mathbf{b}_{\mathrm{MP}} &\neq \mathbf{b}, \quad \text{when } s \neq 0.
\end{aligned}
\tag{11}
$$

Rearranging $\mathbf{b} = \mathbf{b}_{\mathrm{MP}} + s\mathbf{v}$ and taking the expectation, we know that $\mathbb{E}(\mathbf{b}_{\mathrm{MP}}) = \mathbf{b} - s\mathbf{v}$. Thus, the expected value of the MP estimate will not equal the true, unknown temporal effects unless $s = 0$. Assuming $s = 0$ is equivalent to assuming that $\mathbf{v}^T\mathbf{b} = 0$ because if $s = 0$, then $\mathbf{b} = \mathbf{b}_{\mathrm{MP}} + s\mathbf{v}$ becomes $\mathbf{b}_{\mathrm{MP}} = \mathbf{b}$ (see Luo 2013: 1951–3). Hence, $\mathbf{v}^T\mathbf{b}_{\mathrm{MP}} = \mathbf{v}^T\mathbf{b} = 0$. In other words, any MP estimator assumes that the true, unknown temporal effects conform to the linear dependency of its particular design matrix.

Unfortunately, the linear constraints imposed by MP estimators have divergent forms, precluding substantive interpretation or comparison across estimators. To illustrate the differing linear constraints, in Table 1, we show the linear constraints of nine MP estimators applied to data with $I = 3$ age groups and $J = 3$ period groups. We focus here on MP estimators based on contrast coding schemes commonly discussed in the applied statistics literature (e.g., Fox 2002: 126–30; Onyiah 2008: 121–36,148–61; Venables and Ripley 2013: 146–9) as well as those employed in the methodological literature on APC effects in sociology and demography (e.g., Luo et al. 2016; O'Brien 2015: 55–57; Pelzer et al. 2014). We present the parameters in detailed form in Table 1 so as to clarify their substantive interpretation. Specifically, for some parameter β, we let β_k denote the average of the kth category, $\bar{\beta}$ the average of all $k = 1, \ldots k = K$ levels, $\beta_{i:j}$ the average of levels i through j, and β^k the kth orthogonal polynomial term. For example, γ_2 is the average of the second cohort category; $\bar{\gamma}$ is the average across all cohort categories; $\gamma_{2:4}$ is the average of cohort categories two, three, and four; and γ^2 is the second-order orthogonal polynomial term for the cohort variable.

As shown in Table 1, we examine the linear constraints of the following estimators: (1) IE_{First}, or the IE, which uses sum-to-zero effect coding with the first category of each variable omitted (e.g., $\alpha_2 - \bar{\alpha}$ compares the average of the second

Table 1: Linear constraints of MP estimators with $I = 3$ and $J = 3$ groups.

MP Estimator	Linear Constraint	
IE_{First}:	$(0)\mu + (0)(\alpha_2 - \bar{\alpha}) + (1)(\alpha_3 - \bar{\alpha}) + (0)(\pi_2 - \bar{\pi}) + (-1)(\pi_3 - \bar{\pi}) + (-1)(\gamma_2 - \bar{\gamma})$	
	$+(0)(\gamma_3 - \bar{\gamma}) + (1)(\gamma_4 - \bar{\gamma}) + (2)(\gamma_5 - \bar{\gamma})$	$= \quad 0$
IE_{Last}:	$(0)\mu + (1)(\alpha_1 - \bar{\alpha}) + (0)(\alpha_2 - \bar{\alpha}) + (-1)(\pi_1 - \bar{\pi}) + (0)(\pi_2 - \bar{\pi}) + (2)(\gamma_1 - \bar{\gamma})$	
	$+(1)(\gamma_2 - \bar{\gamma}) + (0)(\gamma_3 - \bar{\gamma}) + (-1)(\gamma_4 - \bar{\gamma})$	$= \quad 0$
TE_{First}:	$(0)\mu + (1)(\alpha_2 - \alpha_1) + (2)(\alpha_3 - \alpha_1) + (-1)(\pi_2 - \pi_1) + (-2)(\pi_3 - \pi_1) + (1)(\gamma_2 - \gamma_1)$	
	$+(2)(\gamma_3 - \gamma_1) + (3)(\gamma_4 - \gamma_1) + (4)(\gamma_5 - \gamma_1)$	$= \quad 0$
TE_{Last}:	$(0)\mu + (2)(\alpha_1 - \alpha_3) + (1)(\alpha_2 - \alpha_3) + (-2)(\pi_1 - \pi_3) + (-1)(\pi_2 - \pi_3) + (4)(\gamma_1 - \gamma_5)$	
	$+(3)(\gamma_2 - \gamma_5) + (2)(\gamma_3 - \gamma_5) + (1)(\gamma_4 - \gamma_5)$	$= \quad 0$
$\text{HE}_{\text{Backward}}$:	$(0)\mu + (2)(\alpha_2 - \alpha_1) + (3)(\alpha_3 - \alpha_{1:2}) + (-2)(\pi_2 - \pi_1) + (-3)(\pi_3 - \pi_{2:3}) + (2)(\gamma_2 - \gamma_1)$	
	$+(3)(\gamma_3 - \gamma_{1:2}) + (4)(\gamma_4 - \gamma_{1:3}) + (5)(\gamma_5 - \gamma_{1:4})$	$= \quad 0$
$\text{HE}_{\text{Forward}}$:	$(0)\mu + (3)(\alpha_1 - \alpha_{2:3}) + (2)(\alpha_2 - \alpha_3) + (-3)(\pi_1 - \pi_{2:3}) + (-2)(\pi_2 - \pi_3) + (5)(\gamma_1 - \gamma_{2:5})$	
	$+(4)(\gamma_2 - \gamma_{3:5}) + (3)(\gamma_3 - \gamma_{4:5}) + (2)(\gamma_4 - \gamma_5)$	$= \quad 0$
$\text{DE}_{\text{Backward}}$:	$(0)\mu + (1)(\alpha_2 - \alpha_1) + (1)(\alpha_3 - \alpha_2) + (-1)(\pi_2 - \pi_1) + (-1)(\pi_3 - \pi_2) + (1)(\gamma_2 - \gamma_1)$	
	$+(1)(\gamma_3 - \gamma_2) + (1)(\gamma_4 - \gamma_3) + (1)(\gamma_5 - \gamma_4)$	$= \quad 0$
$\text{DE}_{\text{Forward}}$:	$(0)\mu + (1)(\alpha_1 - \alpha_2) + (1)(\alpha_2 - \alpha_3) + (-1)(\pi_1 - \pi_2) + (-1)(\pi_2 - \pi_3) + (1)(\gamma_1 - \gamma_2)$	
	$+(1)(\gamma_2 - \gamma_3) + (1)(\gamma_3 - \gamma_4) + (1)(\gamma_4 - \gamma_5)$	$= \quad 0$
OE:	$(0)\mu + (1)(\alpha) + (0)(\alpha^2) + (-1)(\pi) + (0)(\pi^2) + (1)(\gamma)$	
	$+(0)(\gamma^2) + (0)(\gamma^3) + (0)(\gamma^4)$	$= \quad 0$

Notes: Based on data with $I = 3$ age, $J = 3$ period, and $K = I + J - 1 = 5$ cohort groups.
Linear constraints are calculated using $\mathbf{v}^T\mathbf{b} = 0$, where $\mathbf{b}$ are the true, unknown temporal effects for a given design matrix. All MP estimators assume that $\mathbf{v}^T\mathbf{b}_{\text{MP}} = \mathbf{v}^T\mathbf{b} = 0$. For some parameter β, we let β_k denote the average of the kth category, $\bar{\beta}$ the average of all $k = 1, \ldots k = K$ levels, $\beta_{i:j}$ the average of levels i through j, and β^k the kth orthogonal polynomial term. For example, γ_2 is the average of the second cohort category; $\bar{\gamma}$ is the average across all cohort categories; $\gamma_{2:4}$ is the average of cohort categories two, three, and four; and γ^2 is the second-order orthogonal polynomial term for the cohort variable.

age category with the average of all age categories); (2) IE_{Last}, also the IE, which uses sum-to-zero effect coding with the last category omitted for each variable (e.g., $\alpha_1 - \bar{\alpha}$ compares the average of the first age category with the average of all age categories); (3) TE_{First}, or the treatment estimator (TE), which uses treatment contrasts with the first category of each variable as a reference (e.g., $\alpha_3 - \alpha_1$ compares

the average of the third age category with the average of the first age category); (4) TE_{Last}, also the TE, which uses treatment contrasts with the last category of each variable as a reference (e.g., $\alpha_1 - \alpha_3$ compares the average of the first age category with the average of the last age category); (5) $\text{HE}_{\text{Backward}}$, or the helmert estimator (HE), which uses backward helmert contrasts with the first category of each variable omitted (e.g., $\alpha_3 - \alpha_{1:2}$ compares the average of the last age category with the average of the two preceding age categories); (6) $\text{HE}_{\text{Forward}}$, also the HE, which uses forward helmert contrasts with the last category of each variable omitted (e.g., $\alpha_1 - \alpha_{2:3}$ compares the average of the first age category with the average of the two subsequent age categories); (7) $\text{DE}_{\text{Backward}}$, or the difference estimator (DE), which uses backward successive differences with the first category of each variable omitted (e.g., $\alpha_2 - \alpha_1$ and $\alpha_3 - \alpha_2$ compare the averages of adjacent age categories); (8) $\text{DE}_{\text{Forward}}$, also the DE, which uses forward successive differences with the last category of each variable omitted (e.g., $\alpha_1 - \alpha_2$ and $\alpha_2 - \alpha_3$ compare the averages of adjacent age categories); and (9) OE, or the orthogonal estimator (OE), which uses sum-to-zero orthogonal polynomial contrasts or, more generally, sum-to-zero effect coding with the linear and nonlinear components orthogonalized (e.g., α is the age linear effect, α^2 is the quadratic age effect, and α^3 is the cubic age effect).

Table 1 reveals that MP estimators differ greatly in terms of the linear constraints they impose on the true temporal effects. The constraint $\mathbf{v}^T \mathbf{b} = 0$ will differ across MP estimators because the parameter vector $\mathbf{b}$ as well as the null vector $\mathbf{v}$ both depend on the structure of the design matrix. As shown in Table 1, not only is it exceedingly difficult to interpret the linear constraint of any particular MP estimator but it is seemingly impossible to compare the constraints across various MP estimators. For instance, as indicated in Table 1, with $I = 3$ age groups, $J = 3$ period groups, and $I + J - 1 = K = 5$ cohort groups, the IE_{Last} assumes that $\mathbf{v}^T \mathbf{b} = (0)\mu + (1)(\alpha_1 - \bar{\alpha}) + (0)(\alpha_2 - \bar{\alpha}) + (-1)(\pi_1 - \bar{\pi}) + (0)(\pi_2 - \bar{\pi}) + (2)(\gamma_1 - \bar{\gamma}) + (1)(\gamma_2 - \bar{\gamma}) + (0)(\gamma_3 - \bar{\gamma}) + (-1)(\gamma_4 - \bar{\gamma}) = 0$. In contrast, when used on a data set with the identical number of age, period, and cohort groups, the TE_{Last} assumes that $\mathbf{v}^T \mathbf{b} = (0)\mu + (2)(\alpha_1 - \alpha_3) + (1)(\alpha_2 - \alpha_3) + (-2)(\pi_1 - \pi_3) + (-1)(\pi_2 - \pi_3) + (4)(\gamma_1 - \gamma_5) + (3)(\gamma_2 - \gamma_5) + (2)(\gamma_3 - \gamma_5) + (1)(\gamma_4 - \gamma_5) = 0$. Due to the disparate forms of these constraints, it is not at all clear how the IE_{Last} and the TE_{Last} are similar and how they are different, let alone how they compare to any of the other MP estimators in Table 1. Fortunately, in the following sections, we develop a canonical form of the linear constraints that facilitates the systematic interpretation and comparison of MP estimators.

The Transformation Matrix

An MP estimator consists of a particular design matrix in conjunction with the Moore–Penrose generalized inverse. Given that there are many such design matrices, potentially an infinite number, a technique is needed for comparing the results of different MP estimators. In this section, we show how, using a special transformation matrix, any APC design matrix can be converted into a canonical form that separates the linear from the nonlinear components. Using this transformation matrix, we then show that the APC identification problem is always restricted to

the linear effects, that any particular parameter can be decomposed into linear and nonlinear components, and that the solution line can always be simplified to three dimensions.

Constructing the Transformation Matrix

If we have two design matrices, $\mathbf{X}$ and $\mathbf{X}^*$, each of which provides a full representation of the age, period, and cohort effects and as such are deficient rank one, then there will always be an invertible $p \times p$ transformation matrix $\mathbf{T}$ of full rank such that (Luo et al. 2016: 947)

$$\mathbf{XT} = \mathbf{X}^* \quad \text{and} \quad \mathbf{X} = \mathbf{X}^* \mathbf{T}^{-1}. \tag{12}$$

However, given a set of MP estimators defined by different design matrices along with their corresponding estimates, we need to decide on a common design matrix $\mathbf{X}^*$ as a basis for comparison. We will let $\mathbf{X}_O$ denote the canonical form of any APC design matrix. The corresponding parameter $\mathbf{b}_O = (\mu, \alpha, \alpha^2 \ldots \alpha^{I-1}, \pi, \pi^2 \ldots \alpha^{J-1}, \gamma, \gamma^2 \ldots \gamma^{K-1})$ represents the full set of orthogonally separated linear and nonlinear components for age, period, and cohort. The null vector of $\mathbf{X}_O$ has the simple representation

$$\mathbf{v}_O = (0, 1, -1, 1, 0 \ldots 0), \tag{13}$$

where the first zero corresponds to the intercept; the elements one, negative one, and one correspond to the age, period, and cohort linear components; and the remaining zeros correspond to the $(I-2) + (J-2) + (K-2)$ nonlinear components. That is, the null vector $\mathbf{v}_O$ encodes the fundamental linear relationship underlying all APC data, namely that a person's age minus their year of measurement plus their birth year equals zero.

To convert any APC data set into canonical form, we will construct a special transformation matrix $\mathbf{T}$ (cf. Luo et al. 2016: 947–52). Let $\mathbf{A}$, $\mathbf{P}$, and $\mathbf{C}$ denote the original, untransformed contrast matrices for age, period, and cohort terms, with dimensions $I \times (I-1)$, $J \times (J-1)$, and $K \times (K-1)$, respectively. Although sum-to-zero effect (or deviation) coding is the most frequently used in the APC literature, these matrices may be coded with any number of schemes without loss of generality. Each of these contrast matrices has full column rank but not full row rank. Hence, we can construct three left inverses

$$\begin{aligned}
\mathbf{A}^L &= (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T, \\
\mathbf{P}^L &= (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T, \\
\mathbf{C}^L &= (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T,
\end{aligned} \tag{14}$$

where the superscript L denotes a left inverse. Let $\mathbf{A}_O$, $\mathbf{P}_O$, and $\mathbf{C}_O$ denote corresponding contrast matrices for age, period, and cohort, in which the linear and nonlinear components are orthogonal to each other (Draper and Smith 2014: 461–72). We are now in position to construct the transformation matrix $\mathbf{T}$, which is a block diagonal matrix in which the main diagonal blocks are square matrices and off-diagonal blocks are zero matrices. The transformation matrix $\mathbf{T}$ has the generic

form

$$\mathbf{T} = 1 \oplus \mathbf{A}^L \mathbf{A}_O \oplus \mathbf{P}^L \mathbf{P}_O \oplus \mathbf{C}^L \mathbf{C}_O \tag{15}$$

or, equivalently,

$$\mathbf{T} = \begin{pmatrix} 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^L \mathbf{A}_O & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P}^L \mathbf{P}_O & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{C}^L \mathbf{C}_O \end{pmatrix} \tag{16}$$

where $\oplus$ is the direct sum. As we demonstrate in the next section, using $\mathbf{T}$, we can convert any design matrix and parameter vector of an APC data set into canonical form.

Critical Insights Using the Transformation Matrix

Using the transformation matrix, we can make several important insights that will greatly simplify the interpretation and comparison of APC estimators. First, we can prove that the results of any constrained estimator, including any MP estimator, will differ only in terms of their estimated linear effects. Specifically, using $\mathbf{T}$, we can convert any design matrix $\mathbf{X}$ as well as its corresponding parameter vector $\mathbf{b}$ into canonical form (cf. Luo et al. 2016: 947)

$$\mathbf{Xb} = (\mathbf{XT})(\mathbf{T}^{-1}\mathbf{b}) = \mathbf{X}_O \mathbf{b}_O, \tag{17}$$

where again $\mathbf{X}_O$ is a design matrix of orthogonal linear and nonlinear components and $\mathbf{b}_O$ is a vector of the new, transformed estimates expressed in terms of linear and nonlinear effects. Because the null vector of $\mathbf{X}_O$ consists of nonzero elements only for the age, period, and cohort linear effects,[20] and we can convert any design matrix into the canonical form $\mathbf{X}_O$, this proves that the APC identification problem is restricted to the linear effects and that the intercept and nonlinear effects of any constrained APC estimator are identified.[21]

Second, using $\mathbf{T}$, we can decompose any particular parameter into its constituent linear and nonlinear components. To decompose a parameter vector, we use the equation $\mathbf{b} = \mathbf{T}\mathbf{b}_O$, where $\mathbf{b}$ is the original set of temporal effects, $\mathbf{T}$ is the transformation matrix defined above, and $\mathbf{b}_O$ are the effects expressed in canonical form. In Table 2, we show how to use the transformation matrix to decompose sum-to-zero effects (with the last category omitted) into a set of linear and nonlinear parameters. For example, in the case of $I = 3$ age groups, the first age parameter with sum-to-zero effect coding represents $\alpha_1 - \bar{\alpha}$, or the difference between the average of the first age category and the average across all age categories. This particular parameter is mathematically equal to a weighted sum of the age slope and the quadratic age effect: $(\alpha_1 - \bar{\alpha}) = (-1)\alpha + (1)\alpha^2$. As we show later, this decomposition allows us to express the linear constraints of MP estimators in a canonical form.

Finally, using the transformation matrix, we can derive a canonical form of the solution line. Researchers have noted that any particular design matrix defines a multidimensional solution line in parameter space (Luo et al. 2016; O'Brien 2011; O'Brien 2015: 59–91). However, it has not been fully appreciated that the solution line for any constrained estimator, including the MP estimator, can be greatly

Table 2: Decomposition of sum-to-zero effects using the transformation matrix.

Sum-to-Zero Effect		Decomposed Effect		
		Linear Component		Nonlinear Component
$(\alpha_1 - \bar{\alpha})$	$=$	$(-1)\alpha$	$+$	$(1)\alpha^2$
$(\alpha_2 - \bar{\alpha})$	$=$	$(0)\alpha$	$+$	$(-2)\alpha^2$
$(\pi_1 - \bar{\pi})$	$=$	$(-1)\pi$	$+$	$(1)\pi^2$
$(\pi_2 - \bar{\pi})$	$=$	$(0)\pi$	$+$	$(-2)\pi^2$
$(\gamma_1 - \bar{\gamma})$	$=$	$(-2)\gamma$	$+$	$(2)\gamma^2 + (-1)\gamma^3 + (1)\gamma^4$
$(\gamma_2 - \bar{\gamma})$	$=$	$(-1)\gamma$	$+$	$(-1)\gamma^2 + (2)\gamma^3 + (-4)\gamma^4$
$(\gamma_3 - \bar{\gamma})$	$=$	$(0)\gamma$	$+$	$(-2)\gamma^2 + (0)\gamma^3 + (6)\gamma^4$
$(\gamma_4 - \bar{\gamma})$	$=$	$(1)\gamma$	$+$	$(-1)\gamma^2 + (-2)\gamma^3 + (-4)\gamma^4$

Notes: Decomposition based on $I = 3$ age, $J = 3$ period, and $K = I + J - 1 = 5$ cohort groups. Inverse of the transformation is used to decompose original sum-to-zero effect coding into linear and nonlinear components such that $\mathbf{b}_{\text{Effect}} = \mathbf{T}\mathbf{b}_O$. Last category is omitted for each variable. For some parameter β, we let β_k denote the average of the kth category, $\bar{\beta}$ the average of all $k = 1, \ldots k = K$ levels, and β^k the kth orthogonal polynomial term. For example, γ_2 is the average of the second cohort category, $\bar{\gamma}$ is the average across all cohort categories, and γ^2 is the second-order orthogonal polynomial term for the cohort variable.

simplified. Using the transformation matrix outlined previously, we can make an important generalization: all constrained APC estimators lie on the same simplified solution line in three-dimensional space, or what we call the *canonical solution line*. Specifically, because $\mathbf{T}^{-1}\mathbf{b} = \mathbf{b}_O$, we can express the equation for the solution line for any APC design matrix as

$$\mathbf{b}_O = \mathbf{b}_O^* + s\mathbf{T}^{-1}\mathbf{v}, \tag{18}$$

where $\mathbf{b}_O^*$ is a constrained set of estimates separated into nonlinear and linear components. Moreover, because $s\mathbf{T}^{-1}\mathbf{v} = s\mathbf{v}_O$, we can write

$$\mathbf{b}_O = \mathbf{b}_O^* + s\mathbf{v}_O \tag{19}$$

or, equivalently,

$$
\begin{aligned}
\alpha^* &= \alpha + s \\
\pi^* &= \pi - s \\
\gamma^* &= \gamma + s,
\end{aligned}
\tag{20}
$$
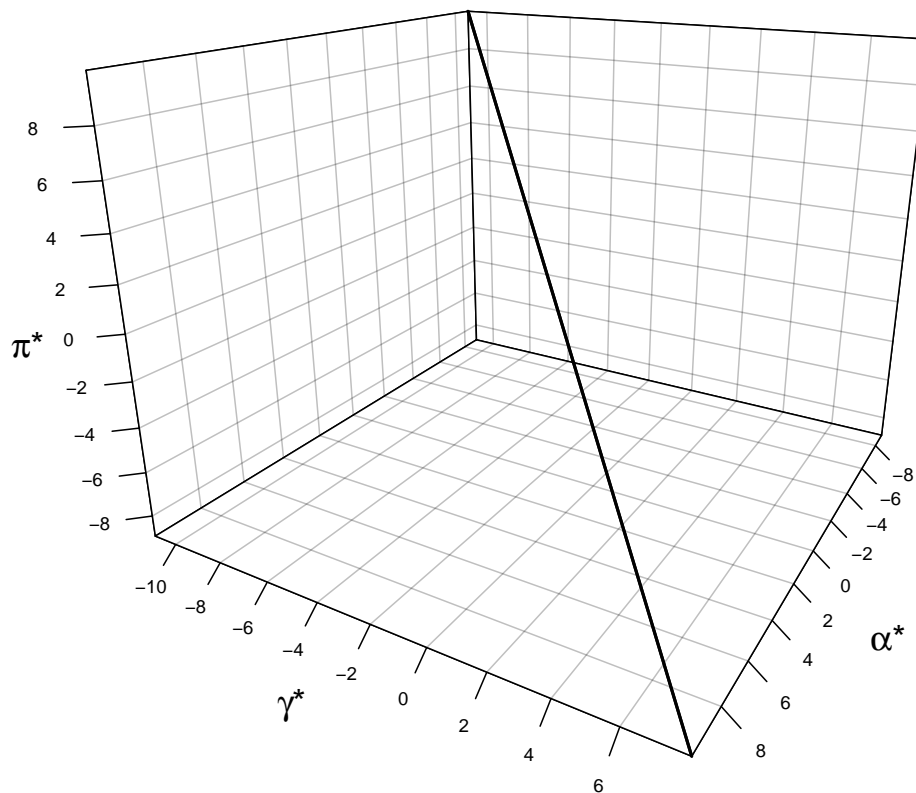
Figure 1: The canonical solution line. *Notes:* Data generated using values of $\alpha = 1.000$, $\alpha^2 = 0.050$, $\pi = -4.000$, $\pi^2 = -0.250$, $\gamma = 6.000$, $\gamma^2 = 0.500$, $\gamma^3 = 0.100$, and $\gamma^4 = 0.050$. Data consist of $I = 8$ age groups, $J = 5$ period groups, and $I + J - 1 = K = 12$ cohort groups.

where α, π, and γ denote the true, unknown data-generating slopes; s is an arbitrary scalar; and the asterisks denote any particular constrained set of estimates on the solution line. Because, as mentioned previously, s can take on any real number, Equation 20 defines a line in a three-dimensional space for the design matrix $\mathbf{XT} = \mathbf{X}_O$ or, in other words, the canonical design matrix. For example, Figure 1 visualizes the solution line for APC data generated using values of $\alpha = 1$, $\pi = -4$, and $\gamma = 6$ along with a set of nonlinear effects. Because the nonlinearities are identified, all possible estimates of the temporal effects lie on the line in the three-dimensional space in Figure 1.

Canonical Linear Constraints of MP Estimators

The major obstacle to understanding and comparing the linear constraints imposed by various MP estimators is that they differ greatly because they are a function of the design matrix. However, using the transformation matrix described in the previous section, we can derive a canonical form of the linear constraints of the

IE and related MP estimators. To obtain the canonical linear constraint of an MP estimator we can use the equation

$$\mathbf{v}^T \mathbf{T} \mathbf{b}_O = 0, \tag{21}$$

where again $\mathbf{b}$ is the original set of parameters; $\mathbf{T}$ is the transformation matrix; $\mathbf{v}$ is the original, untransformed null vector;[22] and $\mathbf{b}_O$ is the set of the parameters expressed in canonical form, with separate linear and nonlinear effects. There are several steps to deriving the canonical linear constraint of a particular MP estimator. First, we use the transformation matrix to decompose each parameter into its linear and nonlinear components: $\mathbf{b} = \mathbf{T} \mathbf{b}_O$. Next, we multiply the parameter vector of decomposed effects by the original, untransformed null vector $\mathbf{v}^T$. Finally, we rearrange terms so that the parameters representing the nonlinear effects are on the right-hand side of the equation and then simplify. This will give us us the canonical linear constraint for that particular MP estimator.

Like linear constraints discussed previously, the canonical linear constraints will differ across estimators. However, the canonical linear constraints of the IE and other MP estimators have a general form that clarifies their assumptions and their sensitivity to changes in the data

$$w_1 \alpha - w_2 \pi + w_3 \gamma = v, \tag{22}$$

where w_1, w_2, and w_3 are weights and v is a scalar. Equation 22 is absolutely crucial in understanding the divergent properties of MP estimators, including the IE. In general, across MP estimators, the estimates of the linear effects vary depending on at least three aspects of the data: first, the number of APC groups in the data set, which alter the weights for α, π, and γ as well as the value of the scalar v; second, the size and sign of the nonlinearities, which shifts the value of v; finally, the choice of the reference category, which also shifts the value of v. If there are no nonlinearities in the data set, then regardless of the MP estimator, $v = 0$; otherwise, the IE and other MP estimators constrain the weighted sum of the slopes to equal some other arbitrary value of v.

We examine the canonical linear contraints of nine MP estimators in Tables 3, 4, and 5. For all tables, we keep the number of age groups fixed at $I = 3$ but vary the number of period groups (and accordingly, the number of cohort groups). Note that in practice, the constraints in Tables 3, 4, and 5 can be simplified because the nonlinearities are identified; that is, the right-hand side of the canonical linear constraints reduces to a scalar when applied to any given set of data. For example, the canonical linear constraint of the IE_{First} in Table 3 simplifies to $\alpha - \pi + 6\gamma = v$, where the scalar $v = -\alpha^2 + \pi^2 - 4\gamma^2 + 2\gamma^3 - 2\gamma^4$.

Mathematically, the canonical linear constraints in these tables reveal that all of the MP estimators examined, except for the OE, will produce differing linear constraints and thus divergent estimates of the true temporal effects because of the following: changes in the category omitted (or similarly, the reference category used), the number of APC categories, and the size and direction of the nonlinearities. Out of all the estimators examined, only the $\text{DE}_{\text{Forward}}$, $\text{DE}_{\text{Backward}}$, and OE produce estimates that will not depend on whether or not the first or last is omitted. However,

Table 3: Canonical form of linear constraints with $I = 3$ and $J = 3$ groups.

MP Estimator	Canonical Linear Constraint		
	Linear Effects		Nonlinear Effects
IE_{First}:	$\alpha - \pi + 6\gamma$	$=$	$-\alpha^2 + \pi^2 - 4\gamma^2 + 2\gamma^3 - 2\gamma^4$
IE_{Last}:	$\alpha - \pi + 6\gamma$	$=$	$\alpha^2 - \pi^2 + 4\gamma^2 + 2\gamma^3 + 2\gamma^4$
TE_{First}:	$\alpha - \pi + 6\gamma$	$=$	$(\frac{3}{5})\alpha^2 - (\frac{3}{5})\pi^2 + 4\gamma^2 - 2\gamma^3 + 2\gamma^4$
TE_{Last}:	$\alpha - \pi + 6\gamma$	$=$	$-(\frac{3}{5})\alpha^2 + (\frac{3}{5})\pi^2 - 4\gamma^2 - 2\gamma^3 - 2\gamma^4$
$\text{HE}_{\text{Backward}}$:	$\alpha - \pi + (\frac{54}{13})\gamma$	$=$	$(\frac{3}{13})\alpha^2 - (\frac{3}{13})\pi^2 + (\frac{22}{39})\gamma^2 - (\frac{17}{78})\gamma^3 + (\frac{5}{26})\gamma^4$
$\text{HE}_{\text{Forward}}$:	$\alpha - \pi + (\frac{54}{13})\gamma$	$=$	$-(\frac{3}{13})\alpha^2 + (\frac{3}{13})\pi^2 - (\frac{22}{39})\gamma^2 - (\frac{17}{78})\gamma^3 - (\frac{5}{26})\gamma^4$
$\text{DE}_{\text{Backward}}$:	$\alpha - \pi + 2\gamma$	$=$	$-\gamma^3$
$\text{DE}_{\text{Forward}}$:	$\alpha - \pi + 2\gamma$	$=$	$-\gamma^3$
OE:	$\alpha - \pi + \gamma$	$=$	0

Notes: Decomposition based on $I = 3$ age, $J = 3$ period, and $K = I + J - 1 = 5$ cohort groups. The transformation matrix is used to decompose original effects into linear and nonlinear components, which is then multiplied by the original null vector. For ease of exposition, null vectors and orthogonal polynomial contrasts with integer elements are used to calculate the linear constraints. For some parameter β, we let β^k denote the kth orthogonal polynomial term. For example, γ^2 is the second-order orthogonal polynomial term for the cohort variable.

only the OE is robust to the size and sign of the nonlinearities as well as the number of APC categories used.

This analysis also informs us when MP estimators will appear to produce reliable results. MP estimators will perform well when (1) there are significant nonlinearities and there are zero or very small linear effects or (2) when the underlying true linear effects have the same relationship among each other as that of the canonical linear constraint. It is important to emphasize that the mathematical constraint imposed by MP estimators, including the IE, will not, generally speaking, recover the underlying data-generating parameters. The reason for this is that the estimates produced by the MP and the true data-generating parameters will equal each other only if the parameters happen to conform to the MP's specific mathematical constraint. We know of no reason why this should ever be the case in any particular substantive application.

Examples

We now turn to several simulations to supplement the mathematical discussion of the previous section. We first compare the various MP estimators when applied

Table 4: Canonical form of linear constraints with $I = 3$ and $J = 4$ groups.

MP Estimator	Canonical Linear Constraint		
	Linear Effects		Nonlinear Effects
IE_{First}:	$\alpha - (\frac{11}{2})\pi + (\frac{45}{2})\gamma$	$=$	$-\alpha^2 + (\frac{3}{2})\pi^2 - (\frac{3}{2})\pi^3 - (\frac{25}{2})\gamma^2 + (\frac{25}{2})\gamma^3 - (\frac{5}{2})\gamma^4 + (\frac{5}{2})\gamma^5$
IE_{Last}:	$\alpha - (\frac{11}{2})\pi + (\frac{45}{2})\gamma$	$=$	$\alpha^2 - (\frac{3}{2})\pi^2 - (\frac{3}{2})\pi^3 + (\frac{25}{2})\gamma^2 + (\frac{25}{2})\gamma^3 + (\frac{5}{2})\gamma^4 + (\frac{5}{2})\gamma^5$
TE_{First}:	$\alpha - (\frac{28}{5})\pi + 22\gamma$	$=$	$(\frac{3}{5})\alpha^2 - (\frac{6}{5})\pi^2 + (\frac{6}{5})\pi^3 + 15\gamma^2 - 15\gamma^3 + 3\gamma^4 - 3\gamma^5$
TE_{Last}:	$\alpha - (\frac{28}{5})\pi + 22\gamma$	$=$	$-(\frac{3}{5})\alpha^2 + (\frac{6}{5})\pi^2 + (\frac{6}{5})\pi^3 - 15\gamma^2 - 15\gamma^3 - 3\gamma^4 - 3\gamma^5$
$\text{HE}_{\text{Backward}}$:	$\alpha - (\frac{58}{13})\pi + (\frac{180}{13})\gamma$	$=$	$(\frac{3}{13})\alpha^2 - (\frac{10}{39})\pi^2 + (\frac{8}{39})\pi^3 + (\frac{20}{13})\gamma^2 - (\frac{15}{13})\gamma^3 + (\frac{1}{5})\gamma^4 - (\frac{12}{65})\gamma^5$
$\text{HE}_{\text{Forward}}$:	$\alpha - (\frac{58}{13})\pi + (\frac{180}{13})\gamma$	$=$	$-(\frac{3}{13})\alpha^2 + (\frac{10}{39})\pi^2 + (\frac{8}{39})\pi^3 - (\frac{20}{13})\gamma^2 - (\frac{15}{13})\gamma^3 - (\frac{1}{5})\gamma^4 - (\frac{12}{65})\gamma^5$
$\text{DE}_{\text{Backward}}$:	$\alpha - 3\pi + 5\gamma$	$=$	$\pi^3 - 5\gamma^3 - \gamma^5$
$\text{DE}_{\text{Forward}}$:	$\alpha - 3\pi + 5\gamma$	$=$	$\pi^3 - 5\gamma^3 - \gamma^5$
OE:	$\alpha - \pi + \gamma$	$=$	0

Notes: Decomposition based on $I = 3$ age, $J = 5$ period, and $K = I + J - 1 = 7$ cohort groups. The transformation matrix is used to decompose original effects into linear and nonlinear components, which is then multiplied by the original null vector. For some parameter β, we let β^k denote the kth orthogonal polynomial term. For example, γ^2 is the second-order orthogonal polynomial term for the cohort variable.

to a set of simulated data. These results are shown in Table 6. For simplicity, and without loss of generality, we assume there is no random error. The results show that for all MP estimators examined, they only differ in their linear components; the IE and related MP estimators all recover the intercept and the nonlinear effects. Note that none of the the estimators recover the true linear effects.[23]

The results in Table 6 raise an important issue regarding the use of fit statistics to determine whether or not an MP estimator should be used, as recommended by Land and colleagues (e.g., Yang and Land 2013a: 125–53; Yang and Land 2013b). They contend that a researcher should adopt a three-step procedure when considering using the IE. In the first step, the researcher conducts a descriptive analysis of the temporal effects using graphical techniques. In the second step, the researcher uses fit statistics to determine whether or not "the data are sufficiently well described by any single factor or two-way combination" of age, period, and cohort (Yang and Land 2013a: 126). If graphical techniques and model fitting suggest that "only one or two of the three effects are operative," then the researcher "can proceed with a reduced model that omits one or two groups of variables" because then "there is no identification problem" (Yang and Land 2013a: 126; Yang and Land 2013b: 1969). However, if these analyses suggest that "all three dimensions are at work," then one should proceed to step three and implement the IE (Yang and Land 2013a: 126). They emphasize that the IE should not be used unless "all three dimensions are operative" (Yang and Land 2013b: 1969).

Table 5: Canonical form of linear constraints with $I = 3$ and $J = 5$ groups.

MP Estimator	Linear Effects		Canonical Linear Constraint — Nonlinear Effects
IE_{First}:	$\alpha - 6\pi + 19\gamma$	$=$	$-\alpha^2 + 4\pi^2 - 2\pi^3 + 2\pi^4 - 15\gamma^2 + 3\gamma^3 - 9\gamma^4 + 3\gamma^5 - 3\gamma^6$
IE_{Last}:	$\alpha - 6\pi + 19\gamma$	$=$	$\alpha^2 - 4\pi^2 - 2\pi^3 - 2\pi^4 + 15\gamma^2 + 3\gamma^3 + 9\gamma^4 + 3\gamma^5 + 3\gamma^6$
TE_{First}:	$\alpha - 6\pi + \left(\frac{91}{5}\right)\gamma$	$=$	$\left(\frac{3}{5}\right)\alpha^2 - 4\pi^2 + 2\pi^3 - 2\pi^4 + 21\gamma^2 - \left(\frac{21}{5}\right)\gamma^3 + \left(\frac{63}{5}\right)\gamma^4 - \left(\frac{21}{5}\right)\gamma^5 + \left(\frac{21}{5}\right)\gamma^6$
TE_{Last}:	$\alpha - 6\pi + \left(\frac{91}{5}\right)\gamma$	$=$	$-\left(\frac{3}{5}\right)\alpha^2 + 4\pi^2 + 2\pi^3 + 2\pi^4 - 21\gamma^2 - \left(\frac{21}{5}\right)\gamma^3 - \left(\frac{63}{5}\right)\gamma^4 - \left(\frac{21}{5}\right)\gamma^5 - \left(\frac{21}{5}\right)\gamma^6$
$\text{HE}_{\text{Backward}}$:	$\alpha - \left(\frac{54}{13}\right)\pi + \left(\frac{139}{13}\right)\gamma$	$=$	$\left(\frac{3}{13}\right)\alpha^2 - \left(\frac{22}{39}\right)\pi^2 + \left(\frac{17}{78}\right)\pi^3 - \left(\frac{5}{26}\right)\pi^4 + \left(\frac{5}{3}\right)\gamma^2 - \left(\frac{19}{78}\right)\gamma^3 + \left(\frac{81}{130}\right)\gamma^4 - \left(\frac{37}{195}\right)\gamma^5 + \left(\frac{7}{39}\right)\gamma^6$
$\text{HE}_{\text{Forward}}$:	$\alpha - \left(\frac{54}{13}\right)\pi + \left(\frac{139}{13}\right)\gamma$	$=$	$-\left(\frac{3}{13}\right)\alpha^2 + \left(\frac{22}{39}\right)\pi^2 + \left(\frac{17}{78}\right)\pi^3 + \left(\frac{5}{26}\right)\pi^4 - \left(\frac{5}{3}\right)\gamma^2 - \left(\frac{19}{78}\right)\gamma^3 - \left(\frac{81}{130}\right)\gamma^4 - \left(\frac{37}{195}\right)\gamma^5 - \left(\frac{7}{39}\right)\gamma^6$
$\text{DE}_{\text{Backward}}$:	$\alpha - 2\pi + 3\gamma$	$=$	$-\gamma^3 - \gamma^5$
$\text{DE}_{\text{Forward}}$:	$\alpha - 2\pi + 3\gamma$	$=$	$-\gamma^3 - \gamma^5$
OE:	$\alpha - \pi + \gamma$	$=$	0

Notes: Decomposition based on $I = 3$ age, $J = 5$ period, and $K = I + J - 1 = 7$ cohort groups. The transformation matrix is used to decompose original effects into linear and nonlinear components, which is then multiplied by the original null vector. For some parameter β, we let β^k denote the kth orthogonal polynomial term. For example, γ^2 is the second-order orthogonal polynomial term for the cohort variable.

Table 6: Comparison of MP estimators: nonlinear and linear effects.

	μ	α	α^2	π	π^2	γ	γ^2	γ^3	γ^4
True Effects:	1.000	1.000	0.250	2.000	0.050	-1.000	-0.250	0.200	-0.100
IE_{First}:	1.000	2.050	0.250	0.950	0.050	0.050	-0.250	0.200	-0.100
IE_{Last}:	1.000	1.800	0.250	1.200	0.050	-0.200	-0.250	0.200	-0.100
TE_{First}:	1.000	1.690	0.250	1.310	0.050	-0.310	-0.250	0.200	-0.100
TE_{Last}:	1.000	1.960	0.250	1.040	0.050	-0.040	-0.250	0.200	-0.100
$\text{HE}_{\text{Backward}}$:	1.000	1.812	0.250	1.188	0.050	-0.188	-0.250	0.200	-0.100
$\text{HE}_{\text{Forward}}$:	1.000	1.849	0.250	1.151	0.050	-0.151	-0.250	0.200	-0.100
$\text{DE}_{\text{Backward}}$:	1.000	1.700	0.250	1.300	0.050	-0.300	-0.250	0.200	-0.100
$\text{DE}_{\text{Forward}}$:	1.000	1.700	0.250	1.300	0.050	-0.300	-0.250	0.200	-0.100
OE:	1.000	1.667	0.250	1.333	0.050	-0.333	-0.250	0.200	-0.100

Notes: Number of age and period groups is set at $I = 3$ and $J = 3$, respectively, for all simulations. Sample size for each simulation is $n = 100 \times (I \times J)$. Shaded columns indicate estimates of the intercept and nonlinearities, both of which are identified for all estimators. For simplicity, and without loss of generality, we assume no random error.

Unfortunately, it is impossible to determine from the data alone whether or not all three temporal variables are operating. Believing otherwise can seriously mislead researchers. For example, consider the models in Table 7. The underlying data-generating process consists of nonzero slopes for all three temporal variables: $\alpha = 0.500$, $\pi = 1.500$, and $\gamma = 2.000$. However, whereas the age and period variables have nonzero nonlinear effects in the data-generating process, all of the cohort nonlinearities are zero. Land and colleagues argue that one should use a reduced model when it fits the data equally well or better than a full model with all three temporal variables (Yang and Land 2013a: 109). The fit statistics in Table 7 suggest that one should, according to Land and colleagues, fit a two-factor age–period (AP) model rather than a three-factor APC model. However, by doing so, one does not avoid the nonidentifiability of the three linear effects. Rather, by fitting the two-factor model with only age and period effects, one is imposing the identification assumption that the cohort linear effect is zero even though its true linear effect is $\gamma = 2.000$. This zero-linear trend constraint on the cohort variable is external to the data, imposed by the researcher. Depending on the substantive application, it may or may not be reasonable to assume that because the nonlinear effects of cohort are observed to be zero, its linear effect is also zero. However, this is an assumption that can only be justified by appealing to theory or the inclusion of additional data.

We now compare two estimators in particular: the IE with the last category omitted (the most widely used of the MP estimators) and the OE (which, as we have shown mathematically, is robust to the number of APC categories, the use of the reference group, as well as the size and direction of the nonlinearities). The general canonical linear constraint of the previous section tells us that mathematically, the IE_{Last} is sensitive to the number of APC groups. Table 8 shows how the IE_{Last}'s estimates of the slopes change as the number of period groups increases from $J = 3$

Table 7: Estimates and goodness-of-fit statistics for IE$_{Last}$ and two-factor models.

Parameter	APC	AP	AC	PC
μ	0.984	0.988	1.375	0.328
$\alpha_1 - \bar{\alpha}$	−0.543	0.114	−7.272	
$\alpha_2 - \bar{\alpha}$	2.780	3.130	−0.103	
$\alpha_3 - \bar{\alpha}$	2.038	2.048	2.459	
$\alpha_4 - \bar{\alpha}$	1.302	0.942	4.467	
$\pi_1 - \bar{\pi}$	−5.416	−6.072		−2.498
$\pi_2 - \bar{\pi}$	−1.853	−2.213		−1.058
$\pi_3 - \bar{\pi}$	−2.943	−2.933		−3.480
$\pi_4 - \bar{\pi}$	1.555	1.905		0.116
$\gamma_1 - \bar{\gamma}$	−1.213		−13.045	−9.053
$\gamma_2 - \bar{\gamma}$	−1.059		−9.680	−4.397
$\gamma_3 - \bar{\gamma}$	−0.688		−7.687	−1.836
$\gamma_4 - \bar{\gamma}$	−0.343		−4.580	0.014
$\gamma_5 - \bar{\gamma}$	0.042		−0.349	0.698
$\gamma_6 - \bar{\gamma}$	0.265		2.735	3.045
$\gamma_7 - \bar{\gamma}$	0.794		5.890	4.113
$\gamma_8 - \bar{\gamma}$	1.011		10.532	4.374
R^2	0.812	0.812	0.744	0.705
Adjusted R^2	0.812	0.812	0.744	0.705
AIC	125963	125958	133678	137189
BIC	126109	126040	133792	137303

Notes: Based on simulated data with $I = 5$ age groups, $J = 5$ period groups, and $I + J - 1 = K = 9$ cohort groups. Sample size is $N = 25,000$. Data-generating parameters are $\mu = 1.000$, $\alpha = 0.500$, $\alpha^2 = −1.450$, $\alpha^3 = −0.200$, $\alpha^4 = −0.150$, $\pi = 1.500$, $\pi^2 = 0.900$, $\pi^3 = 0.700$, $\pi^4 = −0.200$, and $\gamma = 2.000$. Data-generating model includes disturbances drawn from a normal distribution with a mean of zero and standard deviation of 3.000. Zero-sum effect coding with the last category omitted is used for the two-factor AP (age–period), AC (age–cohort), and PC (period–cohort) models. Three-factor age–period–cohort (APC) model is estimated using IE$_{Last}$. Shaded column indicates model with lowest AIC and BIC scores.

to $J = 1000$.[24] For all simulations in Table 8, we keep the data-generating process the same with $\alpha = −2.000$, $\pi = 4.000$, and $\gamma = 1.000$. For simplicity, we keep the number of age groups constant at $I = 3$ as we increase the number of period groups (and accordingly, the number of cohorts). A different number of age groups does not alter our findings regarding the sensitivity of the IE$_{Last}$ to the number of APC groups. We purposely constructed the data-generating process so that it initially conforms to the IE's constraint. With $I = 3$ age groups and $J = 3$ period groups, as well as no nonlinearities, $w_1 = 1$, $w_2 = 1$, $w_3 = 6$, and $v = 0$. We purposely chose values of α, π, and γ so that $\alpha - \pi + 6\gamma = -2 - (4) + 6(1) = 0$. This is indicated in the first row of Table 8, which is shaded.

Table 8: Sensitivity of the IE_{Last} to number of period (and cohort) groups.

J Groups	α_{IE}	π_{IE}	γ_{IE}	Canonical Linear Constraint		
3	-2.000	4.000	1.000	$\alpha - \pi + 6\gamma$	$=$	0
5	-1.731	3.731	1.269	$\alpha - 6\pi + 19\gamma$	$=$	0
8	-1.368	3.368	1.632	$\alpha - \left(\frac{119}{4}\right)\pi + \left(\frac{249}{4}\right)\gamma$	$=$	0
10	-1.214	3.214	1.786	$\alpha - \left(\frac{249}{4}\right)\pi + \left(\frac{451}{4}\right)\gamma$	$=$	0
15	-0.990	2.990	2.010	$\alpha - 231\pi + 344\gamma$	$=$	0
25	-0.798	2.798	2.202	$\alpha - 1156\pi + 1469\gamma$	$=$	0
50	-0.650	2.650	2.350	$\alpha - \left(\frac{39,249}{4}\right)\pi + \left(\frac{44,251}{4}\right)\gamma$	$=$	0
100	-0.575	2.575	2.425	$\alpha - \left(\frac{323,499}{4}\right)\pi + \left(\frac{343,501}{4}\right)\gamma$	$=$	0
500	-0.515	2.515	2.485	$\alpha - (10,354,375)\pi + (10,479,375)\gamma$	$=$	0
1000	-0.507	2.507	2.493	$\alpha - (83,083,750)\pi + (83,583,750)\gamma$	$=$	0

True Linear Effects			
	$\alpha = -2.000$	$\pi = 4.000$	$\gamma = 1.000$

Notes: Number of age groups is set at $I = 3$ for all simulations. Sample size for each simulation is $n = 100 \times (I \times J)$. Shaded row indicates initial simulated data in which, by construction, the IE constraint is satisfied. Due to rounding, some IE constraints displayed here will not equal zero exactly.

As we increase the number of period groups, the values of the weights change, thereby altering the IE_{Last}'s canonical linear constraint. The canonical linear constraints are shown in last column of Table 8. Because the data-generating process contains no nonlinearities and remains the same as we increase the number of groups, the value of v for each constraint is zero. However, the values of the w's for the period and cohort slopes increase as we increase the number of period (and cohort) groups.[25] Although our initial data set satisfies the IE_{Last}'s canonical linear constraint, as the number of period groups increases, the IE_{Last} estimates diverge dramatically from the true data-generating parameters. In this particular case, as we increase the number of period groups, the IE_{Last}'s estimates of the age slope increases, the period slope decreases, and the cohort slope increases. The reason is that as the number of period and cohort groups increase, their values of w become very similar. With $J = 1000$ period groups, the weights for the period and cohort slopes are approximately the same. Thus, the claim that the IE's constraint on the true, unknown slopes is consistent as the number of time periods increases towards infinity is not generally true (Fu 2016).

The last column of Table 8 underscores the complicated, nonintuitive, and highly variable nature of the IE_{Last}'s canonical linear constraint as well as the extreme difficulty of interpreting it substantively. For example, with $J = 10$ groups, the constraint in Table 8 is $\alpha - \left(\frac{249}{4}\right)\pi + \left(\frac{451}{4}\right)\gamma = 0$, with weights of $w_1 = 1$, $w_2 = \left(\frac{249}{4}\right)$, and $w_3 = \left(\frac{451}{4}\right)$. However, with $J = 15$ groups, the IE constraint becomes $\alpha - 231\pi + 344\gamma$, with weights of $w_1 = 1$, $w_2 = 231$, and $w_3 = 344$. It is

Table 9: Robustness of the OE to number of period (and cohort) groups.

J Groups	α_{OE}	π_{OE}	γ_{OE}	Canonical Linear Constraint		
3	-2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
5	-2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
8	-2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
10	-2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
15	-2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
25	-2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
50	-2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
100	-2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
500	-2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
1000	-2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
True Linear Effects	$\alpha = -2.000$	$\pi = 1.000$	$\gamma = 3.000$			

Notes: Number of age groups is set at $I = 3$ for all simulations. Sample size for each simulation is $n = 100 \times (I \times J)$. Shaded row indicates initial simulated data in which, by construction, the OE constraint is satisfied.

exceedingly difficult, if not impossible, to muster a reason why these specific values of w have any theoretical importance. In contrast, as shown in Table 9, regardless of the number of groups, the OE's canonical linear constraint is $\alpha - \pi + \gamma = 0$.

The canonical linear constraint also clarifies that the IE_{Last}'s estimates depend on the size (e.g., large or small in absolute value) and sign (e.g., positive or negative) of the nonlinearities. For example, using the same data-generating slopes as in the previous section, Table 10 shows how the IE_{Last}'s estimates alter depending on the magnitude and direction of the age nonlinearity. For all simulations, we keep the number of age and period groups at $I = 3$ and $J = 3$, respectively. Again, because we specifically constructed a data-generating process that conforms to the IE constraint when there are $I = 3$ age groups and $J = 3$ period groups as well as zero nonlinearities, the IE_{Last} indeed recovers the true slopes when the age nonlinearity is zero. This row is shaded in Table 10.

Because the number of age groups is set at $I = 3$ for all simulations in Table 10, the age weight is set at $w_1 = 1$, and the value of v changes directly with the age nonlinearity. As the age nonlinearity becomes more positive, the age and cohort slopes move towards positive infinity on the real number line, whereas the period slope moves towards negative infinity. In contrast, as the age nonlinearity becomes more negative, the age and cohort slopes move towards negative infinity on the real number line, whereas the period slope moves towards positive infinity. For the same underlying age, period, and cohort linear effects in the population, the IE_{Last} will give radically different estimates of the slopes depending on the nonlinear

Table 10: Sensitivity of the IE_{Last} to age nonlinearities.

α^2	α_{IE}	π_{IE}	γ_{IE}	Canonical Linear Constraint		
-20.000	-4.500	6.500	-1.500	$\alpha - \pi + 6\gamma$	$=$	-20.000
-10.000	-3.250	5.250	-0.250	$\alpha - \pi + 6\gamma$	$=$	-10.000
-5.000	-2.625	4.625	0.375	$\alpha - \pi + 6\gamma$	$=$	-5.000
-1.000	-2.125	4.125	0.875	$\alpha - \pi + 6\gamma$	$=$	-1.000
-0.500	-2.062	4.062	0.938	$\alpha - \pi + 6\gamma$	$=$	-0.500
-0.250	-2.031	4.031	0.969	$\alpha - \pi + 6\gamma$	$=$	-0.250
-0.050	-2.006	4.006	0.994	$\alpha - \pi + 6\gamma$	$=$	-0.050
0.000	-2.000	4.000	1.000	$\alpha - \pi + 6\gamma$	$=$	0.000
0.050	-1.994	3.994	1.006	$\alpha - \pi + 6\gamma$	$=$	$+0.050$
0.250	-1.969	3.969	1.031	$\alpha - \pi + 6\gamma$	$=$	$+0.250$
0.500	-1.938	3.938	1.063	$\alpha - \pi + 6\gamma$	$=$	$+0.500$
1.000	-1.875	3.875	1.125	$\alpha - \pi + 6\gamma$	$=$	$+1.000$
5.000	-1.375	3.375	1.625	$\alpha - \pi + 6\gamma$	$=$	$+5.000$
10.000	-1.075	2.750	2.250	$\alpha - \pi + 6\gamma$	$=$	$+10.000$
20.000	0.500	1.500	3.500	$\alpha - \pi + 6\gamma$	$=$	$+20.000$
True Linear Effects	$\alpha = -2.000$	$\pi = 4.000$	$\gamma = 1.000$			

Notes: For all simulations, number of age, period, and cohort groups is set at $I = 3$, $J = 3$, and $K = I + J - 1 = 5$, respectively. Sample size for each simulation is $n = 1{,}000 \times (I \times J) = 9{,}000$. Shaded row indicates simulated data in which, by construction, the IE constraint is satisfied.

effects. This is in contrast to the OE, which (as shown in 11) provides the same constraint regardless of the nonlinearity of age.

The last column of Table 10 again brings to light the complex and variable nature of the IE_{Last}'s constraint, even in canonical form. For instance, with $\alpha^2 = -1.000$, the constraint in Table 10 is $\alpha - \pi + 6\gamma = -1.000$, but with $\alpha^2 = +1.000$, the constraint becomes $\alpha - \pi + 6\gamma = +1.000$. Again, it is very difficult, if not impossible, to give a theoretical reason why this particular linear combination of the temporal slopes in the population should equal $v = +1.000$ rather than $v = -1.000$ simply because the quadratic age trend is a positive one rather than a negative one. In contrast, the OE imposes the same linear constraint regardless of the magnitude or direction of the nonlinearities. As demonstrated in Table 11, the OE's canonical linear constraint is always $\alpha - \pi + \gamma = 0$ despite changes in the size or sign of the age-squared effect.

Table 11: Robustness of the OE to age nonlinearities.

α^2	α_{OE}	π_{OE}	γ_{OE}	Canonical Linear Constraint		
−20.000	−2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
−10.000	−2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
−5.000	−2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
−1.000	−2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
−0.500	−2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
−0.250	−2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
−0.050	−2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
0.000	−2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
0.050	−2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
0.250	−2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
0.500	−2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
1.000	−2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
5.000	−2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
10.000	−2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
20.000	−2.000	1.000	3.000	$\alpha - \pi + \gamma$	$=$	0
True Linear Effects	$\alpha = -2.000$	$\pi = 1.000$	$\gamma = 3.000$			

Notes: For all simulations, number of age, period, and cohort groups is set at $I = 3$, $J = 3$, and $K = I + J - 1 = 5$, respectively. Sample size for each simulation is $n = 1{,}000 \times (I \times J) = 9{,}000$. Shaded row indicates simulated data in which, by construction, the OE constraint is satisfied.

Conclusion

This article makes a number of important conclusions not recognized or fully appreciated in the current APC literature. First, we compare the similarities and differences of MP estimators, clarifying that although all MP estimators share the same desirable statistical properties, they diverge in a decisive way: because they are based on varying design matrices, they impose differing linear constraints on the true, unknown temporal effects. Second, we show how to explicitly construct a transformation matrix that allows one to convert any APC design matrix in canonical form with the linear and nonlinear components orthogonally separated. Third, using this transformation matrix and canonical form of the design matrix, we then prove that all constrained estimators, including MP estimators, generate the same set of nonlinear effects but differing linear effects. Moreover, we show that the solution line of any APC data set can always be simplified to a canonical form that spans just three dimensions. Fourth, we show mathematically that the IE and related MP estimators have a general canonical linear constraint, parallel to the fundamental linear dependency among the time scales, such that $w_1\alpha -$

$w_2\pi + w_3\gamma = v$, where the w's are weights and v is a scalar. To our knowledge, this is the first analysis to reveal this general canonical linear constraint and use it to compare multiple MP estimators. Finally, we show, both mathematically and using simulations, that the IE and a number of other MP estimators produce varying canonical linear constraints depending on the size and sign of the nonlinear effects, the number of APC categories, and the choice of reference category. However, two MP estimators, the OE and DE, are both easier to interpret and more robust than the IE. In particular, we find that the OE's constraint can always be expressed as $\alpha - \pi + \gamma = 0$.

We now conclude by offering a set of practical guidelines for APC researchers. In general, we do not recommend using MP estimators, including the IE, in an attempt to recover the true APC effects. However, if a researcher chooses to use an MP estimator, including the IE, then we have several recommendations. First, regardless of the MP estimator one uses, the full set of linear and nonlinear effects should be reported. This will allow the researcher to evaluate the legitimacy of the constraint imposed on the true, unknown linear APC effects. Fit statistics should be used judiciously, with full appreciation that they cannot discern whether or not two or three linear effects are operating in any particular data set. Ultimately, any linear constraint should be grounded in an underlying social, cultural, or biological theory.

Second, if using an MP estimator, then we recommend using the OE. The OE's linear constraint on the true temporal effects does not vary based on changes in the design matrix, such as the number of period groups or the size and sign of the nonlinearities. Moreover, the OE has the minimum sampling variance with respect to the canonical solution line spanning three dimensions and has the null vector with the smallest number of nonzero elements. In this way, the OE is the most parsimonious representation of the linear dependency inherent in any APC data set.

Finally, we caution that researchers who use the IE or related MP estimators in the hopes of uncovering the true data-generating parameters are unlikely to attain their goal. If using the IE, one should state explicitly why it is reasonable to assume that the model parameters in the population have the same linearly dependent relationship among each other as their corresponding columns in the design matrix. In general, we suspect that this constraint will be difficult, if not impossible, to justify. At this point, it is unclear what type of logic would be used to connect a set of theoretical considerations to a set of assumptions about the linear dependency in the design matrix. To emphasize, there is, to our knowledge, no social, biological, or cultural theory that claims that the true, unknown APC effects must conform to some specific linear dependence in the data.

Although we think that the OE is preferable to the IE, this in no way implies that it is preferable to other approaches to analyzing APC data. In general, one can divide the set of methods for identifying APC effects into statistical and theoretical approaches. The strength of a statistical approach, such as the OE or IE, is that it is not based on any explicit theoretical or substantive assumptions that researchers may strongly disagree about. The weakness of a statistical approach is that there is, in general, no reason to believe that it estimates the true parameters for the model

that generated the data. A variety of theoretically based strategies for identifying APC effects are possible. For example, one could assume, using social, biological, or cultural theory, that the linear effect of the age, period, or cohort variables is positive or negative. As Fosse and Winship (2018) show, this can be used to bound estimates of the data-generating parameters, in some cases leading to quite narrow bounds despite weak assumptions. An alternative theoretical approach entails specifying the mechanisms through which age, period, and cohort impact the outcome. As Winship and Harding (2008) demonstrate, if one has variables measuring all the pathways through which at least one of the age, period, or cohort variables operate, then it is possible to identify the underlying data-generating parameters. To be sure, an appreciable issue with any theoretical approach is that researchers may well disagree on the validity of particular assumptions. Although this can be a serious problem, at least when there is disagreement, it will be clear why different approaches lead to different estimates.

Ultimately, a theoretical approach is to be preferred to a purely statistical one. In most cases, researchers are interested in the true but unknown APC effects. We appreciate that others see more value in a statistical approach. Nonetheless, if one is going to use a statistical approach to identification, there is merit in using an estimator, such as the OE, that provides separate estimates of the nonidentified linear and identified nonlinear parameters. As we have demonstrated, doing so produces estimates of the linear effects that are not affected by the number of periods and cohorts available to the researcher, the choice of reference category, or the size and sign of the nonlinear effects.

Notes

1 Following the convention in the APC literature, we use the term "effects" when referring to age, period, and cohort processes (e.g., Glenn 1981: 249; Fienberg and Mason 1979: 133; Mason et al. 1973: 243; O'Brien 2015: 1; Yang and Land 2013a: 1–2). These "effects" need not refer to causal effects in the sense of parameters associated with well-defined potential outcomes (Morgan and Winship 2014: 37–76).

2 For our purposes here, we focus on MP estimators based on contrast coding schemes commonly discussed in the applied statistics literature (e.g., Fox 2002: 126–30; Onyiah 2008: 121–36,148–61; Venables and Ripley 2013: 146–9), in addition to those examined in the methodological literature on APC effects in sociology and demography (e.g., Luo et al. 2016; O'Brien 2015: 55–57; Pelzer et al. 2014).

3 As Land and colleagues (Land et al. 2016: 964) have asserted, "only the sum-to-zero [effect] coding is used to define and estimate the IE." However, as we discuss later, this definition differs form that of Fu (2016).

4 In the context of the present article, a set of parameters are *identifiable* if there is an estimator that would produce the true, underlying data-generating parameters on a sample of infinite size. In contrast, *estimability* implies that an estimator gives some unique value whether or not the estimate is equal to the true, underlying data-generating parameters. Thus, identification implies estimability, but estimability does not imply identification. For a succinct but clear discussion on defining identifiability and estimability, see Greenland (2005).

5 For the purposes of this article, by "norm," we refer to the Euclidean or, equivalently, L_2 norm.

6 For simplicity, we also assume that the age and period categories are of equal width.

7 Note that I is added to $j - i$ so that the cohort index begins at $k = 1$. This ensures that, for example, $i = j = k = 1$ refers to the first group for all three temporal scales. One could just as easily index the cohorts using $k = j - i$, but this identity would be lost.

8 Without loss of generality, in the discussion that follows, we assume that there are no disturbances such that $\epsilon = 0$.

9 Because $\mathbf{X}s\mathbf{v} = \mathbf{X}^T\mathbf{X}s\mathbf{v} = \mathbf{0}$, we know that $\mathbf{X}^T\mathbf{y} = \mathbf{X}^T\mathbf{X}\mathbf{b}^* + \mathbf{X}^T\mathbf{X}s\mathbf{v}$. Rearranging, we can thus write $\mathbf{X}^T\mathbf{y} = \mathbf{X}^T\mathbf{X}(\mathbf{b}^* + s\mathbf{v})$. Because $\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{X}(\mathbf{b}^* + s\mathbf{v})$, it follows that $\mathbf{b} = \mathbf{b}^* + s\mathbf{v}$.

10 The fact that $\mathbf{X}^+\mathbf{y} = \mathbf{b}_{\text{OLS}}$ if $\mathbf{X}$ is of full rank may look odd to some readers, but this can be demonstrated algebraically. The first and third conditions imply that $\mathbf{X}^T\mathbf{X}\mathbf{X}^+ = \mathbf{X}^T$. Substituting $\mathbf{b} = \mathbf{X}^+\mathbf{y}$ into the normal equations $\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y}$, we obtain $\mathbf{X}^T\mathbf{X}\mathbf{X}^+\mathbf{y} = \mathbf{X}^T\mathbf{y}$. If $\mathbf{X}$ is of full rank, then we can take the regular inverse to solve for $\mathbf{X}^+\mathbf{y}$, which results in $\mathbf{X}^+\mathbf{y} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{b}_{\text{OLS}}$.

11 There are other possible design matrices that are deficient by more than rank one. These would be matrices that do not provide a full representation of the linear and nonlinear effects of age, period, and cohort. For our purposes here, such matrices are not of concern.

12 A similar decomposition exists for the design matrix $\mathbf{X}$ in which the eigenvalues will be the square roots of the diagonal values of $\mathbf{\Lambda}$. It thus also has a zero eigenvalue, and it will have the same null vector as $\mathbf{X}^T\mathbf{X}$.

13 For any particular constrained set of estimates $\mathbf{b}^*$, an MP estimator minimizes the Euclidean length of $\mathbf{b}^* + s\mathbf{v}$ or, equivalently, its squared length $(\mathbf{b}^* + s\mathbf{v})^T(\mathbf{b}^* + s\mathbf{v})$. Using calculus, it can be shown that this squared length is at its minimum when $s = -\mathbf{v}^T\mathbf{b}^*$.

14 The normalized null vector is given by $\hat{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|_2} = \frac{\mathbf{v}}{(\mathbf{v}^T\mathbf{v})^{1/2}}$, where the double bars ($\|.\|_2$) denote the Euclidean norm (or length) of the vector and the caret ($\hat{}$) indicates normalization. The caret notation, which is commonly used in the mathematics literature to denote a normalized vector, should not be confused with the predicted values of a statistical model. The normalized null vector is like $\mathbf{v}$ except that it has the additional property that its Euclidean length is one, such that $\hat{\mathbf{v}}^T\hat{\mathbf{v}} = 1$. Because $\hat{\mathbf{v}}$ and $\mathbf{v}$ differ only by multiplication by a scalar, for ease of exposition, we will refer to $\mathbf{v}$ unless necessary for computational purposes.

15 Combining these two stages, in the limit, we can write any given MP estimator as minimize $\|\mathbf{X}\mathbf{b} - \mathbf{y}\|_2^2 + \lambda^2\|\mathbf{b}\|_2^2$, where λ is a regularization parameter and $\lambda \to 0$ (cf. Fu 2016: 185). As λ approaches zero, both the MP and ridge (or Tikhonov) regression solutions converge to that set of estimates with the smallest Euclidean norm.

16 Formally, an MP estimator is that set of estimates that minimizes the sum of the squared parameter estimates, or $(\mathbf{b} + s\mathbf{v})^T(\mathbf{b} + s\mathbf{v})$, for a particular design matrix. This is the mathematical function that an MP estimator applies to obtain a set of estimates.

17 A crucial point of confusion is that stating a function is unbiased does not mean that it is an unbiased function for some specific property of the population. To clarify this point, consider the following simple example. The function five times the mean of a variable is an unbiased estimator. If we were to calculate it over an infinite number of samples and take its average, it will equal five times the mean in the population. Formally, this is because five times the mean is a linear function. Obviously, five times the mean is not an unbiased estimator for the mean in the population itself. The fact that MP estimators are unbiased (meaning that for any particular MP estimator, its average across an infinite

number of samples is equal to its value calculated in the population) does *not* indicate that it produces an unbiased estimate of the parameters of the underlying model that generated the data (that is, the true, unknown temporal effects).

18 This can be further revealed by simply examining the various proofs showing that the IE has these properties and noting that none depend on using a zero-sum, effects-coded design matrix (for example, see Fu 2000: 263–8,276–7; Yang et al. 2004: 107–8; Yang and Land 2013a: 75–123).

19 Recall that a linear combination is any mathematical expression that entails adding a set of terms each multiplied by a constant, where the constant can include one. For example, the well-known relationship between temperature in degrees Fahrenheit (F) and Celsius (C) is a linear combination: $F = C \times \frac{9}{5} + 32$. When we convert from degrees Celsius to Fahrenheit, we are not changing the temperature; rather, we are simply recentering (by adding 32) and rescaling (by multiplying by $\frac{9}{5}$) the distribution of the temperature in degrees Celsius. Crucially, once we know the temperature in degrees Celsius, we know the temperature in Fahrenheit because it is a simple linear combination. In a similar way, regardless of the coding scheme for a set of APC data, we can express at least one variable as a linear relationship of the other variables.

20 Note that the nonzero null vector elements correspond to the columns of a design matrix that are not identified.

21 Yang and colleagues (Yang et al. 2008: 1703) state that the "conventional wisdom is that only the nonlinear, but not the linear, components of APC models can be estimable... however, there have been only numeric demonstrations, but no rigorous proofs, to support the idea that no estimable function exists." Although the linear trends are not individually identifiable, there are "estimable functions" in the sense that particular linear combinations of the linear trends are identifiable.

22 For simplicity of exposition, we have expressed the null vectors in unnormalized form.

23 From our experience, it is a very simple matter to set up a data-generating process with linear effects that are not identified using the IE or other MP estimators. This is not surprising because for any given constrained estimator, there is just one point on the solution line corresponding to its linear constraint. To put it another way, there is just one way the IE or other MP estimator can recover the true slopes for a given solution line, whereas there are vastly many more ways for it to fail to do so.

24 This reflects the fact that in actual studies conducted over time, the number of age groups are constant due to the limits of human lifespan, but the number of period (and cohort) groups will increase.

25 The weight for the age slope would similarly increase if we increased the number of age groups across the simulations.

References

Ben-Israel, Adi. 2002. "The Moore of the Moore-Penrose Inverse." *Electronic Journal of Linear Algebra* 9:150–7. https://doi.org/10.13001/1081-3810.1083.

Draper, Norman R., and Harry Smith. 2014. *Applied Regression Analysis.* New York, NY: John Wiley & Sons.

Fienberg, Stephen E., and William M. Mason. 1979. "Identification and Estimation of Age-Period-Cohort Models in the Analysis of Discrete Archival Data." *Sociological Methodology* 10:1–67. https://doi.org/10.2307/270764.

Fosse, Ethan, and Christopher Winship. 2018. "Bounding Analyses of Age-Period-Cohort Effects." Working Paper.

Fox, John. 2002. *Linear Mixed Models: Appendix to An R and S-PLUS Companion to Applied Regression*. Thousand Oaks, CA: Sage Publications, Inc.

Fu, Qiang, and Kenneth C. Land. 2015. "The Increasing Prevalence of Overweight and Obesity of Children and Youth in China, 1989-2009: An Age-Period-Cohort Analysis." *Population Research and Policy Review* 34:901–21. https://doi.org/10.1007/s11113-015-9372-y.

Fu, Wenjiang J. 2000. "Ridge Estimator in Singular Design with Application to Age-Period-Cohort Analysis of Disease Rates." *Communications in Statistics - Theory and Methods* 29:263–78. https://doi.org/10.1080/03610920008832483.

Fu, Wenjiang J. 2008. "A Smoothing Cohort Model in Age-Period-Cohort Analysis with Applications to Homicide Arrest Rates and Lung Cancer Mortality Rates." *Sociological Methods & Research* 36:327–61. https://doi.org/10.1177/0049124107310637.

Fu, Wenjiang J. 2016. "Constrained Estimators and Consistency of a Regression Model on a Lexis Diagram." *Journal of the American Statistical Association* 111:180–99. https://doi.org/10.1080/01621459.2014.998761.

Fu, Wenjiang J., and Peter Hall. 2006. "Asymptotic Properties of Estimators in Age-Period-Cohort Analysis." *Statistics & Probability Letters* 76:1925–9. https://doi.org/10.1016/j.spl.2006.04.051.

Fu, Wenjiang J., Kenneth C. Land, and Yang Yang. 2011. "On the Intrinsic Estimator and Constrained Estimators in Age-Period-Cohort Models." *Sociological Methods & Research* 40:453–66. https://doi.org/10.1177/0049124111415355.

Glenn, Norval D. 1981. "The Utility and Logic of Cohort Analysis." *The Journal of Applied Behavioral Science* 17:247–57. https://doi.org/10.1177/002188638101700209.

Greenland, Sander. 2005. "Identifiability." In *Encyclopedia of Biostatistics*. Hoboken, NJ: John Wiley & Sons, Ltd. https://doi.org/10.1002/0470011815.b2a15055.

Hu, Anning. 2015. "A Loosening Tray of Sand? Age, Period, and Cohort Effects on Generalized Trust in Reform-Era China, 1990-2007." *Social Science Research* 51:233–46. https://doi.org/10.1016/j.ssresearch.2014.09.008.

Keyes, Katherine M., Dahsan S. Gary, Jordan Beardslee, Seth J. Prins, Patrick M. O'Malley, Caroline Rutherford, and John Schulenberg. 2017. "Age, Period, and Cohort Effects in Conduct Problems among American Adolescents from 1991 through 2015." *American Journal of Epidemiology* 187:548–57. https://doi.org/10.1093/aje/kwx268.

Kramer, Michael R., Amy L. Valderrama, and Michele L. Casper. 2015. "Decomposing Black-White Disparities in Heart Disease Mortality in the United States,

1973-2010: An Age-Period-Cohort Analysis." *American Journal of Epidemiology* 182:302–12. https://doi.org/10.1093/aje/kwv050.

Land, Kenneth C., Qiang Fu, Xin Guo, Sun Y. Jeon, Eric N. Reither, and Emma Zang. 2016. "Playing With the Rules and Making Misleading Statements: A Response to Luo, Hodges, Winship, and Powers." *American Journal of Sociology* 122:962–73. https://doi.org/10.1086/689853.

Li, Chunhui, Chuanhua Yu, and Peigang Wang. 2015. "An Age-Period-Cohort Analysis of Female Breast Cancer Mortality from 1990-2009 in China." *International Journal for Equity in Health* 14:76. https://doi.org/10.1186/s12939-015-0211-x.

Luo, Liying. 2013. "Assessing Validity and Application Scope of the Intrinsic Estimator Approach to the Age-Period-Cohort Problem." *Demography* 50:1945–67. https://doi.org/10.1007/s13524-013-0243-z.

Luo, Liying, James Hodges, Christopher Winship, and Daniel Powers. 2016. "The Sensitivity of the Intrinsic Estimator to Coding Schemes: Comment on Yang, Schulhofer-Wohl, Fu, and Land." *American Journal of Sociology* 122:930–61. https://doi.org/10.1086/689830.

Mason, Karen Oppenheim, William M. Mason, H. H. Winsborough, and W. Kenneth Poole. 1973. "Some Methodological Issues in Cohort Analysis of Archival Data." *American Sociological Review* 38:242–58. https://doi.org/10.2307/2094398.

Morgan, Stephen L., and Christopher Winship. 2014. *Counterfactuals and Causal Inference*. Cambridge, UK: Cambridge University Press, 2nd edition. https://doi.org/10.1017/CBO9781107587991.

O'Brien, Robert. 2015. *Age-Period-Cohort Models: Approaches and Analyses with Aggregate Data*. Boca Raton, FL: CRC Press.

O'Brien, Robert M. 2011. "Constrained Estimators and Age-Period-Cohort Models." *Sociological Methods & Research* 40:419–52. https://doi.org/10.1177/0049124111415367.

Onyiah, Leonard C. 2008. *Design and Analysis of Experiments: Classical and Regression Approaches with SAS*. Boca Raton, FL: CRC Press. https://doi.org/10.1201/b15920.

Pelzer, Ben, Manfred te Grotenhuis, Rob Eisinga, and Alexander W. Schmidt-Catran. 2014. "The Non-Uniqueness Property of the Intrinsic Estimator in APC Models." *Demography* 52:315–27. https://doi.org/10.1007/s13524-014-0360-3.

Price, Joseph, Rich Patterson, Mark Regnerus, and Jacob Walley. 2016. "How Much More XXX is Generation X Consuming? Evidence of Changing Attitudes and Behaviors Related to Pornography Since 1973." *The Journal of Sex Research* 53:12–20. https://doi.org/10.1080/00224499.2014.1003773.

Venables, William N., and Brian D. Ripley. 2013. *Modern Applied Statistics with S-PLUS*. New York, NY: Springer Science & Business Media.

Winship, Christopher, and David J. Harding. 2008. "A Mechanism-Based Approach to the Identification of Age-Period-Cohort Models." *Sociological Methods & Research* 36:362–401. https://doi.org/10.1177/0049124107310635.

Yang, Yang, Wenjiang J. Fu, and Kenneth C. Land. 2004. "A Methodological Comparison of Age-Period-Cohort Models: The Intrinsic Estimator and Conventional Generalized Linear Models." *Sociological Methodology* 34:75–110. https://doi.org/10.1111/j.0081-1750.2004.00148.x.

Yang, Yang, and Kenneth C. Land. 2013a. *Age-Period-Cohort Analysis: New Models, Methods, and Empirical Applications*. Boca Raton, FL: CRC Press. https://doi.org/10.1201/b13902.

Yang, Yang, and Kenneth C. Land. 2013b. "Misunderstandings, Mischaracterizations, and the Problematic Choice of a Specific Instance in Which the IE Should Never Be Applied." *Demography* 50:1969–71. https://doi.org/10.1007/s13524-013-0254-9.

Yang, Yang, Sam Schulhofer-Wohl, Wenjiang J. Fu, and Kenneth C. Land. 2008. "The Intrinsic Estimator for Age-Period-Cohort Analysis: What It Is and How to Use It." *American Journal of Sociology* 113:1697–736. https://doi.org/10.1086/587154.

Ethan Fosse: Department of Sociology, Princeton University.
E-mail: efosse@princeton.edu.

Christopher Winship: Department of Sociology, Harvard University.
E-mail: cwinship@wjh.harvard.edu.