

Manuscript submitted to
AIMS' Journals
Volume X, Number 0X, XX 200X

doi:10.3934/xx.xx.xx.xx

pp. X-XX

**THE USES AND ABUSES OF AN AGE-PERIOD-COHORT
METHOD: ON THE LINEAR ALGEBRA AND STATISTICAL
PROPERTIES OF INTRINSIC AND RELATED ESTIMATORS**

QIANG FU*

Department of Sociology, The University of British Columbia,
Vancouver, BC V6T 1Z1, Canada

XIN GUO

Department of Applied Mathematics, The Hong Kong Polytechnic University,
Hung Hom, Kowloon, Hong Kong; and
School of Mathematics and Physics, The University of Queensland,
Brisbane, QLD 4072, Australia

SUN YOUNG JEON

School of Medicine, University of California, San Francisco,
San Francisco, CA 94121, USA

ERIC N. REITHER

Department of Sociology, Social Work, and Anthropology, Utah State University,
Logan, UT 84322, USA

EMMA ZANG

Department of Sociology, Yale University,
New Haven, CT 06511, USA

KENNETH C. LAND

Department of Sociology and Social Science Research Institute, Duke University,
Durham, NC 27708, USA

(Communicated by the associate editor name)

2010 *Mathematics Subject Classification*. Primary: 91D20; Secondary: 15A06.

Key words and phrases. Intrinsic estimator, Age-period-cohort analysis, Statistics, Asymptotic property, Ridge regression, Principal components regression.

This research was partially supported by the Research Grants Council of Hong Kong [Project No. PolyU 15334616] and partially based on class notes provided by Qiang Fu during the course "Age-Period-Cohort Analysis: Principles, Models and Application", given in the Institute for Empirical Social Science Research (IESSR) at Xi'an Jiaotong University (July 2015) and in the School of Public Administration at Zhongnan University of Economics and Law (July 2018).

* Corresponding author: Qiang Fu.

ABSTRACT. As a sophisticated and popular age-period-cohort method, the Intrinsic Estimator (IE) and related estimators have evoked intense debate in demography, sociology, epidemiology and statistics. This study aims to provide a more holistic review and critical assessment of the overall methodological significance of the IE and related estimators in age-period-cohort analysis. We derive the statistical properties of the IE from a linear algebraic perspective, provide more precise mathematical proofs relevant to the current debate, and demonstrate the essential, yet overlooked, link between the IE and classical statistical tools that have been employed by scholars for decades. This study offers guidelines for the future use of the IE and related estimators in demographic research. The exposition of the IE and related estimators may help redirect, if not settle, the logic of the debate.

1. Introduction. For several decades, the classical *age-period-cohort accounting (multiple) classification model (APCMC model [26])* has been a subject of methodological attention and commentary. The reason for this is clear: Age-period-cohort (APC) analysis has played a critical role in studying time-specific phenomena in epidemiology, demography, and sociology for the past 90 years [28]. This approach to analysis distinguishes three types of time-related variation/temporal dimensions in the phenomena of interest: *age effects*, or variation associated with different age groups, *period effects*, or variation over time periods that affects all age groups simultaneously, and *cohort effects*, or changes across groups of individuals who experience an initial event such as birth in the same year or years.

These distinctions have important implications for measurement and analysis. The considerable regularity of age variations in many outcomes across time and place reflects the developmental nature of age changes across the life course. In contrast, period and cohort effects reflect the influences of social, demographic, economic, and other structural forces. Period variations often result from shifts in social, historical, and cultural environments. Cohort variations are conceived as the essence of social change and may reflect the effects of early life exposure to socioeconomic, behavioral, and environmental factors that act persistently over time to produce differences in life-course outcomes for specific cohorts [52].

The APCMC model serves as a general framework for cohort analysis when all three temporal dimensions of age, period, and cohort are potentially of interest and the data to be analyzed are in the form of tables of percentages or occurrence/exposure rates of events such as births, deaths, disease incidence, crimes, etc. In spite of its theoretical merits and conceptual relevance, APC analysis of tabulated data suffers from the *identification problem* induced by the exact linear dependency between age, period, and (birth) cohort: $\text{Period} = \text{Age} + \text{Cohort}$. This can be viewed as a special case of collinear regressors that produces, in this case, a singular matrix (of one less than full rank) used in the statistical estimation process. Since a singular matrix produces multiple estimators of the three effects, it is difficult to identify the unique true age, period, and cohort effects.

A number of methodological contributions to the specification and estimation of APCMC models have been developed in a wide variety of disciplines, including social and demographic research (e.g., [6, 7, 14, 18, 32, 49]) and biostatistics and epidemiology (e.g., [5, 19, 35, 37, 46, 45]). Developments in APCMC methodology in biostatistics emphasized the utility of *estimable functions* that are invariant to the selection of constraints on parameters [5, 19, 22, 21, 38]. This is the approach

applied by Fu [10] in the derivation of a new APC estimator — termed the *intrinsic estimator* (IE).

Yang, Fu, and Land [51] compared two approaches to the identification problem in APCMC models, namely, the IE method and the *constrained generalized linear models* (CGLM) estimator that has been conventional among demographers and other social scientists for more than two decades [6, 7, 27]. Based on their results, Yang et al. [51] concluded that the IE offered a useful alternative to conventional methods for the APC analysis of tables of rates. Yang, Schulhofer-Wohl, Fu, and Land [53] further described the IE algebraically, geometrically, and verbally, reported results of model validation assessments of the IE both from empirical examples and simulation experiments, and introduced a software package that interested users can readily access. While there have been extensive debates over statistical properties of the IE [24, 33] and its applicability in the APC context [25], more recent studies possibly suggest that the IE remains a useful tool for estimating age, period, cohort effects [11, 8, 9].

The publication of these and related articles on the IE for APCMC models stimulated considerable commentary in the past decade as well as the development of related statistical estimators. The purpose of this paper is to investigate the linear algebra and statistical properties of the IE and related estimators¹ of the APCMC model, provide further mathematical proofs that are essential but currently lacking in the existing literature, assess fundamental perspectives in the application of the IE and related estimators, and suggest future directions for the IE approach to APC analysis.

2. Algebra of the APCMC Model and the Intrinsic and Related Estimators. The APCMC model can fall into the class of generalized linear models (GLM; see McCullagh and Nelder [29] or McCulloch and Searle [30] for expositions) and take a *log-linear regression form* via a *log link* as:

$$\log(E_{ij}) = \log(P_{ij}) + \mu + \alpha_i + \beta_j + \gamma_k, \quad (1)$$

where E_{ij} denotes the expected number of deaths in cell (i, j) that is assumed to be distributed as a Poisson variate, and $\log(P_{ij})$ is the logarithm of the exposure P_{ij} in (1) and is called the “offset” or an adjustment for the log-linear contingency table model. μ denotes the intercept (e.g., adjusted mean death rate), α_i denotes the i -th row age effect or the coefficient for the i -th age group; β_j denotes the j -th column period effect or the coefficient for the j -th time period; γ_k denotes the k -th diagonal cohort effect or the coefficient for the k -th cohort for $k = 1, \dots, (a + p - 1)$ cohorts, with $k = a - i + j$. Models of this type are widely used in demography and epidemiology where the counts of demographic events such as deaths or the incidence of diseases generally follow Poisson distributions and the rates are estimated through log-linear models [1].

Regression model (1) can be treated as *fixed-effects generalized linear models* after a re-parametrization to have centered parameters:

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0. \quad (2)$$

¹As noted later, several approaches to estimation of the APCMC model based on different principles of statistical estimation yield estimates that are equivalent to the IE. We thus refer to them throughout as related estimators.

After re-parameterization in (4), these APCMC models can be written in the conventional matrix form of a least-squares regression:

$$Y = Xb + \varepsilon, \quad (3)$$

where Y is a vector of mortality rates or log-transformed rates, X is the regression design matrix consisting of dummy-variable column vectors for the vector of model parameters b , which is of dimension $m = 1 + (a - 1) + (p - 1) + (a + p - 2)$:

$$b = (\mu, \alpha_1, \dots, \alpha_{a-1}, \beta_1, \dots, \beta_{p-1}, \gamma_1, \dots, \gamma_{a+p-2})^T. \quad (4)$$

The ordinary least squares/maximum likelihood estimator (OLS/MLE) of the matrix regression model (3) is the solution b of the normal equations:

$$\hat{b} = (X^T X)^{-1} X^T Y. \quad (5)$$

But this estimator *does not exist* (i.e., there is no uniquely defined vector of coefficient estimates). This is because the design matrix X is singular with one less than full column rank (Kupper et al. [21]), which further originates from the perfect linear relationship between the age, period and cohort effects:

$$\text{Period} - \text{Age} = \text{Cohort}.$$

Therefore, $(X^T X)^{-1}$ does not exist. This is the *model identification problem* of APC analysis. It implies that there are an infinite number of possible solutions of the matrix Equation (5), one for each possible linear combination of column vectors that result in a vector identical to one of the columns of X . Therefore, it is not possible to estimate the separate effects of cohort, age, and period without imposing at least one constraint on the coefficients in addition to the re-parameterization in (2). Since the work of Fienberg and Mason [6, 7], the conventional approach to APCMC models has been a *coefficients-constraints approach*, which takes the form of placing (at least) one additional identifying constraint on the parameter vector in (4), e.g., constraining the effect coefficients of the first two periods to be equal, $\beta_1 = \beta_2$. With this additional constraint, the model (3) is just-identified, the matrix $X^T X$ becomes non-singular, and the least-squares estimator in (5) exists.

Since the design matrix X of the unconstrained APCMC model is one less than full column rank, the parameter space of the model can be decomposed into the direct sum of two linear subspaces that are perpendicular to each other (Yang et al. 2004, 2008). One subspace corresponds to the unique zero eigenvalue of the matrix $X^T X$ of Equation (5) and is of dimension one; it is termed the *null space* of the design matrix X ,² and the other subspace is the complement subspace orthogonal to the null space and is of dimension one less than the number of columns of the design matrix.

Due to this orthogonal decomposition of the parameter space, each of the infinite number of solutions of the unconstrained APC accounting model can be written as

$$\hat{b} = B + sB_0, \quad (6)$$

where s is a scalar corresponding to a specific solution and B_0 is a unique eigenvector of the Euclidean norm or length one (Yang et al. [51, 53]). The eigenvector B_0 does not depend on the observed rates Y , it only depends on the design matrix X and thus is completely determined by the numbers of age groups and period groups regardless of the event rates. It is important to note that the vector B_0 is fixed or

²Since the design matrix X of the APCMC model is of rank one less than the number of columns of X , the null space of X is spanned by a single column vector, the *null vector* B_0 .

non-random, because this vector is a function solely of the dimension of the design matrix X , or the number of age groups (a) and periods (p). The fact that the fixed vector B_0 is independent of the response variable Y suggests that it should not play any role in the estimation of effect coefficients. But this principle is frequently violated in the conventional CGLM approach if the scalar s in Equation (6) is nonzero [21, 51, 53]. This is a key point, as intuition suggests that the eigenvector corresponding to the zero eigenvalue should be an arbitrary vector. And, indeed, sB_0 is arbitrary. On the other hand, B_0 is not arbitrary; it is fixed by the design matrix. Furthermore, as indicated in Equation (6):

- any APC estimator, obtained by placing any identifying constraint(s) on the design matrix can be written as a linear combination $B + sB_0$,
- where B is the *intrinsic estimator (IE)* that lies in the parameter subspace that is orthogonal to the null space.

The IE B in Equation (6) is determined by the Moore-Penrose generalized inverse [12].³

Corresponding to the decomposition of the estimators in equation (6), the unconstrained parameter vector (4) can be decomposed into

$$b = b_0 + sB_0, \quad (7)$$

where the parameter vector $b_0 = P_{\text{proj}}b$ is a linear map of b corresponding to the projection of the unconstrained parameter vector (6) to the row space of X [51, 53]. Specifically, the constrained parameter vector b_0 corresponding to $s = 0$ satisfies the geometric projection:

$$b_0 = (I - B_0B_0^T)b, \quad (8)$$

where b is the parameter vector (6). Since B_0 is a normalized vector orthogonal to b_0 , Equation (8) holds because $B_0^Tb_0 = 0$ and $B_0^TB_0 = 1$ such that

$$\begin{aligned} (I - B_0B_0^T)b &= (I - B_0B_0^T)(b_0 + sB_0) = b_0 - B_0B_0^Tb_0 + sB_0 - B_0B_0^TsB_0 \\ &= b_0 - 0 + sB_0 - sB_0 = b_0. \end{aligned}$$

Yet, it is also important to understand Equation (8) from a linear algebra perspective. Since $B_0(B_0^TB_0)^{-1}B_0^T$ defines a projection matrix which projects b onto the null vector B_0 , we obtain b_0 once b subtracts its component in the direction of B_0 :

$$\begin{aligned} b_0 &= b - sB_0 = b - P_{\text{proj}}b = b - B_0(B_0^TB_0)^{-1}B_0^Tb \\ &= b - B_0B_0^Tb = (I - B_0B_0^T)b. \end{aligned} \quad (9)$$

This projection is illustrated geometrically in Yang and Land [52]. The projection is independent of the arbitrary real number/scalar s . Corresponding to the projection of the parameter vector b onto b_0 , we have the following projection of the estimators of equation (8) onto the intrinsic estimator B (e.g., [10, 51]):

$$B = (I - B_0B_0^T)\hat{b}. \quad (10)$$

In brief, the basic idea of the IE is to remove the influence of the null space of the design matrix on coefficient estimates [51, 53]. This is done by projecting the APCMC coefficient vector b onto the non-null space of the design matrix X , which is equivalent to setting the scalar s (through which the design matrix affects

³See, e.g., Searle [41, pages 16-19] for a definition of the Moore-Penrose generalized inverse and its properties. In some expositions of generalized inverse matrices (e.g., [3]) the Moore-Penrose generalized inverse is termed a pseudoinverse.

the vector b) in Equation (7) equal to zero and yields the constrained parameter vector b_0 to be estimated. Corresponding to this projection of b onto b_0 , the IE B is obtained by projecting any equality constrained APCMC estimator \hat{b} onto the non-null space of X . The IE also can be viewed as a special form of principal components regression estimator⁴ that removes the influence of the null space of the design matrix X on the estimator. In order to facilitate interpretability of the estimated A, P, and C coefficients, the IE uses an extra step of inverse orthonormal transformation of the coefficient estimates of the principal components regression back to the original space of age, period, and cohort coordinates.

Remark 1 (Related Estimators). *As O'Brien [34, page 88] noted, several other approaches to estimation of the APCMC model yield estimates that are equivalent to those estimated by the Moore-Penrose generalized inverse matrix and the IE. This includes the principal components estimator of Kupper et al. [21], the singular value decomposition solution [36], the partial least squares estimator of Tu et al. [47], and the maximum entropy estimator of Browning et al. [4]. And while Fu [10] derived the IE as the limit of a ridge regression estimator for singular design matrices as the shrinkage penalty diminishes, Xu and Powers [50] showed that a Bayesian ridge regression model with a common prior for the ridge parameter yields estimates of age, period, and cohort effects similar to those based on the IE and to those based on a ridge estimator.*

3. The Linear Algebra of the Intrinsic Estimator. Given the foregoing definitions and algebraic properties, we now place the intrinsic and related estimators in a broader context of matrix analysis and demonstrate their essential (albeit often overlooked) link with established regression models. More specifically, we prove:

Theorem 1. *The IE is an estimator of principal component regression (PCR) after the principal component corresponding to eigenvalue 0 is removed.*

Proof. To express principal components of variables in the design matrix, we first perform a singular value decomposition of $X_{n \times m}$, which has n rows (observations) and m columns (variables):

$$X_{n \times m} = P_{n \times n} \Delta_{n \times m} Q_{m \times m}^T, \quad (11)$$

where $P_{n \times n} = (p_1, p_2, \dots, p_n)$ contains n eigenvectors of XX^T , entries of the diagonal matrix $\Delta_{n \times m}$ are the singular values of $X_{n \times m}$, and $Q_{m \times m} = (q_1, q_2, \dots, q_m)$ contains m eigenvectors of $X^T X$. Since $Q_{m \times m}$ is an orthogonal matrix, we have

$$Q^T Q = Q Q^T = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} = I, \quad (12)$$

where the transpose of Q is also the inverse of Q , i.e., $Q^T = Q^{-1}$. To describe the relations among the square (symmetric) matrix $X^T X$, its eigenvectors and

⁴For a standard exposition of principal components regression, see, e.g., [42].

eigenvalues, we have [44]:

$$\begin{aligned} X^T X Q &= X^T X (q_1, q_2, \dots, q_m) = (X^T X q_1, X^T X q_2, \dots, X^T X q_m) \\ &= (q_1, q_2, \dots, q_m) \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{pmatrix} = Q \Lambda, \end{aligned} \quad (13)$$

where λ_i is the corresponding eigenvalue of the eigenvector q_i and $\Lambda = \Delta^T \Delta$ is a diagonal matrix consisting of all these eigenvalues. Or,

$$X^T X = Q \Lambda Q^{-1}. \quad (14)$$

Since $Q^T = Q^{-1}$ and $Q^{-1}Q = I$, each side of Equation (14) can be pre-multiplied by Q^T and post-multiplied by Q to yield:

$$Q^T X^T X Q = Q^{-1} Q \Lambda Q^{-1} Q = \Lambda = Z^T Z, \quad (15)$$

where

$$\begin{aligned} Z &= X Q = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} (q_1 \quad q_2 \quad \dots \quad q_m) \\ &= \begin{pmatrix} \langle x_1, q_1 \rangle & \langle x_1, q_2 \rangle & \dots & \langle x_1, q_m \rangle \\ \langle x_2, q_1 \rangle & \langle x_2, q_2 \rangle & \dots & \langle x_2, q_m \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle x_n, q_1 \rangle & \langle x_n, q_2 \rangle & \dots & \langle x_n, q_m \rangle \end{pmatrix} \end{aligned}$$

actually projects the original design matrix $X_{n \times m}$ onto eigenvectors of $X^T X$ (or right-singular vectors of $X_{n \times m}$) and produces principal components of X . In this regard, eigenvectors q_i are also known as principal component directions of X [17]. If we rewrite the design matrix $X_{n \times m}$ in terms of its principal components, Equation (5) can now be rewritten as:

$$Y = X b + \varepsilon = X Q Q^T b + \varepsilon = Z b_Z + \varepsilon, \quad (16)$$

where $Z = X Q$ and $b_Z = Q^T b$ are the new design matrix and the new coefficient vector, respectively. The regression estimator of Equation (16) can be denoted by Moore-Penrose generalized inverse matrices (see the Appendix) as:

$$\begin{aligned} \hat{b}_Z &= Q^T \hat{b} = Q^T (X^T X)^\dagger X^T Y = Q^T (Q \Lambda Q^T)^\dagger X^T Y \\ &= Q^T Q \Lambda^\dagger Q^T X^T Y = \Lambda^\dagger Q^T X^T Y \\ &= (Q^T X^T X Q)^\dagger Q^T X^T Y = (Z^T Z)^\dagger Z^T Y. \end{aligned} \quad (17)$$

Once \hat{b}_Z is calculated, the PCR estimator \hat{b}_{PCR} with all the principal components can be obtained by:

$$\hat{b}_{\text{PCR}} = Q \hat{b}_Z = Q Q^T \hat{b}. \quad (18)$$

Given that $Q Q^T = I$, it is easy to see $\hat{b}_{\text{PCR}} = \hat{b}$. Considering the eigenvector in the null space of $X_{n \times m}$, one can address the identification problem by removing

the eigenvector B_0 from $Q_{m \times m}$ and then perform principal-component-regression analysis. More specifically, we define:

$$Q_* = (q_2, \dots, q_m), \quad (19)$$

where without loss of generality we assume that the first eigenvector q_1 is associated with eigenvalue $\lambda_1 = 0$. Correspondingly, the PCR estimator becomes the following since Λ_* is now nonsingular:

$$\hat{b}_{\text{PCR}*} = Q_* \hat{b}_{Z*} = Q_* \Lambda_*^{-1} Q_*^T X^T Y. \quad (20)$$

Next, we prove that $\hat{b}_{\text{PCR}*}$ is the intrinsic estimator B , namely, $\hat{b}_{\text{PCR}*} = (I - B_0 B_0^T) \hat{b}$ in the following two steps:

- (1). $\hat{b}_{\text{PCR}*}$ is a least squares solution to equation (5). Since $\lambda_1 = 0$, we have

$$Q_* \Lambda_*^{-1} Q_*^T = \sum_{i=2}^m \lambda_i^{-1} q_i q_i^T = \sum_{i=2}^m \lambda_i^{-1} q_i q_i^T + 0 q_1 q_1^T = Q \Lambda^\dagger Q^T,$$

and $\hat{b}_{\text{PCR}*}$ is thus a least squares solution to equation (5):

$$\hat{b}_{\text{PCR}*} = Q_* \Lambda_*^{-1} Q_*^T X^T Y = Q \Lambda^\dagger Q^T X^T Y = (X^T X)^\dagger X^T Y.$$

- (2). Next, we further show that $\hat{b}_{\text{PCR}*} = (I - B_0 B_0^T) \hat{b}_{\text{PCR}*}$. This equation holds when $B_0^T \hat{b}_{\text{PCR}*} = 0$. Since $B_0^T q_i = 0$ for any $i \geq 2$, we have $B_0^T \hat{b}_{\text{PCR}*} = B_0^T Q_* \Lambda_*^{-1} Q_*^T X^T Y = 0$.

Therefore, we conclude that $\hat{b}_{\text{PCR}*} = (I - B_0 B_0^T) \hat{b}_{\text{PCR}*} = (I - B_0 B_0^T) \hat{b}$ and $\hat{b}_{\text{PCR}*}$ is the intrinsic estimator B . \square

4. Statistical Properties of the Intrinsic and Related Estimators. The previous section on the linear algebra of the intrinsic estimators clearly shows the essential link between the IE and an established regression method, principal component regression, which has been widely applied in areas such as ecology, chemistry and genomics for years (e.g., [2, 16, 31]). Given that the recent discussion on the IE has been characterized by sometimes heated debates about its statistical properties [23, 24, 33], we further critically investigate statistical properties of the IE based on the mathematical proof of Theorem 1 presented above.

The fact that these various principles of statistical estimation yield estimates of the coefficient vectors of the APCMC model that are equivalent suggests that they are identifying certain fundamental features of the resulting coefficient vector estimates. For context, we consider an APC dataset for a finite number of time periods, p . That is, suppose that an APCMC analysis is to be conducted for a fixed matrix of observed rates or event counts. This implies that the corresponding design matrix X is fixed (i.e., X has a fixed number of age groups and time periods).

Property 1 – Estimable: For a finite number of time periods p , the IE and related estimators of the APCMC model are estimable, where *estimable functions*⁵ are invariant with respect to whatever solution is obtained to the normal equations.⁶

⁵See [41, pages 180-188] or [30, pages 120-121] for expositions of this concept.

⁶In the history of discussions of the APC accounting model in sociology, Rodgers [39] was early to argue that analysts should seek estimable functions of the unidentified parameter vector (Equation (6)); see also the comment by Smith, Mason, and Fienberg [43] and the response by Rodgers [40]. In some respects, the IE can be regarded as providing a practical method for calculating estimates of estimable functions from data in the form of age by time period tables of rates, as called for by Rodgers over two decades ago. The estimability referred by Rodgers,

Remark 2. *This property of the IE and related estimators of the APCMC accounting model has been a source of contention. Yang et al. [51, page 101] stated that the IE satisfies a condition for estimability of linear functions of the parameter vector b that was established by Kupper et al. [21, Appendix B]. Specifically, the condition for estimability of a constraint on the parameter vector that was established algebraically by Kupper et al. [21] is, in the notation defined above, that $l^T B_0 = 0$, where l^T is a constraint vector (of appropriate dimension) that defines a linear function $l^T b$ of b . Yang et al. [51, page 101] noted the IE imposes the constraint that $s = 0$, i.e., that the arbitrary vector B_0 has zero influence, $l^T = (I - B_0 B_0^T)$ for the IE and since $B_0^T B_0 = 1$, it then follows that*

$$l^T B_0 = (I - B_0 B_0^T) B_0 = B_0 - B_0 B_0^T B_0 = B_0 - B_0 = 0, \quad (21)$$

i.e., the Kupper et al. condition holds for the IE. We have (a) Any coefficient equality constraint on an APCMC model suffices to identify the model and is associated with a generalized inverse matrix that yields a corresponding estimated coefficient vector. (b) All such generalized inverse matrices correspond to a single, unique Moore-Penrose generalized inverse matrix that is associated with the IE and related estimators of a specific APCMC model with a corresponding specific design matrix. Thus, these estimators meet the estimability condition of being invariant to the constraint used to solve the normal equations. Finally, it should be noted that, using linear model theory, a proof of the estimability of the IE has recently been provided by Fu [11].

Property 2 – Unbiased: For a finite number of time periods p , the IE and related estimators of the APCMC model are *unbiased estimators* of the parameter vector $b_0 = P_{\text{proj}} b$ defined above in Equation (12), that is, of the projection of the unidentified coefficient vector of the APCMC model onto the non-null space of the design matrix X .

This property follows from Property 1 and the property of estimable functions that they are linear functions of the unidentified parameter vector that can be estimated without bias, i.e., the Moore-Penrose generalized inverse matrix yields unbiased estimates of b_0 (see Yang et al. [51, page 107]). Note that this unbiased estimator property is specific to the projection of the unidentified coefficient vector b of the APCMC model onto the vector b_0 in the non-null space of the design matrix X .

Property 3 – Relative Efficiency: For a finite number of time periods p , the IE and related estimators of the APCMC model have a *variance smaller than that of any Constrained Generalized Linear Model APCMC estimator \hat{b}* , i.e., $\text{Var}(\hat{b}) - \text{Var}(B)$ is non-negative for a nontrivial identifying constraint that corresponds to \hat{b} . This property states that the IE is relatively efficient compared to any possible CGLM-class estimator. It was proven in Yang et al. [51, page 108].

Property 4 – Minimum Quadratic Norm and Best Approximate Solution: (a) For a finite number of time periods p , the IE and related estimators of the APCMC model are *minimum quadratic norm estimators*, where the quadratic norm of an estimated coefficient vector of the APCMC model is defined as the square root of the sum of the estimated elements of the coefficient vector. (b) For a finite number of time periods, the IE and related estimators of the APCMC

however, essentially means identifiability that can be achieved by any linear constraints, which differs from statistical estimability defined for APC models by Kupper et al. [21].

model are the *best approximate solutions* of the APCMC model (see Appendix for a proof).

Remark 3. *The minimum quadratic norm (Property 4a) is a property of the Moore-Penrose generalized inverse matrix of the unidentified APCMC model design matrix X [13, page 238]. The best approximate solution property (4b) responds to an observation of Norval Glenn [15, page 20], a long-time critic of attempts to provide general solutions to the APCMC identification and estimation problem stated that such a general method “may prove to be useful ...if it yields approximately correct estimates ‘more often than not,’ if researchers carefully assess the credibility of the estimates by using theory and side information, and if they keep their conclusions about the effects tentative”.*⁷

Property 5 – Uniqueness: The IE and related estimators, that is, estimators that are based on the Moore-Penrose generalized inverse matrix, are the *unique solutions* to the estimation of the APCMC model that satisfy the normal equations of the model (Equation (7) above) and the collinearity constraints of the covariates of the model (i.e., Age – Period + Cohort = 0). This property was stated and proved by Tu et al. [47].

Remark 4. *In age-period-cohort analysis, the asymptotic properties of estimators raise two different yet related questions. First, as the time periods of observations increase without bound, will an estimator converge to the “true” population parameter? Second, with an increase in the number of observations, does an estimator converge towards the “true” population parameter? For the first question, it has been adequately proven in a recent article that as the number of time periods increases without bound, the IE produces consistent estimation of population parameters [11, Section 4]. The second question is nevertheless more relevant to typical empirical applications of the APCMC model since only a finite and usually moderate number of time periods of data are available for analysis. The IE’s consistency as the number of observations increases can be demonstrated as follows:*

$$\begin{aligned} B &= Q_* \Lambda_*^{-1} Q_*^T X^T Y = Q_* \Lambda_*^{-1} Q_*^T X^T (Xb + \varepsilon) \\ &= (X^T X)^\dagger X^T Xb + (X^T X)^\dagger X^T \varepsilon = b_0 + (X^T X)^\dagger X^T \varepsilon, \end{aligned} \quad (22)$$

where $(X^T X)^\dagger$ denotes the Moore-Penrose generalized inverse of $X^T X$. Since $X^T \varepsilon$ converges to zero as $N \rightarrow \infty$ under the strict exogeneity assumption and $(X^T X)^\dagger$ does not depend on N , we know from the Slutsky’s theorem (see, e.g., [48, page 11]) that $(X^T X)^\dagger X^T \varepsilon$ also converges to zero as $N \rightarrow \infty$, which shows the consistency of IE as sample size increases.

5. Conclusions and Discussion. In this paper, we have reviewed and further demonstrated the mathematical foundations and statistical properties of the Intrinsic Estimator for Age-Period-Cohort Multiple Classification models that originally was derived as the limit of a ridge regression estimator for singular design matrices as the shrinkage penalty diminishes [10]. We also have noted that the IE is essentially equivalent to estimators of the APCMC model that have been derived by applications of various generally accepted principles of estimation of statistical models, including the principal components estimator, the singular value estimator, the partial least squares estimator, the maximum entropy estimator, and Bayesian

⁷For extensive simulation analyses that demonstrate the veracity of Glenn’s observation for the IE, see Jeon [20].

ridge regression. In particular, we illustrate the inherent link between the IE and principal-component-regression estimators.

The IE is a sophisticated yet practical age-period-cohort method. It is useful not because its concept is new. On the contrary, the IE is inherently related to classical statistical tools that have been employed by scholars for decades. By and large, existing critiques of the IE focus on two methodological possibilities, namely, whether the IE method per se is valid, and/or whether it is valid to apply the IE to APC data. For the first issue, the statistical properties of the IE (e.g., unbiasedness and consistency) have been questioned [24, 33]; for the second issue, it has been suggested that the IE is sensitive to the choice of coding schemes in age-period-cohort analysis [9, 25]. Because a flawed method and an inappropriate application of a valid method can both contribute to misleading results, the ongoing debates over the validity of the IE exactly highlight our failure in ruling out either one of these two methodological possibilities. By thoroughly investigating its statistical properties, this research demonstrates that the IE per se remains a valid estimator. Once we rule out the first methodological possibility, new IE-related methods that are specifically designed to address the second methodological challenge can provide a reasonable solution. In particular, the Orthogonal Estimator [8, 9], which is not affected by the choice of coding schemes, can provide a promising alternative when reasonable assumptions are employed following Glenn’s [15] recommendation to use “theory and side information” as a basis for APC analysis and perform good enough to meet his criterion of “approximately correct estimates, ‘more often than not’ ”.

Moreover, a key lesson we learn from this debate is that a statistically sound tool does not readily render itself as a perfect or universal solution in empirical research. Results of statistical models are also a function of procedures, algorithms and metrics chosen by scholars, which may not reflect different social realities at stake: 1.609344 kilometers and 1 mile may look different but they refer to the same geographic distance. It is also noteworthy to mention that parametric estimation with constraints can hardly produce a satisfactory solution to the nonlinearity issue in age-period-cohort analysis; instead, non-parametric methods such as kernel density estimation may be considered to address this issue. Likewise, more powerful statistical packages, such as the “glm” function in R, may provide a more viable way for scholars to overcome challenges in existing tools for computing the IE. Since mathematical proof is the ultimate arbiter of methodological truth, this study critically investigates the linear algebra and statistical properties of the IE to provide a more holistic understanding of this method.

REFERENCES

- [1] A. Agresti, *An introduction to categorical data analysis*, John Wiley & Sons, 2019.
- [2] A.-L. Boulesteix and K. Strimmer, Partial least squares: a versatile tool for the analysis of high-dimensional genomic data, *Briefings in Bioinformatics*, **8** (2006), 32–44.
- [3] T. L. Boullion and P. L. Odell, *Generalized inverse matrices*, Wiley-Interscience [A Division of John Wiley & Sons, Inc.], New York-London-Sydney, 1971.
- [4] M. Browning, I. Crawford and M. Knoef, *The age-period cohort problem: set identification and point identification*, Technical report, cemmap working paper, 2012.
- [5] D. Clayton and E. Schifflers, Models for temporal variation in cancer rates. ii: Age-period-cohort models, *Statistics in Medicine*, **6** (1987), 469–481.
- [6] S. E. Fienberg and W. M. Mason, Identification and estimation of age-period-cohort models in the analysis of discrete archival data, *Sociological Methodology*, **10** (1979), 1–67.
- [7] S. E. Fienberg and W. M. Mason, *Specification and Implementation of Age, Period and Cohort Models*, 45–88, Springer New York, New York, NY, 1985.

- [8] E. Fosse and C. Winship, Moore-penrose estimators of age-period-cohort effects: Their inter-relationship and properties, *Sociological Science*, **5** (2018), 304–334.
- [9] E. Fosse and C. Winship, Analyzing age-period-cohort data: A review and critique, *Annual Review of Sociology*, **45** (2019), 467–492.
- [10] W. J. Fu, Ridge estimator in singular design with application to age-period-cohort analysis of disease rates, *Communications in Statistics - Theory and Methods*, **29** (2000), 263–278.
- [11] W. J. Fu, Constrained estimators and consistency of a regression model on a Lexis diagram, *Journal of the American Statistical Association*, **111** (2016), 180–199.
- [12] W. J. Fu and P. Hall, Asymptotic properties of estimators in age-period-cohort analysis, *Statistics & Probability Letters*, **76** (2006), 1925–1929.
- [13] F. Girosi and G. King, *Demographic Forecasting*, Princeton University Press, 41 William Street, Princeton, New Jersey 08540, 2008.
- [14] N. D. Glenn, Cohort analysts’ futile quest: Statistical attempts to separate age, period and cohort effects, *American Sociological Review*, **41** (1976), 900–904.
- [15] N. D. Glenn, *Cohort Analysis*, 2nd edition, SAGE Publications, Inc., Thousand Oaks, California, 2005.
- [16] M. H. Graham, Confronting multicollinearity in ecological multiple regression, *Ecology*, **84** (2003), 2809–2815.
- [17] T. Hastie, R. Tibshirani and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, 2nd edition, Springer Science & Business Media, 2016.
- [18] J. Hobcraft, J. Menken and S. Preston, Age, period, and cohort effects in demography: A review, *Population Index*, **48** (1982), 4–43.
- [19] T. R. Holford, The estimation of age, period and cohort effects for vital rates, *Biometrics*, **39** (1983), 311–324.
- [20] S. Y. Jeon, *Do Data Structures Matter? A Simulation Study for Testing the Validity of Age-Period-Cohort Models*, PhD thesis, Utah State University, 2017.
- [21] L. L. Kupper, J. M. Janis, A. Karmous and B. G. Greenberg, Statistical age-period-cohort analysis: A review and critique, *Journal of Chronic Diseases*, **38** (1985), 811 – 830.
- [22] L. L. Kupper, J. M. Janis, I. A. Salama, C. N. Yoshizawa, B. G. Greenberg and H. H. Winsborough, Age-period-cohort analysis: an illustration of the problems in assessing interaction in one observation per cell data, *Communications in Statistics - Theory and Methods*, **12** (1983), 201–217.
- [23] K. C. Land, Q. Fu, X. Guo, S. Y. Jeon, E. N. Reither and E. Zang, Playing With the Rules and Making Misleading Statements: A Response to Luo, Hodges, Winship, and Powers, *American Journal of Sociology*, **122** (2016), 962 – 973.
- [24] L. Luo, Assessing validity and application scope of the intrinsic estimator approach to the age-period-cohort problem, *Demography*, **50** (2013), 1945–1967.
- [25] L. Luo, J. Hodges, C. Winship and D. Powers, The sensitivity of the intrinsic estimator to coding schemes: Comment on yang, schulhofer-wohl, fu, and land, *American Journal of Sociology*, **122** (2016), 930–961.
- [26] K. O. Mason, W. M. Mason, H. H. Winsborough and W. K. Poole, Some methodological issues in cohort analysis of archival data, *American Sociological Review*, **38** (1973), 242–258.
- [27] W. M. Mason and H. L. Smith, *Age-Period-Cohort Analysis and the Study of Deaths from Pulmonary Tuberculosis*, 151–227, Springer New York, New York, NY, 1985.
- [28] W. Mason and N. Wolfinger, Cohort analysis, in *International Encyclopedia of the Social & Behavioral Sciences* (eds. N. J. Smelser and P. B. Baltes), Pergamon, Oxford, 2001, 2189 – 2194.
- [29] P. McCullagh and J. Nelder, *Generalized Linear Models*, Chapman and Hall, 1989.
- [30] C. E. McCulloch and S. R. Searle, *Generalized, linear, and mixed models*, Wiley Series in Probability and Statistics: Texts, References, and Pocketbooks Section, Wiley-Interscience [John Wiley & Sons], New York, 2001.
- [31] T. Næs and H. Martens, Principal component regression in NIR analysis: Viewpoints, background details and selection of components, *Journal of Chemometrics*, **2** (1988), 155–167.
- [32] R. M. O’Brien, Age period cohort characteristic models, *Social Science Research*, **29** (2000), 123 – 139.
- [33] R. M. O’Brien, Constrained estimators and age-period-cohort models, *Sociological Methods & Research*, **40** (2011), 419–452.
- [34] R. M. O’Brien, Age-period-cohort models and the perpendicular solution, *Epidemiologic Methods*, **4** (2015), 87–99.

- [35] C. Osmond and M. J. Gardner, Age, period and cohort models applied to cancer mortality rates, *Statistics in Medicine*, **1** (1982), 245–259.
- [36] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical recipes: The art of scientific computing*, 3rd edition, Cambridge University Press, 2007.
- [37] C. Robertson and P. Boyle, Age-period-cohort analysis of chronic disease rates. i: modelling approach, *Statistics in Medicine*, **17** (1998), 1305–1323.
- [38] C. Robertson, S. Gandini and P. Boyle, Age-period-cohort models: A comparative study of available methodologies, *Journal of Clinical Epidemiology*, **52** (1999), 569–583.
- [39] W. L. Rodgers, Estimable functions of age, period, and cohort effects, *American Sociological Review*, **47** (1982), 774–787.
- [40] W. L. Rodgers, Reply to comment by smith, mason, and fienberg, *American Sociological Review*, **47** (1982), 793–796.
- [41] S. R. Searle, *Linear models*, John Wiley & Sons, Inc., New York-London-Sydney, 1971.
- [42] A. Sen and M. Srivastava, *Regression analysis: theory, methods, and applications*, Springer-Verlag, New York, NY, USA, 1990.
- [43] H. L. Smith, W. M. Mason and S. E. Fienberg, More chimeras of the age-period-cohort accounting framework: Comment on rodgers, *American Sociological Review*, **47** (1982), 787–793.
- [44] G. Strang, *Introduction to linear algebra*, 5th edition, Wellesley-Cambridge Press Wellesley, MA, USA, 2016.
- [45] R. Tarone and K. Chu, Age-period-cohort analyses of breast-, ovarian-, endometrial- and cervical-cancer mortality rates for caucasian women in the USA, *Journal of Epidemiology and Biostatistics*, **5** (2000), 221–231.
- [46] R. E. Tarone and K. C. Chu, Implications of Birth Cohort Patterns in Interpreting Trends in Breast Cancer Rates, *JNCI: Journal of the National Cancer Institute*, **84** (1992), 1402–1410.
- [47] Y.-K. Tu, N. Krämer and W.-C. Lee, Addressing the identification problem in age-period-cohort analysis: a tutorial on the use of partial least squares and principal components analysis, *Epidemiology*, **23** (2012), 583–593.
- [48] A. W. van der Vaart, *Asymptotic statistics*, vol. 3 of Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, UK, 1998.
- [49] J. R. Wilmoth, Variation in vital rates by age, period, and cohort, *Sociological Methodology*, **20** (1990), 295–335.
- [50] M. Xu and D. A. Powers, *Bayesian Ridge Estimation of Age-Period-Cohort Models*, 337–359, in *Dynamic Demographic Analysis*, Springer International Publishing, Cham, Switzerland, 2016.
- [51] Y. Yang, W. J. Fu and K. C. Land, A methodological comparison of age-period-cohort models: The intrinsic estimator and conventional generalized linear models, *Sociological Methodology*, **34** (2004), 75–110.
- [52] Y. Yang and K. C. Land, *Age-period-cohort analysis: New models, methods, and empirical applications*, Chapman and Hall/CRC, Boca Raton, FL, USA, 2013.
- [53] Y. Yang, S. Schulhofer-Wohl, W. J. Fu and K. C. Land, The intrinsic estimator for age-period-cohort analysis: What it is and how to use it, *American Journal of Sociology*, **113** (2008), 1697–1736.

Appendix A. Generalized Inverses and Intrinsic Estimators. For a real matrix A , in the following matrix equations (adopted from Boullion and Odell [3]), $(\cdot)^\dagger$ is used to denote generalized inverse and $(\cdot)^T$ denotes the matrix transpose.

$$AXA = A, \quad (23)$$

$$XAX = X, \quad (24)$$

$$(XA)^T = XA, \quad (25)$$

$$(AX)^T = AX. \quad (26)$$

Definition 1. A generalized inverse of a matrix A is a matrix $X = A^g$ satisfying (23).

Definition 2. A reflexive generalized inverse of a matrix A is a matrix $X = A^r$ satisfying (23) and (24).

Definition 3. A left weak generalized inverse of a matrix A is a matrix $X = A^{lw}$ satisfying (23), (24), and (25).

Definition 4. A right weak generalized inverse of a matrix A is a matrix $X = A^{rw}$ satisfying (23), (24), and (26).

Definition 5. A Moore-Penrose generalized inverse of a matrix A is a matrix $X = A^\dagger$ satisfying (23) through (26).

Best Approximate Solution: Adapted from Boullion and Odell [3, pages 42-43], if A is an $m \times n$ real matrix, the notation $\|A\|^2$ denotes the non-negative square root of the sum of squares of the elements of A . Note that $\|A\|^2 = \text{tr}(A^T A)$ and $\|A\| > 0$ unless $A = 0$, then $\|A\| = 0$.

Definition 6. The matrix X_0 is a best approximate solution of the equation $f(X) = G$ if for all X , either

$$\|f(X) - G\| > \|f(X_0) - G\|, \quad (27)$$

or

$$\|f(X) - G\| = \|f(X_0) - G\| \quad \text{and} \quad \|X\| \geq \|X_0\|. \quad (28)$$

Theorem 2. The best approximate solution of the equation $AX = B$ is $X_0 = A^\dagger B$, where A^\dagger denotes the Moore-Penrose generalized inverse matrix.

Proof. It is readily established for matrices P and Q that

$$\|AP + (I - AA^\dagger)Q\|^2 = \|AP\|^2 + \|(I - AA^\dagger)Q\|^2. \quad (29)$$

In particular, then

$$\begin{aligned} \|AX - B\|^2 &= \|A(X - A^\dagger B) + (I - AA^\dagger)(-B)\|^2 \\ &= \|A(X - A^\dagger B)\|^2 + \|(I - AA^\dagger)(-B)\|^2 \\ &= \|AX - AA^\dagger B\|^2 + \|AA^\dagger B - B\|^2 \\ &\geq \|AA^\dagger B - B\|^2. \end{aligned}$$

Equality holds only when $\|AX - AA^\dagger B\| = 0$ or

$$AX = AA^\dagger B. \quad (30)$$

Replacing A by A^\dagger in (29) and using the fact that $AA^\dagger A = A$, it follows that $\|A^\dagger B + (I - A^\dagger A)X\|^2 = \|A^\dagger B\|^2 + \|(I - A^\dagger A)X\|^2$. Then if (30) holds, $A^\dagger AA^\dagger$ gives $\|X\|^2 = \|A^\dagger B\|^2 + \|X - A^\dagger B\|^2$, which is minimal if $\|X - A^\dagger B\| = 0$ or $X - A^\dagger B = 0$ implying $X_0 = A^\dagger B$. \square

Note: Applied to estimation of the APCMC model, the matrix equation of the theorem, $AX = B$, is the normal equations in matrix form of the APCMC model $X^T X b = X^T Y$ and the best approximate solution $X_0 = A^\dagger B$ becomes $b = (X^T X)^{-1} X^T Y = (X^T X)^\dagger X^T Y$, where $(X^T X)^\dagger$ is the Moore-Penrose generalized inverse matrix.

E-mail address: qiang.fu@ubc.ca

E-mail address: xin.guo@uq.edu.au

E-mail address: SunYoung.Jeon@ucsf.edu

E-mail address: eric.reither@usu.edu

E-mail address: emma.zang@yale.edu

E-mail address: kland@duke.edu