DYNAMIC PANEL DATA MODELS WITH IRREGULAR SPACING: WITH AN APPLICATION TO EARLY CHILDHOOD DEVELOPMENT

DANIEL L. MILLIMETa,b* AND IAN K. MCDONOUGHc
a Department of Economics, Southern Methodist University, Dallas, TX, USA
b IZA, Bonn, Germany
c Department of Economics, University of Nevada, Las Vegas, NV, USA

SUMMARY

With the increased availability of longitudinal data, dynamic panel data models have become commonplace. Moreover, the properties of various estimators of such models are well known. However, we show that these estimators break down when the data are irregularly spaced along the time dimension. Unfortunately, this is an increasingly frequent occurrence as many longitudinal surveys are collected at non-uniform intervals and no solution is currently available when time-varying covariates are included in the model. In this paper, we propose two new estimators for dynamic panel data models when data are irregularly spaced and compare their finite-sample performance to the naïve application of existing estimators. We illustrate the practical importance of this issue in an application concerning early childhood development. Copyright © 2016 John Wiley & Sons, Ltd.

Received 11 November 2013; Revised 8 August 2016

Supporting information may be found in the online version of this article.

1. INTRODUCTION

Dynamic panel data (DPD) models were first analyzed in Balestra and Nerlove (1966) and have since become commonplace in economics. Since then, much progress has been made in terms of understanding the properties of various estimators of these models. However, a relatively overlooked issue, and one that is becoming more prominent with the increased availability of longitudinal surveys, pertains to panel data designs with irregular spacing. For example, an often used dataset, the Early Childhood Longitudinal Survey, Kindergarten Cohort (ECLS-K), is a survey of roughly 20,000 children who entered kindergarten in the USA in fall 1998. Information is collected on this sample over seven waves: Fall and Spring Kindergarten, Fall and Spring First Grade, Spring Third Grade, Spring Fifth Grade and Spring Eighth Grade. Thus the first four waves are spaced roughly 6 months apart; waves four, five and six are spaced 2 years apart; and waves six and seven are spaced 3 years apart. Other examples from developed countries are provided in Table I; see also McKenzie (2001). As we demonstrate below, irregularly spaced data generate a host of difficulties for the estimation of DPD models that researchers cannot ignore.

Before continuing, a formal definition of irregular spacing is required as it is broader than the ECLS-K example suggests. A longitudinal survey is irregularly spaced if successive waves do not conform to successive periods as defined by the underlying data-generating process (DGP). In time series, the distance between successive waves is referred to as the observation interval, whereas the unit period denotes the reference unit of time for the underlying process (Fuleky, 2012). Thus irregularly spaced longitudinal surveys are spaced such that the observation interval deviates from the unit period. Importantly, under this definition, longitudinal data collected at uniform intervals but with gaps are

* Correspondence to: Daniel L. Millimet, Department of Economics, Southern Methodist University, Dallas, TX, USA. E-mail: millimet@smu.edu

Copyright © 2016 John Wiley & Sons, Ltd.
Table I. Examples of irregularly spaced longitudinal surveys in developed countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Survey</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>Australian Longitudinal Study on Women's Health (ALSWH)</td>
<td>Waves for different cohorts are separated by 2–4 years</td>
</tr>
<tr>
<td></td>
<td>Longitudinal Study of Australian Children (LSAC)</td>
<td>Biennial from 2003 to present</td>
</tr>
<tr>
<td></td>
<td>Longitudinal Survey of Immigrants to Australia</td>
<td>Wave 1 covers first 6 months post-immigration; Wave 2 covers 6–18 months post-immigration; Wave 3 covers 18–42 months post-immigration</td>
</tr>
<tr>
<td>Canada</td>
<td>National Longitudinal Survey of Children and Youth (NLSCY)</td>
<td>Biennial from 1994 to 2009</td>
</tr>
<tr>
<td>France</td>
<td>French Longitudinal Study of Children (ELFE)</td>
<td>Individuals at birth, 2 months old, 10 months old, annual until 5 years old, then every 2–3 years thereafter</td>
</tr>
<tr>
<td>Japan</td>
<td>Nihon University Japanese Longitudinal Study of Aging (NUJLSOA)</td>
<td>Biennial for first three waves; Waves 3 and 4 separated by 3 years</td>
</tr>
<tr>
<td>UK</td>
<td>1958 National Child Development Study (NCDS)</td>
<td>Individuals at birth, 7 years old, 11 years old, 16 years old, 23 years old, 33 years old, 42 years old, 46 years old, 50 years old and 55 years old</td>
</tr>
<tr>
<td></td>
<td>1970 British Cohort Study (BCS70)</td>
<td>Individuals at birth, 5 years old, 10 years old, 16 years old, 26 years old, 30 years old, 34 years old, 38 years old and 42 years old</td>
</tr>
<tr>
<td></td>
<td>Millennium Cohort Study (MCS)</td>
<td>Individuals at birth, 9 months old, 3 years old, 5 years old and 7 years old</td>
</tr>
<tr>
<td></td>
<td>National Pupil Database (NPD)</td>
<td>Annual data on all students attending state primary and secondary schools, but national achievement assessments are only for ages 7, 11, 14 and 16</td>
</tr>
<tr>
<td>USA</td>
<td>Current Population Survey</td>
<td>4 consecutive months, 8 month gap, 4 consecutive months</td>
</tr>
<tr>
<td></td>
<td>Early Childhood Longitudinal Survey, Kindergarten Cohort (ECLS-K)</td>
<td>Fall and Spring Kindergarten, Fall and Spring 1st Grade, Spring 3rd Grade, Spring 5th Grade, Spring 8th Grade</td>
</tr>
<tr>
<td></td>
<td>Early Childhood Longitudinal Survey, Birth Cohort (ECLS-B)</td>
<td>Children at 9 months of age, 2 years old, 4 years old, Fall Kindergarten (either age 5 or 6)</td>
</tr>
<tr>
<td></td>
<td>Education Longitudinal Study of 2002 (ELS:2002)</td>
<td>10th Grade, 12th Grade, 4 years post-baseline, 10 years post-baseline</td>
</tr>
<tr>
<td></td>
<td>General Social Survey (GSS)</td>
<td>Biennial</td>
</tr>
<tr>
<td></td>
<td>Health and Retirement Study (HRS)</td>
<td>2 cohorts surveyed every 2 years for 6 years, with the younger cohort also surveyed 12 years post-baseline</td>
</tr>
<tr>
<td></td>
<td>High School &amp; Beyond (HS&amp;B)</td>
<td>Fall 9th Grade, Spring 11th Grade, Spring 12th Grade, 6 years post-baseline, 12 years post-baseline</td>
</tr>
<tr>
<td></td>
<td>High School Longitudinal Study of 2009 (HSL:09)</td>
<td>8th Grade, 10th Grade, 12th Grade, 6 years post-baseline, 12 years post-baseline</td>
</tr>
<tr>
<td></td>
<td>National Education Longitudinal Study of 1988 (NELS:88)</td>
<td>Waves 1 and 2 are separated by 1 year; Wave 3 is 6 years post-baseline; Wave 4 is 13 years post-baseline</td>
</tr>
<tr>
<td></td>
<td>National Longitudinal Study of Adolescent Health (Add Health)</td>
<td>Annual from 1979 to 1994, biennial thereafter</td>
</tr>
<tr>
<td></td>
<td>National Longitudinal Survey of Youth 1979 (NLSY79)</td>
<td>Annual from 1968 to 1997, biennial thereafter</td>
</tr>
<tr>
<td></td>
<td>Panel Study of Income Dynamics (PSID)</td>
<td>Waves 1 and 2 are separated by 3 years; Waves 2 and 3 are separated by 2 years</td>
</tr>
<tr>
<td></td>
<td>Second Longitudinal Study of Aging (LSOAII)</td>
<td>Variable design, but 8 equally spaced waves are administered over a 32-month window</td>
</tr>
<tr>
<td></td>
<td>Survey of Income and Program Participation</td>
<td></td>
</tr>
</tbody>
</table>

Note: As some of these surveys are ongoing, the reported structure refers to the intended collection design. See McKenzie (2001) for data examples from developing countries.

Irregularly spaced as the unit period is smaller than the observation interval. For example, if the true DGP is based on periods representing a year but the data are only collected biennially, then the data are irregularly spaced. Equally spaced longitudinal surveys are not necessarily regularly spaced.

If the observation interval does not equal the unit period over the entire sample, then a missing data problem is introduced, where the pattern of missing data is dictated by the survey design and is not observation specific. In other words, if data are missing for a particular time period, they are missing for the entire sample and for all variables. This differs from most of the literature on missing data,
where the focus is on observation-specific missing data (e.g. Little and Rubin, 2002). Nonetheless, missing data due to irregularly spaced survey designs invalidate now standard DPD estimators. After discussing the problems created, we present several new estimators and assess their finite-sample performance in a Monte Carlo study. We then illustrate the practical importance of addressing irregular spacing when estimating a dynamic model of early childhood health (as measured by body mass index) using data from the ECLS-K.

As noted above, the existing literature on DPD models with irregularly spaced data is limited. Rosner and Munoz (1988) address the issue in the context of a DPD model with no time-invariant, unobserved heterogeneity. The solution proposed is a pooled nonlinear least squares (NLS) estimator. Jones and Boadi-Boateng (1991) derive an exact maximum likelihood estimator using the Kalman filter to estimate a static, random coefficients panel data model with serially correlated errors and unequal spacing. To address the serial correlation, the authors assume a continuous-time first-order autoregressive, AR(1), error structure. Baltagi and Wu (1999) present a feasible generalized least squares (GLS) estimator for a static panel data model with AR(1) errors and irregularly spaced data. Finally, in the paper most similar to ours, McKenzie (2001) analyzes irregular spacing in the context of dynamic pseudo-panel models. The author shows that consistent estimation is feasible as the number of observations per cohort goes to infinity. However, while covariates other than the lagged dependent variable are allowed in the model, the estimation strategy requires that these covariates be observed in the missing periods as well. This limits the researcher to only time-invariant covariates or time-varying covariates that are obtained from outside data sources.

The issue of irregular spacing has received more attention in the time series literature. In one strand, Savin and White (1978), Jones (1980, 1985, 1986), Dunsmuir and Robinson (1981), Harvey and Piersse (1984), Palm and Nijman (1984), Dufour and Dagenais (1985), Robinson (1985), Kohn and Ansley (1986) and Shively (1993) confront the problem of irregularly spaced data in the context of ARMA(p, q), ARIMA(p, d, q) and ARMAX(p, q, r) models. The solutions generally center on using a state-space representation along with the Kalman filter to derive the exact likelihood function. In another strand, Ryan and Giles (1998), building on Shin and Sarkar (1994a, 1994b), are interested in testing for the presence of unit roots with missing data. The authors assess the performance of various solutions relying on imputation. Finally, the problem of missing data due to utilizing mixed frequency data has received much attention. Eraker et al. (2015), for example, are interested in estimating a multivariate vector autoregression (VAR) with mixed frequency data. The authors address the issue in a Bayesian framework using data augmentation to simulate the missing data for the variable observed at the lower frequency.

In light of this background, our paper is the first, to our knowledge, to explicitly address the issue of irregular spacing in a standard (i.e. non-pseudo) DPD model with unobserved effects and individual, time-varying covariates. We obtain several striking findings. First, traditional DPD estimators are inconsistent in the presence of irregularly spaced data. This arises for three reasons: (i) typical transformations no longer eliminate the time-invariant, observation-specific unobserved effect as the effect has a time-varying factor structure; (ii) the coefficient on the lagged dependent variable depends on the gap structure; and (iii) covariates (and idiosyncratic errors) from the missing time periods are relegated to the error term. Moreover, this inconsistency matters in practice; the finite sample performance of the commonly used DPD estimators can be extremely poor.

Second, if the covariates are strictly exogenous and serially uncorrelated, we propose two consistent estimators. In finite samples, we obtain superior performance by our new estimators: a quasi-differenced generalized method of moments (GMM) estimator and an extended, nonlinear version of an estimator recently proposed in Everaert (2013). Third, in the presence of strictly exogenous but serially correlated covariates, we find that our two new estimators, augmented to include imputation of the covariates from the missing periods, may still perform well. Finally, our application reveals...
a meaningful impact of failing to account for irregular spacing in the data, particularly as it relates to the health–income gradient.

These conclusions, in combination with the fact that even uniformly spaced longitudinal data may be irregularly spaced, should prompt researchers to think more carefully about what constitutes the ‘true’ length of a period in the underlying DGP before estimating dynamic models. The remainder of the paper is organized as follows. Section 2 presents the DPD model, along with the various estimators considered. Section 3 describes the Monte Carlo Study. Section 4 contains the applications to early childhood development. Finally, Section 5 concludes.

2. MODEL

2.1. Setup

The DGP in the standard DPD framework is given by

\[ y_{it} = \gamma y_{i(t-1)} + x_{it} \beta + \alpha_i + \epsilon_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T \tag{1} \]

where \( y_{it} \) is the outcome for individual \( i \) in period \( t \), \( \gamma \) is the autoregressive parameter (\(|\gamma| < 1\)), \( x \) is a \( 1 \times K \) vector of covariates with parameter vector \( \beta \), \( \alpha_i \) is the individual-specific unobserved effect, and \( \epsilon_{it} \) is the idiosyncratic, mean zero error term. The initial condition is given by \( y_{i0} \). We focus on the case where \( \gamma \neq 0 \).

Given a random sample, \( \{y_{it}, x_{it}\}_{i=1, \ldots, N; t=1, \ldots, T} \), along with data on the initial condition, equation (1) may be estimated using a number of techniques depending on the nature of the dependence between \( x \) and the error components, \( \alpha \) and \( \epsilon \). Even if \( x \) is independent of both error components (conditional on \( y_{i(t-1)} \)), pooled ordinary least squares (POLS), fixed effects (FE) and first-differenced (FD) estimation of (1) are biased and inconsistent for fixed \( T \) (Nickell, 1981). However, instrumental variable (IV) estimation of equation (1) is consistent as \( x_{it-1} \) is a valid instrument for \( y_{it-1} \). In fact, further lags of \( x \) may also be used as instruments in a GMM framework.

If \( x \) is dependent on \( \alpha \), then the preceding strategy is no longer valid. The approach based on Anderson and Hsiao (1981) is to eliminate \( \alpha \) via FD and estimate the model using IV. Specifically, given the model

\[ \Delta y_{it} = \gamma \Delta y_{i(t-1)} + \Delta x_{it} \beta + \Delta \epsilon_{it}, \quad i = 1, \ldots, N; \quad t = 2, \ldots, T \tag{2} \]

where \( \Delta \) represents the difference operator and \( x_{it-2} \) is a valid instrument for \( \Delta y_{i(t-1)} \). Again, additional moment conditions may be incorporated into a GMM framework (Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998). Other proposed solutions include long-differences (LD) combined with IV (Hahn et al., 2007), deviations from backward means combined with a Hausman and Taylor (1981) IV approach (Evereart, 2013) or a bias-corrected least squares dummy variable (LSDV) approach (Kiviet, 1995; Hahn and Kuersteiner, 2002; Bun and Carree, 2005).

When the observed data are irregularly spaced, FE or FD transformations no longer succeed in eliminating the unobserved effect. LD will also not be successful unless the data are regularly spaced at the beginning and end periods of the sample. To proceed, begin by noting the following result obtained trivially from repeated substitution in equation (1):

\[ y_{it} = \gamma^s y_{i(t-s)} + \sum_{j=0}^{s-1} x_{it-j} \gamma^j \beta + \sum_{j=0}^{s-1} \gamma^j \alpha_i + \sum_{j=0}^{s-1} \gamma^j \epsilon_{i(t-j)}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T \tag{3} \]

\[ = \gamma^s y_{i(t-s)} + \sum_{j=0}^{s-1} x_{it-j} \gamma^j \beta + \left( \frac{1 - \gamma^s}{1 - \gamma} \right) \alpha_i + \sum_{j=0}^{s-1} \gamma^j \epsilon_{i(t-j)} \tag{4} \]

for all \( s \in [1, t] \). Let \( m = 0, 1, 2, \ldots, M \) index the \( M + 1 \) periods of data observed in the sample, where \( M < T \). Note that the same periods are assumed to be observed for all individuals, \( i \). For example, Figure 1 illustrates the case where the DGP defined by equation (1) applies to periods \( t = 1, \ldots, 8 \) (with period 0 representing the initial period). However, the sample only includes data from periods \( t = 0, 2, 4, 5, 8 \). Thus, in this example, \( M = 4 \) while \( T = 8 \).

Given the DGP in equation (1), the result in equation (3) and irregularly spaced observed data, the model defined over the observed periods is given by

\[
y_{im} = y_{i1}^{g_m} y_{i(m-1)} + x_{im} \beta + \left[ \sum_{j=1}^{g_m-1} x_{i(t(m)-j)} \gamma^j \beta + \frac{(1 - y_{i1}^{g_m})}{1 - \gamma} \alpha_i + \sum_{j=0}^{g_m-1} y_{i1}^{g_m} \epsilon_{i(t(m)-j)} \right],
\]

where \( g_m, m = 1, \ldots, M \), is the gap size or the number of periods between observed periods \( m \) and \( m - 1 \), \( t(m) \) is the actual period reflected by observed period \( m \) and the term in brackets contains all unobserved determinants of \( y_{im} \). For example, in Figure 1, \( t(0) = 0, t(1) = 2, t(2) = 4, t(3) = 5, t(4) = 8 \) and \( g_m = t(m) - t(m-1) \).

Several observations are noteworthy. First, if \( g_m = 1 \) for all \( m \), then the data are regularly spaced and equation (5) simplifies to equation (1). Second, the coefficient on the lagged dependent variable is not constant; it depends on the gap size. Third, the error term in brackets contains the covariates, \( x \), and the idiosyncratic errors, \( \varepsilon \), from the missing periods between periods \( m - 1 \) and \( m \), and the contemporaneous error. Fourth, the unobserved effect, \( \alpha \), is scaled by a period-specific factor loading that depends on the autoregressive parameter and the gap size. Thus the model is closely related to interactive fixed-effects models (Pesaran, 2006; Bai, 2009) and time-varying inefficiency in panel data stochastic frontier models (Cornwell et al., 1990). Finally, if \( g_m = g > 1 \) for all \( m \), where \( g \) is a finite integer, then the data are equally spaced with gaps. Now, FE, FD and LD will eliminate \( \alpha \). However, complications still arise in that the coefficient on \( y_{i(m-1)} \) is \( y^g \) and the \( x \)’s from the missing periods are in the transformed error term.

### 2.2. Estimation

#### 2.2.1. Traditional Estimators

To begin, we evaluate the behavior of perhaps the three most commonly used DPD estimators: Anderson and Hsiao (1981), Arellano and Bond (1991) and Blundell and Bond (1998). Hereafter, we refer to these as AH, AB and BB, respectively. To proceed, rewrite the model in equation (5) as

\[
y_{im} = y_{i1}^{g_m} y_{i(m-1)} + x_{im} \beta + \theta_m \alpha_i + \tilde{\epsilon}_{im}, \quad i = 1, \ldots, N; \quad m = 1, \ldots, M
\]

There is a minor issue with our choice to use the notations \( y_{it} \) and \( y_{im} \) to refer to actual and observed periods, respectively, in that a term such as \( y_{i0} \) has an ambiguous meaning. That said, we prefer our notation as we believe it is easiest for the reader. When we reference data in a specific period, such as \( y_{i0} \), throughout the paper, we will be clear whether the numerical period corresponds to the \( t \) or \( m \) index.

---

**Figure 1.** Illustration of irregularly spaced panel data. [Colour figure can be viewed at wileyonlinelibrary.com]

---

\[t: \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8\]

\[m: \quad 0 \quad 1 \quad 2 \quad 3 \quad 4\]
where
\[ \theta_m \equiv \left( \frac{1 - \gamma^{g_m}}{1 - \gamma} \right); \quad \tilde{\varepsilon}_{im} \equiv \sum_{j=1}^{g_m-1} x_{i,t(m)-j} \gamma^j \beta + \sum_{j=0}^{g_m-1} \gamma^j \varepsilon_{i,t(m)-j} \]

First-differencing the model yields
\[ \Delta y_{im} = \gamma^{g_m} y_{im-1} - \gamma^{g_m-1} y_{im-2} + \Delta x_{im} \beta + \alpha_i \Delta \theta_m + \Delta \tilde{\varepsilon}_{im}, \quad i = 1, \ldots, N; \quad m = 2, \ldots, M \] (7)
where \( \Delta \) represents the difference between consecutive, observed periods (indexed by \( m \)) and \( \Delta \theta_m = (\gamma^{g_m-1} - \gamma^{g_m})/(1 - \gamma) \). However, ignoring the irregular spacing issue entails naively focusing on
\[ \Delta y_{im} = \gamma_0 \Delta y_{im-1} + \Delta x_{im} \beta + [\alpha_i \Delta \theta_m + \gamma^{g_m} y_{im-1} - \gamma^{g_m-1} y_{im-2} - \gamma_0 \Delta y_{im-1} + \Delta \tilde{\varepsilon}_{im}] \] (8)
Not only does \( \gamma_0 \) lack any structural interpretation, but none of the estimators will produce a consistent estimate of \( \gamma_0 \). This arises because the instruments used by all three estimators to address the endogeneity of \( \Delta y_{im-1} \) are not valid since the error term includes \( y_{im-1} \) and \( y_{im-2} \). Moreover, estimates of \( \beta \) will be inconsistent owing to the dependence between \( \Delta x_{im} \) and \( \Delta y_{im-1} \) and the fact that FD does not eliminate the unobserved effect from equation (7). Serial correlation in \( x \) is also likely to contribute to inconsistency due to dependence between \( \Delta x_{im} \) and \( \Delta \tilde{\varepsilon}_{im} \).\(^2\) Finally, note that the additional moment conditions utilized in the BB estimator are also not valid. For instance, utilizing \( \Delta y_{im-1} \) as an instrument for \( y_{im-1} \) in equation (6) is not valid since \( \Delta y_{im-1} \) is dependent on \( \alpha_i \).

2.2.2. Proposed Estimators
We propose two estimators of the model in equation (6). The first is a GMM estimator based on quasi-differencing (QD). The second is an extension of an estimator recently proposed in Everaert (2013).

Quasi-differencing.
While FD does not eliminate the unobserved effect, a QD approach does. To see this, we set up the following quasi-differenced equation:
\[ y_{im} - \varphi_m y_{im-1} = \gamma^{g_m} y_{im-1} - \varphi_m \gamma^{g_m-1} y_{im-2} + (x_{im} - \varphi_m x_{im-1}) \beta \\
+ \alpha_i (\theta_m - \varphi_m \theta_{m-1}) + \tilde{\varepsilon}_{im} - \varphi_m \tilde{\varepsilon}_{im-1}, \quad i = 1, \ldots, N; \quad m = 2, \ldots, M \] (9)
Defining
\[ \varphi_m \equiv \frac{\theta_m}{\theta_{m-1}} = \frac{1 - \gamma^{g_m}}{1 - \gamma^{g_m-1}} \] (10)
implies that equation (9) simplifies to
\[ y_{im} - \varphi_m y_{im-1} = \gamma^{g_m} y_{im-1} - \varphi_m \gamma^{g_m-1} y_{im-2} + (x_{im} - \varphi_m x_{im-1}) \beta + \tilde{\varepsilon}_{im} \] (11)
where
\[ \tilde{\varepsilon}_{im} \equiv \tilde{\varepsilon}_{im} - \varphi_m \tilde{\varepsilon}_{im-1} \]
\[ = \left[ \sum_{j=1}^{g_m-1} x_{i,t(m)-j} \gamma^j - \varphi_m \sum_{j=1}^{g_m-1} x_{i,t(m-1)-j} \gamma^j \right] \beta + \sum_{j=0}^{g_m-1} \gamma^j \varepsilon_{i,t(m)-j} - \varphi_m \sum_{j=0}^{g_m-1} \gamma^j \varepsilon_{i,t(m-1)-j} \] (12)
\[^2\] Serial correlation in \( x \) implies that the observed \( x_{im} \) is not independent of \( x_{i,t(m)-1} \), \ldots, \( x_{i,t(m)-(g_m-1)} \).
With \( \varrho_m \) unknown, the model may be estimated by GMM as in Nauges and Thomas (2003). The moment conditions used are given by \( E[Z_i \widetilde{e}_i] = 0 \), where

\[
Z_i = \begin{bmatrix}
y_{i0} & 0 & 0 & \cdots & 0 & x_{i1} & \cdots & x_{iM} & 0 & \cdots & 0 & \cdots & 0 & 0 \\
0 & y_{i1} & \cdots & \cdots & 0 & 0 & \cdots & 0 & x_{i1} & \cdots & x_{iM} & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & y_{iM-2} & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & x_{iM} & 0
\end{bmatrix}
\]

and \( \widetilde{e}_i = \left[ \widetilde{e}_{i2} \cdots \widetilde{e}_{iM} \right]' \). Our estimator, referred to as QD-GMM, is consistent as \( N \to \infty \) if \( x \) is strictly exogenous and serially uncorrelated. Inference is based on cluster-robust standard errors given the within-group serial correlation in \( \widetilde{e}_{im} \) (Cameron and Miller, 2015).

**Orthogonal to backward mean transformation.**

Everaert (2013) proposed an alternative technique for estimating DPD models. The approach estimates the model in levels via IV. With regularly spaced data, the model is given by equation (1). The proposed instrument for \( y_{it-1} \) is the OLS residual of \( y_{it-1} \) regressed on its backward mean, \( \hat{y}^{b}_{it-1} \), defined as

\[
\hat{y}^{b}_{it-1} = \frac{1}{T} \sum_{s=0}^{T-1} y_{is}
\]

(13)

If \( x_{it} \) is independent of \( \alpha_i \) and \( \varepsilon_{it} \), then this estimator is consistent as \( T \to \infty \). However, the inconsistency for fixed \( T \) is shown to be small in practice.

If \( x \) is correlated with the unobserved effect, then Everaert (2013) suggests using Hausman and Taylor (1981) type instruments for \( x \); namely, deviations from individual sample means, \( x_{it} - \bar{x}_i \), where \( \bar{x}_i = (1/T) \sum_t x_{it} \). A Mundlak (1978) correlated random effects (CRE) approach in combination with use of the residual from the backward mean regression as an instrument for \( y_{it-1} \) is also consistent as \( T \to \infty \) and avoids the need to instrument for each covariate suspected of being dependent on \( \alpha \). Formally, the Mundlak (1978) approach assumes \( E[\alpha_i | x_i] = \bar{x}_i \delta \). We can then write \( \alpha_i = \bar{x}_i \delta + \upsilon_t \) and equation (1) becomes

\[
y_{it} = y y_{it-1} + x_{it} \beta + \bar{x}_i \delta + \upsilon_t + \varepsilon_{it}
\]

(14)

This is estimable using the Everaert (2013) proposed instrument for \( y_{it-1} \).

With irregularly spaced data, the model is given in equation (6). Our proposed estimator extends Everaert’s (2013) approach to account for the time-varying coefficient on the lagged dependent variable. Specifically, we focus on the CRE approach and estimate

\[
y_{im} = y y_{im-1} + x_{im} \beta + \bar{x}_i \left( \frac{1 - y g_{im}}{1 - \gamma} \right) \delta + \left[ \left( \frac{1 - y g_{im}}{1 - \gamma} \right) \upsilon_t + \tilde{\varepsilon}_{im} \right]
\]

(15)

by NLS-IV using orthogonal deviations from its backward mean as an instrument for \( y_{im-1} \). This estimator, referred to as E-CRE, requires \( T \to \infty \) and \( x \) to be strictly exogenous and serially uncorrelated for consistency. Inference is based on cluster-robust standard errors given the within-group serial correlation due to \( \upsilon_t \) (Cameron and Miller, 2015).

### 2.3. Imputation of Missing Covariates

Perhaps the most troublesome issue created by irregularly spaced panel data is the relegation of the covariates from the missing periods to the error term. Missing data on the covariates implies that serial
correlation in \( x \) causes all the estimators considered thus far to be inconsistent due to the endogeneity of \( x \) and the inability to obtain suitable instruments for the lagged dependent variable.

To address this issue, we incorporate imputation of the missing covariates into our estimators. This is analogous to the strategy pursued in Ryan and Giles (1998), who analyze the problem of unit root testing with irregularly spaced time series. In that context, the authors consider imputing data for the missing periods using two strategies: linear interpolation and carrying the last value forward. Our extended estimators are also similar to the Bayesian approach to mixed-frequency VAR models based on data augmentation suggested in Eraker et al. (2015).

For ease of exposition, we first consider imputation in the context of the E-CRE estimator and then discuss the QD-GMM estimator.

**Orthogonal to backward mean transformation.**

The E-CRE estimator is based on equation (15), where the composite error term includes

\[
\sum_{j=1}^{g_m-1} x_{i,t(m)-j} y^j \beta 
\]

While one could impute distinct values for \( x \), \( QD \)-GMM.

information). \( \text{process. Denoted E-CRE-AR1, we relegate the details to the supplementary Appendix (supporting} \)

\[
\text{Further, we also consider a fourth method assuming that each element of } x \text{ follows its own AR(1) process. Denoted E-CRE-AR1, we relegate the details to the supplementary Appendix (supporting information).}

**QD-GMM.**

The QD-GMM estimator is based on equation (11), where the composite error term includes

\[
\sum_{j=1}^{g_{m-1}} x_{i,t(m)-j} y^j \beta
\]

3 As an alternative, one might consider imputation of the missing lagged dependent variable. The advantage of this approach is that there is only a single variable to impute. However, this introduces some complications in that the measurement error—representing the deviation between the actual \( y_{it} \) and the imputed \( \hat{y}_{it-1} \)—will (likely) depend on the deviation between the covariates from period \( t-1 \) and the covariates used to impute the missing value, causing \( x_{it} \) to be correlated with the composite error if \( x \) is serially correlated.
Using a single imputed value of \( x \) for all periods strictly between \( m \) and \( m - 1 \), denoted again by \( x_{im}^{0} \), the estimating equation becomes

\[
y_{im} - \phi_{m}y_{im-1} = \gamma^{g_{m}}y_{im-1} - \phi_{m}y_{im-2} + (x_{im} - \phi_{m}x_{im-1})\beta + D_{im}x_{im}^{0}\left(\frac{\gamma^{g_{m}}}{1 - \gamma}\right)\beta
\]

\[\quad - \phi_{m}D_{im-1}x_{im-1}^{0}\left(\frac{\gamma^{g_{m-1}}}{1 - \gamma}\right)\beta + \sum_{j=0}^{g_{m-1}} \gamma^{j}e_{i,t(m-j)} - \phi_{m} \sum_{j=0}^{g_{m-1}} \gamma^{j}e_{i,t(m-1)-j} + \sigma_{im}
\]

(20)

where \( \sigma_{im} \) represents the approximation error from equation (19). Using the three possible values for \( x_{im}^{0} \) discussed previously, we refer to these estimators as QD-GMM-L, QD-GMM-C and QD-GMM-A, respectively. We also consider a fourth method assuming that each element of \( x \) follows its own AR(1) process. Denoted QD-GMM-AR1, we relegate the details to the supplementary Appendix.

3. MONTE CARLO STUDY

3.1. Design of the DGP

To compare the finite-sample performance of the various estimators discussed above, we utilize the basic Monte Carlo design in Everaert (2013). The DGP, with a single covariate, is as follows:

\[
y_{it} = \gamma y_{it-1} + \beta x_{it} + \alpha_{i} + \epsilon_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T
\]

\[
y_{i0} = \frac{\alpha_{i} + \beta \theta \alpha_{i}(1 - \rho)^{-1}}{1 - \gamma} + \xi_{i0}
\]

\[
\xi_{i0} \overset{\text{i.i.d.}}{\sim} N\left(0, \sigma_{\xi_{0}}^{2}\right)
\]

\[
\sigma_{\xi_{0}}^{2} = \frac{1}{1 - \gamma^{2}} + \frac{\beta^{2}\sigma_{\epsilon}^{2}(1 + \gamma \rho)}{(1 - \gamma \rho)(1 - \gamma^{2})(1 - \rho^{2})}
\]

\[
\alpha_{i} \overset{\text{i.i.d.}}{\sim} N\left(0, \sigma_{\alpha}^{2}\right)
\]

\[
\sigma_{\alpha}^{2} = \mu_{\alpha}(1 - \gamma)^{2}
\]

\[
x_{it} = \theta \alpha_{i} + \rho x_{i,t-1} + \xi_{it}
\]

\[
x_{i0} = \frac{\theta \alpha_{i}}{1 - \rho} + \xi_{i0}\left(\frac{1}{1 - \rho^{2}}\right)^{1/2}
\]

\[
\xi_{it} \overset{\text{i.i.d.}}{\sim} N\left(0, \sigma_{\xi}^{2}\right)
\]

\[
\sigma_{\xi}^{2} = \left(\sigma_{\xi_{0}}^{2} - \frac{\gamma^{2}}{1 - \gamma^{2}}\right)\left[\frac{(1 - \gamma \rho)(1 - \gamma^{2})(1 - \rho^{2})}{\beta^{2}(1 + \gamma \rho)}\right]
\]

We vary the values of \( \sigma_{\xi_{0}}^{2}, \mu_{\alpha}, \theta \) and \( \rho \). By varying \( \sigma_{\xi_{0}}^{2} \) we manipulate the variance of the signal explaining \( y_{it} \) contained in the within variation of \( x_{it} \) and \( y_{it-1} \) relative to the noise contained in \( \alpha_{i} \) and \( \epsilon_{it} \). By adjusting \( \mu_{\alpha} \) we alter the relative importance of \( \alpha \) versus \( \epsilon \) in the determination of \( y \). By altering \( \theta \) and \( \rho \) we control the degree of serial correlation in the covariate, as well as correlation between \( x_{it} \) and \( \alpha_{i} \). For parameter values, we set \( \beta = 0.1 \) and \( \gamma = \{0.4, 0.8\} \), allowing for (relatively) low and high persistence. The long-run coefficient, \( \beta/(1 - \gamma) \), is either 0.17 or 0.5.

We induce a pattern of irregular spacing to mimic the ECLS-K, where a ‘true’ period is approximately 6 months. We also do a limited exploration of the role of the initial condition. First,
when simulating the data, we set \( t = -99, -98, \ldots, -1, 0, 1, \ldots, 17 \), and then retain periods \{0, 1, 2, 3, 7, 11, 17\}. Second, we set \( t = 0, 1, \ldots, 17 \), and then retain periods \{0, 1, 2, 3, 7, 11, 17\}. Thus, in both cases, \( M = 6 \) in our notation, with \( m = 0 \) representing the initial period in the data sample. For all experiments, we set \( N = 500 \) and perform 500 simulations.

### 3.2. Simulation Results

The results are relegated to the supplementary Appendix. Appendices A and B report the median bias, standard deviation (SD), mean (cluster-robust) standard error (SE) and coverage rate for \( \gamma, \beta, \) and the long-run coefficient, \( \beta/(1 - \gamma) \), where Appendix A (B) sets the initial period as \( t = -99 \) (\( t = 0 \)). Appendices C and D display Pitman’s (1937) nearness measure, PN, for the select comparisons. Formally, this measure is given by

\[
PN = \Pr \left[ \left| \hat{\Theta}_1 - \Theta \right| < \left| \hat{\Theta}_2 - \Theta \right| \right]
\]

where \( \hat{\Theta}_j, j = 1, 2, \) represents estimates of the parameter \( \Theta, \Theta \in \{\gamma, \beta, \beta/(1 - \gamma)\} \), produced estimator \( j \). Thus \( PN > 0.5 \) indicates superior performance of the first estimator. The advantage of PN is that it summarizes the entire sampling distribution of an estimator. It is estimated by its empirical counterpart.

In the following discussion, we highlight our primary takeaways from the experimental designs most likely to conform to our application. Specifically, we focus on Tables A11, A12, A15, and A16 in supplementary Appendix A. These experiments utilize a relatively high value for \( \theta \) and a relatively high value for \( \mu_a \). In Tables A11 and A12, \( \gamma \) is equal to 0.4. In Tables A15 and A16, \( \gamma \) is equal to 0.8. In Table A11 \((\sigma_x^2, \mu_a) = (2, 5)\), in Table A12 \((\sigma_x^2, \mu_a) = (0.5, 5)\), in Table A15 \((\sigma_x^2, \mu_a) = (3, 5)\) and in Table A16 \((\sigma_x^2, \mu_a) = (2, 5)\). Within each table, DGP1 sets \( \rho \) such that there is no serial correlation in \( x \) (in the population), while DGP2 (DGP3) adjust \( \rho \) such that there is modest (strong) serial correlation in \( x \); \( \text{corr}(x_{it}, x_{i(t-1)}) = 0.00 \) in DGP1 and 0.40 (0.80) in DGP2 (DGP3).

#### 3.2.1. Estimating \( \gamma \) and \( \beta \)

We reach four primary conclusions from the experiments considered. First, the traditional estimators (AH, AB, BB) for \( \gamma \) are severely biased, particularly when \( \gamma \) is of moderate size. Among these estimators, the BB estimator tends to be the most biased (even when \( \gamma \) is relatively close to unity). The traditional estimators for \( \beta \) are also severely (and similarly) biased in the majority of experiments with \( x \) serially correlated (DGP2 and DGP3). Second, our proposed estimators without imputation, QD-GMM and E-CRE, generally perform well in terms of bias and SD with \( x \) serially uncorrelated (DGP1). That said, E-CRE tends to outperform QD-GMM in terms of estimating \( \gamma \) and \( \beta \) when \( \gamma \) is relatively high \((\gamma = 0.8)\). QD-GMM tends to outperform E-CRE in terms of estimating \( \gamma \) and the long-run effect, \( \beta/(1 - \gamma) \) when \( \gamma \) is of moderate size \((\gamma = 0.4)\).

Third, with \( x \) serially correlated (DGP2 and DGP3), the QD-GMM and E-CRE estimators for \( \gamma \) and \( \beta \) can be severely biased. The median bias for \( \gamma \) and \( \beta \) often increases at least two- or threefold relative to DGP1. Fourth, our proposed estimators with imputation using averages, QD-GMM-A and E-CRE-A, or the current value, QD-GMM-C and E-CRE-C, perform consistently well. These estimators generally outperform our estimators without imputation and the traditional estimators. As before, the main exceptions are in Tables A11 and A12, where the E-CRE-A and E-CRE-C estimates of \( \gamma \) are severely biased when \( \gamma \) is of moderate size.

Finally, comparison among the estimators that seem to perform best according to the PN measure provides an ambiguous picture; see panels III and IV in Tables C3 and C4 in supplementary Appendix C. With \( x \) serially uncorrelated (DGP1), QD-GMM outperforms QD-GMM-A and QD-GMM-C and
E-CRE outperforms E-CRE-A and E-CRE-C. Between QD-GMM and E-CRE, the former (latter) tends to perform better when $\gamma$ is 0.4 (0.8). With $x$ highly serially correlated (DGP3), QD-GMM-A and QD-GMM-C outperform QD-GMM, whereas E-CRE-A and E-CRE-C outperform E-CRE only when $\gamma$ is 0.8. When comparing the different imputation estimators, we find that QD-GMM-A outperforms QD-GMM-C in terms of estimating $\beta$ and $\beta/(1-\gamma)$, but not $\gamma$. E-CRE-A outperforms E-CRE-C in terms of estimating all of the parameters except $\gamma$ when $\gamma$ is 0.8. Across estimators, QD-GMM-A is generally preferable to E-CRE-A, particularly when $\gamma$ is 0.4. QD-GMM-C is generally preferable to E-CRE-C, although the latter is preferred in terms of estimating $\beta$ when $\gamma$ is 0.4.

3.2.2. Inference on $\gamma$ and $\beta$

Inference is based on a comparison of the mean (cluster-robust) standard error to the empirical SD of the estimates, as well as on the coverage rates of 95% confidence intervals utilizing these same standard errors. Again, we highlight our primary findings. First, the mean SEs are very close to the empirical SDs for QD-GMM and E-CRE with $x$ serially uncorrelated (DGP1) and QD-GMM-A and E-CRE-A with $x$ serially correlated (DGP2 and DGP3). The mean SEs for QD-GMM, QD-GMM-A and QD-GMM-C tend to be slightly smaller than the corresponding empirical SDs, whereas the mean SEs for E-CRE, E-CRE-A and E-CRE-C tend to be slightly larger.

Second, the coverage rates for $\gamma$ for QD-GMM and E-CRE with $x$ serially uncorrelated (DGP1) are severely distorted when $\gamma$ is of moderate size. The coverage rates for $\beta$ in these cases are closer to 0.95, particularly for QD-GMM. In the experiments with a relatively large $\gamma$, the coverage rates for $\gamma$ improve. Specifically, the coverage rates for QD-GMM exceed 0.7 and the coverage rates for E-CRE approach unity. The coverage rates for $\beta$ are close to 0.95 for both estimators.

Third, the coverage rates for QD-GMM-A, QD-GMM-C, E-CRE-A and E-CRE-C with $x$ serially correlated (DGP2 and DGP3) can also be severely distorted in the experiments with $x$ of moderate size. In particular, the coverage rates for E-CRE-A and E-CRE-C are rarely close to 0.95 for the three parameters. The coverage rates for $\gamma$ for QD-GMM-C are closer to 0.95, whereas the coverage rates for $\beta$ and $\beta/(1-\gamma)$ are closer to 0.95 for QD-GMM-A. With a relatively large $\gamma$, the coverage rates for QD-GMM-A, QD-GMM-C are relatively unchanged. However, the coverage rates for E-CRE-A are much improved. Overall, the coverage rates of our proposed estimators, when distorted, are a reflection of the bias of the estimators and not erroneous standard errors.

3.2.3. Summation

The naïve application of traditional DPD estimators is likely to be grossly misleading in the presence of irregularly spaced data, particularly with serially correlated covariates. Of our proposed estimators, QD-GMM-A and E-CRE-A, represent a significant improvement in the experiments most likely to conform to our application. However, there is not an unambiguous ranking between these estimators. Specifically, QD-GMM-A estimator seems preferable when $\gamma$ is of moderate size. The two estimators are more comparable when $\gamma$ is relatively large.

Because the results discussed to this point represent only a selection of the experiments conducted, Table II offers a straightforward way of aggregating performance across all experiments reported in supplementary Appendices A and B. Within each experiment, parameter and DGP, we rank the 13 estimators considered according to either bias or SD; one indicates the smallest bias or SD and 13 indicates the largest. Table II reports the mean rank, median rank and SD of ranks for each estimator across the 288 experiment–parameter combinations for a given DGP. With $x$ serially uncorrelated (DGP1), E-CRE has the lowest mean and median rank (2.6 and 2, respectively) in terms of bias. QD-GMM is not far behind (3.6 and 3, respectively). With $x$ highly serially correlated (DGP3), QD-GMM-A has the lowest mean and median rank (2.8 and 2.5, respectively) in terms of bias. E-CRE-A is not far behind (4.8 and 4.5, respectively). In terms of SD, the E-CRE estimators generally outrank the QD-GMM estimators regardless of the DGP. In light of this, we focus on QD-GMM-A, QD-GMM-C,
Table II. Simulation results: estimator rankings by DGP across all experiments

<table>
<thead>
<tr>
<th>Estimator</th>
<th>DGP1</th>
<th>DGP2</th>
<th>DGP3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean rank SD</td>
<td>Median rank</td>
<td>Mean rank SD</td>
</tr>
<tr>
<td><strong>Panel 1. Median bias</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AH</td>
<td>6.5 2.3 7</td>
<td>7.3 3.4 7</td>
<td>9.1 3.7 10</td>
</tr>
<tr>
<td>AB</td>
<td>6.8 2.4 6</td>
<td>7.7 3.4 8</td>
<td>9.2 3.4 11</td>
</tr>
<tr>
<td>BB</td>
<td>8.2 3.8 8.5</td>
<td>10.1 2.5 10</td>
<td>9.6 3.5 11</td>
</tr>
<tr>
<td>QD-GMM</td>
<td>3.6 1.8 3</td>
<td>3.6 1.9 4</td>
<td>7.0 3.0 7</td>
</tr>
<tr>
<td>QD-GMM-L</td>
<td>9.9 1.5 10</td>
<td>8.2 2.8 9</td>
<td>5.5 2.4 6</td>
</tr>
<tr>
<td>QD-GMM-A</td>
<td>8.1 2.1 8</td>
<td>5.8 2.4 6</td>
<td>2.8 1.5 2.5</td>
</tr>
<tr>
<td>QD-GMM-C</td>
<td>9.6 3.1 10</td>
<td>8.4 3.9 10</td>
<td>5.4 3.6 6</td>
</tr>
<tr>
<td>QD-GMM-AR1</td>
<td>3.4 1.7 3</td>
<td>4.2 2.7 3</td>
<td>5.4 3.0 5</td>
</tr>
<tr>
<td>E-CRE</td>
<td>2.6 1.6 2</td>
<td>5.0 2.8 4</td>
<td>7.5 4.1 8</td>
</tr>
<tr>
<td>E-CRE-L</td>
<td>10.4 3.3 12</td>
<td>10.3 3.5 12</td>
<td>9.7 2.2 10</td>
</tr>
<tr>
<td>E-CRE-A</td>
<td>8.5 2.2 9</td>
<td>6.9 2.9 8</td>
<td>4.8 2.2 4.5</td>
</tr>
<tr>
<td>E-CRE-C</td>
<td>10.1 3.9 12</td>
<td>9.3 4.4 12</td>
<td>7.4 4.2 8</td>
</tr>
<tr>
<td>E-CRE-AR1</td>
<td>3.2 1.4 3</td>
<td>4.2 2.1 4</td>
<td>7.5 2.9 8</td>
</tr>
</tbody>
</table>

| **Panel 2. Standard deviation** |                |               |               |               |               |               |
| AH           | 10.2 2.2 9    | 10.6 2.2 11   | 11.5 2.1 13   |               |               |               |
| AB           | 8.7 2.6 8     | 8.7 2.8 9     | 9.8 2.6 12    |               |               |               |
| BB           | 6.5 2.4 7     | 6.8 2.7 7     | 7.8 2.8 8     |               |               |               |
| QD-GMM       | 9.9 2.7 11    | 10.0 2.3 11   | 9.5 2.3 9     |               |               |               |
| QD-GMM-L     | 7.9 4.2 9     | 8.0 4.1 9     | 7.6 3.4 8     |               |               |               |
| QD-GMM-A     | 7.4 3.1 7     | 6.9 2.7 7     | 6.9 3.0 7     |               |               |               |
| QD-GMM-C     | 6.4 2.0 6     | 6.6 2.1 6     | 6.2 2.6 6     |               |               |               |
| QD-GMM-AR1   | 8.3 1.6 9     | 7.3 1.8 7     | 5.8 2.4 6     |               |               |               |
| E-CRE        | 6.9 4.8 6     | 6.7 5.0 5.5   | 6.7 3.7 5     |               |               |               |
| E-CRE-L      | 4.5 3.6 3     | 4.5 3.6 3     | 4.1 3.3 2     |               |               |               |
| E-CRE-A      | 4.0 2.9 3     | 3.9 2.6 3     | 3.7 2.6 3     |               |               |               |
| E-CRE-C      | 3.5 3.1 2.5   | 4.0 3.8 2     | 4.3 4.5 1     |               |               |               |
| E-CRE-AR1    | 7.0 4.3 6.5   | 7.1 4.0 6.5   | 7.0 3.2 6     |               |               |               |

Note: Within each experiment, parameter and DGP reported in supplementary Appendices A and B, the rankings of estimators are computed in terms of median bias or standard deviation. A rank of one indicates the lowest median bias or SD. The mean, median and SD of the ranks—computed over $96 \times 3 = 288$ experiments—for each DGP are reported here. See text for further details.

4. APPLICATION

4.1. Motivation

We utilize the ECLS-K to estimate a dynamic model of child health as measured by BMI. The traditional estimators assume the model is given by equation (1), except with the time periods indexed by $m$ instead of $t$. Our proposed estimates assume the model is given by equation (6).


Understanding the evolution of BMI during early childhood is critical. Recently, Millimet and Chernis (2015) document an increase in the persistence of BMI once children enter primary school. As a result, a better understanding of the factors that may affect the BMI of children during primary school...
is crucial to combating not only obesity but also the problem with malnutrition and underweight children. As such, we estimate a dynamic weight production function, viewing BMI as determined by not only contemporaneous health-related inputs but also the complete history of health-related inputs (Lakdawalla et al., 2005; Ng et al., 2012; Cavaco et al., 2014; Fortin and Yazbeck, 2015). Given the lack of data on the complete history of health-related inputs, we further assume that lagged BMI is a sufficient statistic for all prior inputs (Strauss and Thomas, 2008).

Specifically, we focus on a parsimonious set of covariates including socioeconomic status (SES), household size and changes in fast food prices over time (Cawley, 2010). SES has a theoretically ambiguous impact on child BMI (e.g. Fertig et al., 2009). Certainly, SES is associated with a reduction in malnutrition and the incidence of being underweight. However, there is also concern that the propensity for childhood obesity increases in dual-earner households due to, perhaps, less parental oversight and more reliance on quicker, less healthy eating options. Like SES, the impact of household size on childhood obesity is not definitive. For example, multiple children may dilute parents’ time and resources, leading them to implement more opportune means of caring for their children (more television, fast food, etc.). Alternatively, an increase in family size may decrease the amount of requisite nutrients allocated to any individual child, thus reducing BMI (Chen and Escarce, 2010). With respect to food prices, studies have demonstrated an association between real food prices and BMI, especially as it relates to declining fast food prices (Powell, 2009; Cawley, 2010). Lastly, we include time fixed effects to capture broader economic phenomenon, such as technological innovation in food preparation (Cutler et al., 2003).

4.2. Data

The data are from the ECLS-K and have been used previously in Millimet and Tchernis (2015) to analyze persistence in BMI during early childhood. However, their analysis utilized nonparametric measures of mobility over different time intervals rather than a dynamic panel data model. Thus irregular spacing was not an issue but at the expense of different information being learned. Collected by the US Department of Education, the ECLS-K surveys a nationally representative cohort of children throughout the USA in fall and spring kindergarten, fall and spring first grade, spring third grade, spring fifth grade and spring eighth grade. The sample includes data on over 20,000 students from roughly 1000 schools who entered kindergarten during the 1998–99 school year. Information is collected on a host of topics, including family background, teacher and school characteristics, and student height and weight. The fall first-grade wave was only administered to a portion of the sample and thus we ignore this wave.

We utilize a balanced sample of children through eighth grade for whom we have non-missing data on age and gender, and valid measures of height and weight. With information on height and weight, we create BMI z-scores. The z-score is obtained using CDC 2000 growth charts; these are age and gender specific, and are adjusted for normal growth. Covariates included in the model are: an SES index and its square, household size and fast food prices. The SES index is based on father’s and mother’s education, father’s and mother’s occupation, and household income. To control for fast food prices, we construct a price index using data from the US Department of Agriculture (USDA). Specifically, the price index is constructed using regional pricing data from the USDA’s Quarterly Food-Away-from-Home Prices database on hamburgers, fried chicken, pizza and combination meals from limited service restaurants. Lastly, missing values for the covariates are imputed and imputation dummies are added to the control set.

Summary statistics are provided in Table III. The mean BMI z-score is 0.50, which corresponds to the 70th percentile of the relevant distribution based on the growth charts references above. Moreover, nearly 32% of the sample are overweight and more than 16% are obese (defined as having a BMI above the 85th and 95th percentile of the relevant distribution, respectively). Also, the mean household
Table III. Summary statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI z-score</td>
<td>54,930</td>
<td>0.502</td>
<td>1.135</td>
<td>-11.639</td>
<td>4.232</td>
</tr>
<tr>
<td>SES</td>
<td>52,338</td>
<td>0.061</td>
<td>0.810</td>
<td>-4.75</td>
<td>2.88</td>
</tr>
<tr>
<td>Household size</td>
<td>51,272</td>
<td>4.558</td>
<td>1.353</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>Fast food price index</td>
<td>54,930</td>
<td>0.985</td>
<td>0.063</td>
<td>0.827</td>
<td>1.070</td>
</tr>
<tr>
<td>No mother in household (1 = Yes)</td>
<td>54,930</td>
<td>0.022</td>
<td>0.148</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Exercise at least 3 days/wk (1 = Yes)</td>
<td>50,879</td>
<td>0.696</td>
<td>0.460</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Usually eats at school: provided lunch (1 = Yes)</td>
<td>51,247</td>
<td>0.683</td>
<td>0.465</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Health insurance (1 = Yes)</td>
<td>51,479</td>
<td>0.905</td>
<td>0.293</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Number of books at home</td>
<td>50,791</td>
<td>10.47</td>
<td>152.353</td>
<td>0</td>
<td>5000</td>
</tr>
</tbody>
</table>

*Note:* Data from the Early Childhood Longitudinal Survey, Kindergarten Cohort (ECLS-K). N, number of observations; SD, standard deviation; SES, socioeconomic status. See text for further details.

size is between four and five individuals. Fast food prices are indexed to 1.00 in the fall kindergarten wave, with a mean of 0.98.

4.3. Results

The results are displayed in Table IV. We report only the AH, AB, BB, QD-GMM, E-CRE, QD-GMM-A(-C) and E-CRE-A(-C) estimators. Because the covariates are serially correlated in the data, we are most interested in the QD-GMM-A(-C) and E-CRE-A(-C) estimators; the remaining estimators are presented for comparison.

Several interesting findings emerge. First, the traditional estimators, along with QD-GMM and QD-GMM-A(-C), yield an estimate of $\gamma$ around 0.26. In contrast, the E-CRE and E-CRE-A(-C) estimates of $\gamma$ are roughly 0.89. The discrepancy between QD-GMM-A(-C) and E-CRE-A(-C) is unexpected. In the simulations discussed above, the largest discrepancy in median bias occurs in Tables A11 and A12 (where the unobserved effect is relatively more important than the idiosyncratic error). In this case, with $x$ even modestly serially correlated, the median bias of E-CRE-A(-C) is about three to six times larger than the median bias of QD-GMM-A(-C). In light of this, we are apt to give more credence to the QD-GMM-A(-C) estimators of $\gamma$, but we revisit this below.

Second, the estimated health–income gradient differs across the various estimators. Specifically, AB, AH and QD-GMM-A indicates no statistically meaningful association between SES and BMI $z$-score, whereas BB produces a modest negative association that is statistically significant at conventional levels. QD-GMM-C, when evaluated at the mean level of SES, also yields a modest negative association between SES and BMI $z$-score. More precisely, the QD-GMM-C estimator indicates a $U$-shaped relationship, with the trough close to the 75th percentile of the SES distribution. This non-monotonic relationship is consistent with improved health (as measured by BMI $z$-score) as households move out of the lower tail of the SES distribution, but a potential worsening of child health as SES continues to improve. E-CRE and E-CRE-A(-C), however, indicate a small but positive and statistically significant association between SES and BMI $z$-score over the entire range of the data. Thus, while we are cautious not to overstate the findings, our new estimators suggest a different health–income gradient from the traditional estimators. Third, the impact of fast food prices, even when statistically significant, is economically close to zero. Specifically, only AH and E-CRE indicate a statistically significant relationship between fast food prices and BMI $z$-score. The effects,
Table IV. Determinants of child BMI z-scores

<table>
<thead>
<tr>
<th></th>
<th>AH</th>
<th>AB</th>
<th>BB</th>
<th>QD-GMM</th>
<th>E-CRE</th>
<th>QD-GMM-A</th>
<th>E-CRE-A</th>
<th>QD-GMM-C</th>
<th>E-CRE-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>0.259*</td>
<td>0.279*</td>
<td>0.343*</td>
<td>0.267*</td>
<td>0.890*</td>
<td>0.264*</td>
<td>0.886*</td>
<td>0.264*</td>
<td>0.886*</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.008)</td>
<td>(0.013)</td>
<td>(0.008)</td>
<td>(0.013)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>SES</td>
<td>0.011</td>
<td>0.01</td>
<td>-0.030†</td>
<td>0.046†</td>
<td>0.026†</td>
<td>-0.029</td>
<td>0.011‡</td>
<td>-0.023</td>
<td>0.011‡</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.019)</td>
<td>(0.010)</td>
<td>(0.020)</td>
<td>(0.006)</td>
<td>(0.019)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>SES²</td>
<td>0.006</td>
<td>0.005</td>
<td>-0.003</td>
<td>0.026*</td>
<td>0.007</td>
<td>0.013</td>
<td>0.013*</td>
<td>0.014‡</td>
<td>0.007*</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.006)</td>
<td>(0.008)</td>
<td>(0.004)</td>
<td>(0.008)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Household size</td>
<td>-0.003</td>
<td>-0.004</td>
<td>-0.002</td>
<td>-0.008</td>
<td>-0.0046</td>
<td>-0.007</td>
<td>0.000</td>
<td>-0.007</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.005)</td>
<td>(0.007)</td>
<td>(0.002)</td>
<td>(0.006)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Fast food price index</td>
<td>-0.112†</td>
<td>-0.073</td>
<td>-0.087</td>
<td>0.004</td>
<td>0.054†</td>
<td>0.055</td>
<td>0.000</td>
<td>0.038</td>
<td>-0.016</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.069)</td>
<td>(0.071)</td>
<td>(0.039)</td>
<td>(0.026)</td>
<td>(0.095)</td>
<td>(0.031)</td>
<td>(0.086)</td>
<td>(0.020)</td>
</tr>
</tbody>
</table>

Note: ‡ $p < 0.10$, † $p < 0.05$ and * $p < 0.01$. Data from the ECLS-K. Cluster-robust standard errors in parentheses. Number of observations = 54,930 (9155 students over six time periods). $y$ is the coefficient on the lagged dependent variable. SES, index of socioeconomic status. Coefficients on the time fixed effects and imputation dummies for missing covariates not shown. See text for further details.
though, are small; a one SD increase in fast food prices is associated with a decline in BMI $z$-score of 0.006 SD for AH and increase in BMI $z$-score of 0.003 SD for E-CRE. Lastly, the estimated impact of household size is negative and statistically insignificant in all cases.

As a note, it is important to recognize that the assumption of homogeneous parameters over time imposed in both the simulations and application above need not actually hold in the data. As explored in Cavaco et al. (2014), the dynamics governing obesity may vary over the life cycle. As such, we truncate the sample into four waves spanning kindergarten through spring third grade and then five waves spanning kindergarten through spring fifth grade. Doing so yields qualitatively similar results. However, the impact of fast food prices on BMI $z$-score over the time period from kindergarten to spring third grade is now negative and statistically significant at conventional levels across all estimators, except for E-CRE. In particular, a one SD increase in fast food prices is associated with approximately 0.03 SD decrease in BMI $z$-score and 0.02 SD decrease in BMI $z$-score for QD-GMM-A(-C) and E-CRE-A(-C), respectively. Results from the truncated samples are in Tables E2–E5 in supplementary Appendix E.

As a robustness check to mitigate concerns related to omitted variables, we augment the previous specification controlling for additional covariates including the number of children’s books in the household, a binary indicator if the mother is present in the household, a binary indicator if the child exercises at least 3 days a week, a binary indicator if the child usually eats a school-provided lunch, and a binary indicator for health insurance. The estimates from the augmented specification for $\gamma$ and for the coefficients on the covariates included in the parsimonious specification are qualitatively similar in magnitude when compared to the baseline specification; see Table E1 in supplementary Appendix E. The one difference that does emerge is a statistically significant, negative association between SES and BMI $z$-score according to the QD-GMM-A estimator, with this negative relationship holding across most of the data. Specifically, the marginal impact of SES on BMI $z$-score remains negative until one gets close to the 99th percentile of the SES distribution. This result is consistent with the idea that households can shift consumption towards more healthy eating alternatives, or demand overall better health, as income increases (Philipson, 2001; Cawley, 2010). It also aligns with prior findings that children in households with more highly educated parents tend to have a lower prevalence of childhood obesity (Lamerz et al., 2005). However, the E-CRE-A(-C) estimators continue to point to a positive association between SES and BMI $z$-score, consistent with adverse health consequences of dual-earner households.

In sum, our proposed estimators, QD-GMM-A(-C) and E-CRE-A(-C), provide some results that differ from those obtained using traditional estimators. First, the E-CRE-A(-C) estimators suggest that BMI $z$-scores are much more persistent (as measured by the effect of the lagged outcome). Second, the E-CRE-A(-C) estimators provide stronger evidence of a monotonic health–income gradient (as measured by SES and BMI $z$-scores), while the QD-GMM-A(-C) estimators point to a non-monotonic, U-shaped relationship.

While the divergence between our proposed estimators and the traditional estimators is not surprising, the disparate estimates of $\gamma$ and the effects of SES are stark and surprising. In the supplementary Appendix we delve into this through the use of an empirical Monte Carlo study and additional Monte Carlo simulations; see Tables I and II. The additional simulations suggest that the QD-GMM-A estimator is more reliable in the application. The evidence suggests that the performance of the E-CRE-A estimator is adversely impacted by the inclusion of binary covariates in the model. As a result, we conclude that the level of persistence in BMI $z$-scores is roughly 0.3 per period in the ECLS-K sample. Moreover, the contemporaneous effects of the health-related inputs appear small, although the health–income gradient is likely to be positively sloped.
5. CONCLUSION

Problems associated with missing data have a lengthy history in econometrics. In the case of dynamic models, missing data may arise not just from the usual, observation-specific sources but also from a data structure that does not align the observation interval with the unit interval from the underlying DGP. As demonstrated here, irregular spacing invalidates the typical estimators of dynamic panel data models for three reasons. First, the coefficient on the lagged dependent variable is no longer constant. Second, first-differencing no longer eliminates the unobserved effect. Third, covariates (and idiosyncratic errors) from the missing time periods are relegated to the error term, invalidating typical instrumental variable strategies and making the covariates endogenous if they are serially correlated. Simulations reveal that the performance of the commonly used Anderson and Hsiao (1981), Arellano and Bond (1991) and Blundell and Bond (1998) estimators can be quite poor when irregular spacing is ignored.

As an alternative, we propose two new estimators. The first utilizes quasi-differencing to remove the unobserved effect and is estimated via GMM. The second is an extended, nonlinear version of the estimator proposed in Everaert (2013). We then augment both estimators by imputing the missing covariates. Simulations reveal that our new estimators perform significantly better than the traditional estimators in the presence of irregular spacing with strictly exogenous covariates.

While both of our new estimators tend to work well, there exist two important cases—at least in finite samples—where our quasi-differenced estimator outperforms the estimator based on Everaert (2013) in terms of estimating the autoregressive parameter. The first case corresponds to the situation where the unobserved effect is much more important, relative to the idiosyncratic error, in the determination of the outcome. The second case is when the model contains binary covariates. Interestingly, our application is covered by these cases. Turning to the assessment of the evolution of BMI during early childhood, our analysis reveals a meaningful impact of failing to account for irregular spacing in the data. In particular, we obtain different results regarding the health–income gradient.

Much work remains, in our view, on the issue of irregular spacing. First, alternative imputation techniques, in particular multiple imputation methods, should be considered. Second, as our estimators are inconsistent when the covariates are not strictly exogenous, this case must be addressed. Third, the DGPs considered in the simulations here are restricted to (i) models that allow for the dynamics to operate only through effects of the lagged outcome, and (ii) models where the initial condition is drawn from a stationary distribution. Investigation into models that allow for deviations from these conditions is necessary. Finally, since dynamic panel models estimated using evenly spaced data but with gaps suffer from some of the problems illuminated here, research into specification testing regarding the ‘true’ definition of a period would be fruitful.

ACKNOWLEDGEMENTS

The authors benefited from helpful suggestions from the Co-editor, Thierry Magnac, and three anonymous referees, as well as discussions with Manuel Arellano, Badi Baltagi, David Drukker, Paul Frijters, Alicia Rambaldi, Robin Sickles and Rusty Tchernis, as well as seminar/conference participants at Syracuse University, SMU, Queensland University of Technology, University of Queensland, Texas Econometrics Camp XVIII and the Econometric Society World Congress.

REFERENCES


