Does Marriage Boost Men's Wages?: Identification of Treatment Effects in Fixed Effects Regression Models for Panel Data

Author(s): Michael E. Sobel

Source: *Journal of the American Statistical Association*, June 2012, Vol. 107, No. 498 (June 2012), pp. 521-529

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association

Stable URL: https://www.jstor.org/stable/23239590

Does Marriage Boost Men's Wages?: Identification of Treatment Effects in Fixed Effects Regression Models for Panel Data

Michael E. SOBEL

Social scientists have generated a large and inconclusive literature on the effect(s) of marriage on men's wages. Researchers have hypothesized that the wage premium enjoyed by married men may reflect both a tendency for more productive men to marry and an effect of marriage on productivity. To sort out these explanations, researchers have used fixed effects regression models for panel data to adjust for selection on unobserved time-invariant confounders, interpreting coefficients for the time-varying marriage variables as effects. However, they did not define these effects or give conditions under which the regression coefficients would warrant a causal interpretation. Consequently, they failed to appropriately adjust for important time-varying confounders and misinterpreted their results. Regression models for panel data with unobserved time-invariant confounders are also widely used in many other policy-relevant contexts and the same problems arise there. This article draws on recent statistical work on causal inference with longitudinal data to clarify these problems and help researchers use appropriate methods to model their data. A basic set of treatment effects is defined and used to define derived effects. Causal models for panel data with unobserved time-invariant confounders are defined and the treatment effects are reexpressed in terms of these models. Ignorability conditions under which the parameters of the causal models are identified from the regression models are given. Even when these hold, a number of interesting and important treatment effects are typically not identified.

KEY WORDS: Causal inference; Fixed effects; Longitudinal data; Marriage premium; Unobserved confounders; Unobserved variables.

1. INTRODUCTION

Married men have higher hourly and annual earnings than ostensibly similar unmarried men (Schoeni 1995). Economists have advanced various explanations. Hill (1979) argued that employers may favor married men. Under signaling or statistical discrimination, employers pay a premium to married men because they believe them to be more responsible and productive. Reed and Harford (1989) argued that married men receive a compensating wage differential for taking jobs with fewer nonpecuniary benefits. Selection on income can occur if women view higher wage men as more desirable partners.

More recent efforts to account for the marital wage premium have focused on two somewhat different explanations. One is that marriage makes men more productive, for example, by allowing them to specialize in building human capital and engaging in labor force activities, while wives specialize in home activities (Becker 1981). A second is that unobserved variables such as ability and psychological dispositions are related to both productivity (hence wages) and the probability of marrying (Cohen and Haberfeld 1991).

To sort out the contributions of productivity and selection on unobservables to the marital wage premium, most researchers, beginning with Cohen and Haberfeld (1991) and Korenman and Neumark (1991) (hereafter KN), have used panel data, comparing coefficients for marital status variables in pooled cross-sectional regressions that adjust for measured time-varying and time-invariant (baseline) confounders to analogous coefficients in fixed effects regressions that also adjust for unobserved baseline confounders.

The findings are mixed. Using the Panel Study of Income Dynamics for 1979–1982, Cohen and Haberfeld (1991) found no effect of marriage. Using the National Longitudinal Survey (NLS) of Young Men for 1976–1980, with a sample of white males who "completed" (did not obtain additional schooling during the study period) school by 1976, KN argued that 80%–90% of the wage premium is due to the effect of marriage on productivity. But Cornwell and Rupert (1997), who extended KN's analysis by including data for 1971, failed to replicate KN's results. Gray (1997), using similar regressors as KN, compared NLS data for 1976–1980 with National Longitudinal Survey of Youth (NLSY) data for 1989–1993; in the NLSY data, there is no effect of martial status. Rodgers and Stratton (2010), using the NLSY for 1988–1994 with men who "completed" school by 1988, also found no effect of marital status; but using the data for 1985–1992 for white males who "completed" school by 1985, Akerlof (1998) obtained results similar to KN.

Several other studies used fixed effects regression models with sibling (Loh 1996) or twin (Antonovics and Town 2004; Krashinsky 2004; Isaacson 2007) data. A few studies have used other methodologies (Chun and Lee 2001; Mincy, Hill, and Sinkewicz 2009). Again, the findings are mixed.

The mixed results reflect differences, sometimes minor, among studies in samples, places, times, and other details. The results are also questionable. For example, the twin studies control for observed and unobserved baseline variables such as family background and genetic makeup, as well as some observed variables that vary within pairs such as education, but not for others such as mental and physical health status. Antonovics and Town (2004) also excluded part-time workers; this is problematic (Rosenbaum 1984) as hours worked may be affected by marital status.

Michael E. Sobel, Professor in the Department of Statistics, Columbia University, New York, NY 10027 (E-mail: *michael@stat.columbia.edu*). For helpful comments, I am grateful to Aimee Dechter, participants in the New York University Applied Statistics Seminar Series, an associate editor, and several reviewers.

Journal of the American Statistical Association
June 2012, Vol. 107, No. 498, Applications and Case Studies
DOI: 10.1080/01621459.2011.646917

Similar criticisms apply to the longitudinal studies. For example, educational attainment may be affected by marriage, so respondents who obtained additional schooling during the study period should not have been excluded (see Section 5 for a formal analysis). Further, with panel data, there are several effects of possible interest. But researchers neither make clear what effects are of interest nor conditions under which model coefficients warrant a causal interpretation. This has led to confusion and misinterpretation of results.

This article has several aims. The first is to ascertain what is actually known about the effect(s) of marriage on men's wages. Recent policies promoting marriage hinge in part on the belief that marriage enhances productivity, thereby increasing family income and reducing child poverty (Avner and Lerman 2007; Mincy, Hill, and Sinkewicz 2009). More generally, if marriage affects productivity, aggregate labor force quality will depend on the marital distribution of the population.

Second, fixed effects regression models for panel data with unobserved baseline confounders are widely used in criminology, demography, economics, political science, sociology, and other fields to study and inform policy on many significant topics, ranging from the effects of agricultural shocks on the use of child labor (Beegle, Dehejia, and Gatti 2006) to the effects of state governance on violent crime (Neumayer 2003). The same problems that arise in the literature on marriage and men's wages arise here as well.

To clarify these issues, I build on recent statistical work on causal inference with longitudinal data, defining "basic treatment effects." These are also used to define other effects of interest, for example, "contemporaneous" effects. Conditions for the identification of the various effects are also given.

To maximize readability, I take up the most common case of a linear model for a continuous dependent variable. However, similar considerations arise with nonlinear models and other types of dependent variables, for example, binary and count variables. It is hoped these results will help future workers to clearly define estimands of interest, to understand what conditions must be met to estimate these, and to use appropriate methods to do so.

The article proceeds as follows. Section 2 describes current practice, setting out the two most commonly used "fixed effects" regression models. Section 3 defines causal estimands of interest. Fixed effects causal models are then constructed and the estimands are reexpressed using these models. Section 4 gives conditions under which the estimands are identified from the regression models. Section 5 uses the framework constructed in Section 3 and the results on identification to critically reexamine the empirical literature on the marital wage premium. Section 6 concludes.

2. FIXED EFFECTS REGRESSION MODELS

For $i = 1, \ldots, n$ and $t = 1, \ldots, T$, with $T \geq 2$, let Y_{it} be the wage measure for unit i in period t, let $\mathbf{X}_{it}$ be a column vector of time-varying confounders, and let $D_{it} = 1$ if i is married in period t, 0 otherwise (the extension to treatments with more levels is straightforward, as in Section 5). Note that $\mathbf{X}_{it}$ is "prior" to D_{it}, as when (a) $\mathbf{X}_{it}$ is measured before period t and D_{it} is measured during period t, or (b) $\mathbf{X}_{it}$ refers to a time in period t before D_{it}, or (c) $\mathbf{X}_{it}$ and D_{it} are concurrent, but the value of $\mathbf{X}_{it}$ is the same whether or not treatment is received. Also D_{it} is prior to Y_{it}, Y_{it} is prior to $\mathbf{X}_{i,t+1}$, and so on. Let $\mathbf{W}_i$ and $\boldsymbol{\eta}_i$ denote vectors of observed and unobserved baseline confounders, respectively, and let $U_i \equiv f(\mathbf{W}_i, \boldsymbol{\eta}_i)$ and V_t denote unobserved unit and period-specific terms, respectively. Here U_i and $\mathbf{W}_i$ are prior to $t = 1$. Cumulative histories are denoted using a bar over the argument: for example, $\bar{Y}_{i\{j>t\}} = (Y_{i,t+1}, \ldots, Y_{iT})$. For the special case of a cumulative history beginning in period 1 and continuing through period t, the set $\{1, \ldots, t\}$ is indexed only by t: for example, $\bar{D}_{it} = (D_{i1}, \ldots, D_{it})$. The vectors $\mathbf{Q}_i = (\bar{Y}_{iT}, \bar{\mathbf{X}}_{iT}, \bar{D}_{iT}, \mathbf{W}_i, \bar{V}_T, U_i)$ are assumed to be independent and identically distributed (iid) with the same distribution as the random vector $\mathbf{Q} = (\bar{Y}_T, \bar{\mathbf{X}}_T, \bar{D}_T, \mathbf{W}, \bar{V}_T, U)$.

Researchers typically model Y_{it} using a regression of the form

$$Y_{it} = \boldsymbol{\beta}'\vec{\mathbf{X}}_{it} + \tau(\bar{D}_{it}, \mathbf{W}_{1i}) + V_t + U_i + \varepsilon_{it}, \tag{1}$$

where $\vec{\mathbf{X}}_{it} = (1, \mathbf{X}'_{it})'$, $\boldsymbol{\beta}$ is a column vector of coefficients whose first element is an intercept, ε_{it} is an error with mean 0, and $\mathbf{W}_{1i}$ is a subvector of $\mathbf{W}_i$. Functions of $\tau(\bar{D}_{it}, \mathbf{W}_{1i})$ are interpreted as effects. Although this formulation allows the values of these "effects" to vary across subpopulations defined by pretreatment variables, for example, race, in many applications $\tau(\bar{D}_{it}, \mathbf{W}_{1i}) = \tau D_{it}$ and τ is interpreted as the effect. Sometimes more terms are added, for example, $\tau(\bar{D}_{it}, \mathbf{W}_{1i}) = \tau_1 D_{it} + \tau_2 D_{i,t-1}$. Note also that by including interactions of time-varying variables with time and/or elements of $\mathbf{W}_i$, coefficients in $\boldsymbol{\beta}$ can depend on time and/or time-invariant variables $\mathbf{W}_i$.

The properties of ε_{it}, U_i, and V_t in (1) are the key to interpretation, estimation, and inference for the model parameters. When $\mathbf{X}_{it}$ does not contain lagged values of the dependent variable, a "strict exogeneity" assumption (Chamberlain 1982) is often made: for all i and t,

$$E(\varepsilon_{it} \mid \bar{\mathbf{X}}_{iT}, \bar{D}_{iT}, \mathbf{W}_{1i}, \bar{V}_T, U_i) = 0. \tag{2}$$

Equations (1) and (2) jointly imply

$$\begin{aligned} E(Y_{it} \mid \bar{\mathbf{X}}_{iT}, \bar{D}_{iT}, \mathbf{W}_{1i}, \bar{V}_T, U_i) \\ = \boldsymbol{\beta}'\vec{\mathbf{X}}_{it} + \tau(\bar{D}_{it}, \mathbf{W}_{1i}) + V_t + U_i. \end{aligned} \tag{3}$$

When U_i and V_t are treated as constants, a fixed effects regression model with strict exogeneity (FER-STE) is obtained. To identify $\boldsymbol{\beta}$ and $\tau(\bar{d}_t, \mathbf{w}_1)$ under (1) and (2), a restriction on each of the collections of U_i and V_t must be imposed, for example, $U_1 = 0$, $V_1 = 0$; subject to such restrictions, hereafter assumed, $\boldsymbol{\beta}$ and $\tau(\bar{d}_t, \boldsymbol{w}_1)$ are assumed to be identified.

In empirical work, FER-STE models are often used, and if $\tau(\bar{D}_{it}, \mathbf{W}_{1i})$ is specified so that the model is linear, for example, $\tau(\bar{D}_{it}, \mathbf{W}_{1i}) = \tau D_{it} W_{pi}$, where W_{pi} is the p(th) element of $\mathbf{W}_{1i}$, the "least-square dummy variable estimator" yields, with T fixed and $n \to \infty$, consistent and asymptotically normal estimates of $\boldsymbol{\beta}$ and $\tau(\bar{D}_t, \mathbf{w}_1)$ under general conditions; when the errors are also iid normal, ordinary least squares (OLS) is maximum likelihood. Estimates of the U_i are not, however, consistent.

Assumption (2) is quite strong, implying that a shock to wages in period t is uncorrelated with wages and time-varying regressors in later periods. But suppose now that wages in period $t + 1$ are assumed to depend, in addition, on period t wages, prompting the investigator to include the lagged dependent variable Y_{it} as an element of $\mathbf{X}_{i,t+1}$. In this case, as it would be unreasonable to assume that ε_{it} and Y_{it} are uncorrelated, the strict exogeneity assumption (2) will be violated. In such instances, it is more reasonable to assume that ε_{it} is uncorrelated with prior variables only. This is implied by the weaker assumption of "sequential exogeneity" (Chamberlain 1992): for all i and t,

$$E(\varepsilon_{it} \mid \bar{\mathbf{X}}_{it}, \bar{D}_{it}, \mathbf{W}_{1i}, \bar{V}_t, U_i) = 0. \tag{4}$$

Equations (1) and (4) jointly imply

$$E(Y_{it} \mid \bar{\mathbf{X}}_{it}, \bar{D}_{it}, \mathbf{W}_{1i}, \bar{V}_t, U_i) = \boldsymbol{\beta}'\vec{\mathbf{X}}_{it} + \tau(\bar{D}_{it}, \mathbf{W}_{1i}) + V_t + U_i. \tag{5}$$

If assumption (4) is made and U_i and V_t are treated as constants in (1), a fixed effects regression model with sequential exogeneity (FER-SEE) is obtained. To obtain consistent estimates, instrumental variables or the generalized methods of moments can be used. For further material on identification and estimation when strict exogeneity does not hold, see, for example, Arellano and Honoré (2001), Hsiao (2003), and Lai, Small, and Liu (2008).

3. TREATMENT EFFECTS IN FIXED EFFECTS CAUSAL MODELS

Potential outcomes and treatment effects on a basis set of regimes are defined. Fixed effects causal models analogous to the regression models mentioned earlier are constructed and used to reexpress the basic effects. Effects derived from the basic effects are also defined. Conditions under which the various effects are identified using the fixed effects regression models are given in Section 4.

3.1 Basic Effects: Definitions

Assume that all units i can be exposed to any regime $\bar{d}_T$ in a basis set $M_T \subseteq \{0, 1\}^T$. For all i and $\bar{d}_T \in M_T$, suppose that i's outcomes under regime $\bar{d}_T$ do not depend on regimes to which other units may be exposed; this is the stable unit treatment value assumption (SUTVA) (Rubin 1980), allowing us to write i's potential outcomes under regime $\bar{d}_T$ as $\bar{Y}_i(\bar{d}_T) = (Y_{i1}(\bar{d}_T), \ldots, Y_{iT}(\bar{d}_T))$. Also, suppose that the temporal ordering assumption (Neugebauer and van der Laan 2006) $Y_{it}(\bar{d}_T) = Y_{it}(\bar{d}_T^*)$ for all $\bar{d}_T = (\bar{d}_t, \bar{d}_{\{j>t\}})$ and $\bar{d}_T^* = (\bar{d}_t^*, \bar{d}_{\{j>t\}}^*)$ for which $\bar{d}_t = \bar{d}_t^*$ holds, allowing period t potential outcomes to be written as $Y_{it}(\bar{d}_t)$; this assumption is similar to the no-anticipation assumption (Abbring and van den Berge 2003) in the econometric literature. The assumptions that i's period t potential wages depend neither on his subsequent marital status nor on the marital history of other units seem reasonable.

Treatment status may also affect the time-varying confounders. Let $\mathbf{X}_{it}(\bar{d}_T)$ denote i's outcome under regime $\bar{d}_T$, $t = 2, \ldots, T$, and as mentioned earlier, assume $\mathbf{X}_{it}(\bar{d}_T) = \mathbf{X}_{it}(\bar{d}_{t-1})$. For $t = 1$, $\mathbf{X}_{i1}(\bar{d}_0) \equiv \mathbf{X}_{i1}$.

Treatment effects compare potential outcomes under different regimes. The unit effects $Y_{it}(\bar{d}_t) - Y_{it}(\bar{d}_t^*)$, $t = 1, \ldots, T$, cannot be observed, and interest typically centers on average treatment effects. The basic effects considered here are special cases of the form:

$$E(Y_t(\bar{d}_t) - Y_t(\bar{d}_t^*) \mid \{\mathbf{X}_j(\bar{d}_{j-1})\}_{j=1}^t \in S_{1t}, \{\mathbf{X}_j(\bar{d}_{j-1}^*)\}_{j=1}^t \in S_{2t}, \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t), \tag{6}$$

where $\bar{D}_{k(t)}$ is a subsequence of $\bar{D}_T$. This is the average effect of $\bar{d}_t$ versus $\bar{d}_t^*$ for respondents in subpopulation A and principal stratum (S_{1t}, S_{2t}) (Frangakis and Rubin 2002) who followed a regime or subregime in B_t.

Lechner and Miquel (2005) (hereafter LM) and Lechner (2009) considered the special case of (6) with $S_{1t} = \mathrm{Im}(\{\mathbf{X}_j(\bar{d}_{j-1})\}_{j=1}^t)$, $S_{2t} = \mathrm{Im}(\{\mathbf{X}_j(\bar{d}_{j-1}^*)\}_{j=1}^t)$, where $\mathrm{Im}(\boldsymbol{f})$ is the image of $\boldsymbol{f}$. As an example, let A be the set of black men and consider the effect $\bar{d}_t = (0, \ldots, 0, 1)$ versus $\bar{d}_t^* = (0, \ldots, 0)$ on period t wages for black men who have not married by period t $(\bar{D}_{k(t)} = \bar{D}_t, B_t = \{\bar{d}_t^*\})$. For $t = 1$, this reduces to the effect of treatment on the untreated for black men. The case $\bar{D}_{k(t)} = \bar{D}_t$, $B_t = \{0, 1\}^t$ is also treated in the statistical literature on marginal structural models (e.g., Hernán, Brumback, and Robins 2001; Robins and Hernán 2009), and if $\mathrm{A} = \mathrm{Im}(\mathbf{W})$, the average effect of regime $\bar{d}_t$ versus $\bar{d}_t^*$ is obtained; LM call this a "dynamic average treatment effect." For two regimes identical through period $t-1$ ($\bar{d}_t = (\bar{d}_{t-1}, 1)$, $\bar{d}_t^* = (\bar{d}_{t-1}^*, 0)$, $\bar{d}_{t-1} = \bar{d}_{t-1}^*$) and thus $\{\mathbf{X}_j(\bar{d}_{j-1})\}_{j=1}^t = \{\mathbf{X}_j(\bar{d}_{j-1}^*)\}_{j=1}^t$, Almirall, Ten Have, and Murphy (2010) consider (6) with $\bar{D}_{k(t)} = \bar{D}_t$, $B_t = \{0, 1\}^t$, and $S_{1t} = S_{2t} = \{\boldsymbol{x}_j(\bar{d}_{j-1})\}_{j=1}^t$, calling it a conditional intermediate causal effect of treatment.

3.2 Fixed Effects Causal Models

Causal models for the potential outcomes in the basis set M_T may now be constructed: for all i, t, and $\bar{d}_T \in M_T$,

$$Y_{it}(\bar{d}_t) = \boldsymbol{\beta}_c'\vec{\mathbf{X}}_{it}(\bar{d}_{t-1}) + \tau_c(\bar{d}_t, \mathbf{W}_{1i}) + V_t + U_i + \varepsilon_{it}(\bar{d}_t), \tag{7}$$

where $\boldsymbol{\beta}_c$ is a vector of regression coefficients, $\vec{\mathbf{X}}_{it}(\bar{d}_{t-1}) = (1, \mathbf{X}_{it}'(\bar{d}_{t-1}))'$, $\tau_c(\bar{d}_t, \mathbf{W}_{1i})$ is a function of $\bar{d}_t, \mathbf{W}_{1i}$, depending on one or more parameters, and $\varepsilon_{it}(\bar{d}_t)$ is a potential error.

Two assumptions about the potential errors are considered here. Either they are "strictly mean independent" of time-varying confounders, period effects, and baseline variables: for all i and t,

$$E\big(\varepsilon_{it}(\bar{d}_t) \mid \{\mathbf{X}_{i,j}(\bar{d}_{j-1})\}_{j=1}^T, \mathbf{W}_{1i}, \bar{V}_T, U_i\big) = 0, \tag{8}$$

or they are "sequentially mean independent" of prior time-varying confounders, period effects, and baseline variables: for all i and t,

$$E\big(\varepsilon_{it}(\bar{d}_t) \mid \{\mathbf{X}_{i,j}(\bar{d}_{j-1})\}_{j=1}^t, \mathbf{W}_{1i}, \bar{V}_t, U_i\big) = 0. \tag{9}$$

Assumptions (8) and (9) for the potential errors are analogous, respectively, to the strict and sequential exogeneity assumptions (2) and (4) for the errors in regression models FER-STE and FER-SEE.

Combining (7) with (8) and treating U_i as a constant yields a fixed effects causal model with strictly mean independent errors (FEC-STI):

$$E\big(Y_{it}(\bar{d}_t) \mid \{\mathbf{X}_{i,j}(\bar{d}_{j-1})\}_{j=1}^T, \mathbf{W}_{1i}, \bar{V}_T, U_i\big) = \boldsymbol{\beta}_c'\vec{\mathbf{X}}_{it}(\bar{d}_{t-1}) + \tau_c(\bar{d}_t, \mathbf{W}_{i1}) + V_t + U_i. \tag{10}$$

Combining (7) with (9) and treating U_i as a constant yields a fixed effects causal model with sequentially mean independent errors (FEC-SEI):

$$E\big(Y_{it}(\bar{d}_t) \mid \{\mathbf{X}_{i,j}(\bar{d}_{j-1})\}_{j=1}^t, \mathbf{W}_{1i}, \bar{V}_t, U_i\big) = \boldsymbol{\beta}_c' \vec{\mathbf{X}}_{it}(\bar{d}_{t-1}) + \tau_c(\bar{d}_t, \mathbf{W}_{i1}) + V_t + U_i. \quad (11)$$

Under model FEC-STI, potential wages in period t do not depend on past or future time-varying confounders, given current period and baseline confounders. Under model FEC-SEI, potential wages in period t do not depend on past confounders, given current period and baseline confounders. If model FEC-STI holds, model FEC-SEI holds, as is evident from (8) and (9).

3.3 Model-Based Effects: Basic and Derived

Under either causal model, the basic effect (6), with $S_{1t} = \text{Im}(\{\mathbf{X}_j(\bar{d}_{j-1})\}_{j=1}^t)$ and $S_{2t} = \text{Im}\,(\{\mathbf{X}_j(\bar{d}_{j-1}^*)\}_{j=1}^t)$, is

$$E(\tau_c(\bar{d}_t, \mathbf{W}_1) - \tau_c(\bar{d}_t^*, \mathbf{W}_1) + \boldsymbol{\beta}_c'(\vec{\mathbf{X}}_t(\bar{d}_{t-1}) - \vec{\mathbf{X}}_t(\bar{d}_{t-1}^*)) + \varepsilon_t(\bar{d}_t) - \varepsilon_t(\bar{d}_t^*) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t). \quad (12)$$

This is the average effect in period t of partial regime $\bar{d}_t$ versus $\bar{d}_t^*$ in the subpopulation A following a partial regime in B_t, for example, the effect, previously considered, of first marrying in period t for black men who have not married by period t. When A is the subpopulation of respondents with $\mathbf{W}_1 = \mathbf{w}_1$, note that $E(\tau_c(\bar{d}_t, \mathbf{W}_1) - \tau_c(\bar{d}_t^*, \mathbf{W}_1) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t) = \tau_c(\bar{d}_t, \mathbf{w}_1) - \tau_c(\bar{d}_t^*, \mathbf{w}_1)$.

When B_t includes all possible histories through period t ($B_t = \{0, 1\}^t$) and $A = \{\mathbf{w}_1 : \mathbf{w}_1 \in C\}$, where C is a subset of $\text{Im}(\mathbf{W}_1)$, $E(\varepsilon_t(\bar{d}_t) - \varepsilon_t(\bar{d}_t^*) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t) = 0$ in (12). In this case, even if $\tau_c(\bar{d}_t, \mathbf{W}_1) = \tau_c(\bar{d}_t)$, the effect (12) nonetheless depends on $\mathbf{W}$ if the effects of treatments on the time-varying confounders depend on $\mathbf{W}$. For example, let $s < t$, $\bar{d}_t = (\bar{d}_s, 1, \ldots, 1)$, $\bar{d}_t^* = (\bar{d}_s^*, 0, \ldots, 0)$, $\bar{d}_s = \bar{d}_s^* = (0, \ldots, 0)$, $\tau_c(\bar{d}_t, \mathbf{W}_1) = \sum_{j=1}^{t-s} \tau_c^j$. The basic effect (12) is then $\sum_{j=1}^{t-s} \tau_c^j + \boldsymbol{\beta}_c' E(\vec{\mathbf{X}}_t(\bar{d}_{t-1}) - \vec{\mathbf{X}}_t(\bar{d}_{t-1}^*) \mid \mathbf{W} \in A)$. For $t = T, T \to \infty, \sum_{j=1}^{T-s} \tau_c^j = \tau_c/(1 - \tau_c)$ if $|\, \tau_c \,| < 1$.

For $B_t = \{0, 1\}^t$, $A = \text{Im}(\mathbf{W})$, the average effect of $\bar{d}_t$ versus $\bar{d}_t^*$ is obtained:

$$E(\tau_c(\bar{d}_t, \mathbf{W}_1) - \tau_c(\bar{d}_t^*, \mathbf{W}_1) + \boldsymbol{\beta}_c'(\vec{\mathbf{X}}_t(\bar{d}_{t-1}) - \vec{\mathbf{X}}_t(\bar{d}_{t-1}^*))). \quad (13)$$

Under model FEC-SEI, with $B_t = \{0, 1\}^t$, $\bar{d}_{t-1} = \bar{d}_{t-1}^*$, (12) is the conditional intermediate causal effect considered by Almirall, Ten Have, and Murphy (2010).

Finally, note that neither (12) nor (13) depends on the time-varying regressors if $\mathbf{X}_t(\bar{d}_{t-1}) = \mathbf{X}_t(\bar{d}_{t-1}^*)$, which occurs if these variables are not affected by prior treatments and/or $\bar{d}_t$ and $\bar{d}_t^*$ agree through period $t-1$.

Derived effects may be generated from the basic effects. A general exposition is easily given, but at the expense of further complicating the notation. Thus, only two effects of likely interest are considered: (a) the effect of treatment in period 1 on the period t outcome and (b) the contemporaneous effect of treatment in period t.

A partial regime $\bar{d}_t$ is "feasible" if there is a vector $\bar{d}_{\{j>t\}} \in \{0, 1\}^{T-t}$ such that $(\bar{d}_t, \bar{d}_{\{j>t\}})$ belongs to the basis set M_T. Let M_t denote the set of feasible $\bar{d}_t$ and suppose $M_1 = \{0, 1\}$. Let $Y_{it}(D_1 = d_1) \equiv Y_{it}(d_1, \bar{D}_{i\{j:2\le j\le t\}}(d_1))$ denote i's outcome in period t if treatment d_1 is received in period 1. The effect of treatment in period 1, for respondents in subpopulation A with observed history $\bar{D}_{k(t)} \in B_t$, is

$$E(Y_t(D_1 = 1) - Y_t(D_1 = 0) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t). \quad (14)$$

Now let $M_{1t} = \{(1, \bar{d}_{\{j:2\le j\le t\}}) \in M_t\}$, and let $p(\bar{d}_{\{j:2\le j\le t\}}(1) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t)$ denote the (potential) conditional probability that $\bar{D}_{\{j:2\le j\le t\}}(1) = \bar{d}_{\{j:2\le j\le t\}}$ if all units in subpopulation A with treatment history in B_t receive treatment in period 1; M_{0t} and $p(\bar{d}_{\{j:2\le j\le t\}}(0) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t)$ are defined similarly.

The effect (14) is then

$$\begin{aligned}\sum_{M_{1t}} E\left(Y_t\left(1, \bar{d}_{\{j:2\le j\le t\}}\right) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t\right) &\times p(\bar{d}_{\{j:2\le j\le t\}}(1) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t) \\ - \sum_{M_{0t}} E(Y_t(0, \bar{d}_{\{j:2\le j\le t\}}) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t) &\times p(\bar{d}_{\{j:2\le j\le t\}}(0) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t), \quad (15)\end{aligned}$$

which is easily expressed using either causal model by substituting equivalent expressions for $E(Y_t(d_1, \bar{d}_{\{j:2\le j\le t\}}) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t)$.

To define the contemporaneous effect, suppose that there exist $\bar{d}_{t-1}$ and $\bar{d}_{t-1}^*$ such that $(\bar{d}_{t-1}, 0) \in M_t$, $(\bar{d}_{t-1}^*, 1) \in M_t$, and let $Y_{it}(D_t = d_t) \equiv Y_{it}(\bar{D}_{i,t-1}, d_t)$ denote i's outcome in period t under treatment d_t. The contemporaneous effect of treatment in period t on the period t outcome, for respondents in subpopulation A with treatment history in B_t, is

$$E(Y_t(D_t = 1) - Y_t(D_t = 0) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t). \quad (16)$$

The case $A = \text{Im}(\mathbf{W})$, $\bar{D}_{k(t)} = \bar{D}_t$, $B_t = \{0, 1\}^t$ is of special interest: (16) is then the average effect of treatment in period t. Also, typically $(\bar{d}_{t-1}, 0)$ and $(\bar{d}_{t-1}, 1)$ are feasible for all $\bar{d}_{t-1} \in M_{t-1}$; then, under either causal model, (16) is

$$E \sum_{M_{t-1}} (\tau_c(\bar{d}_{t-1}, 1, \mathbf{W}_1) - \tau_c(\bar{d}_{t-1}, 0, \mathbf{W}_1)) p(\bar{d}_{t-1}) \mid \mathbf{W}_1). \quad (17)$$

4. IDENTIFICATION OF TREATMENT EFFECTS FROM FIXED EFFECTS REGRESSION MODELS

Identification of the basic effect (12) depends on identifiability of the causal models, specifically (a) the parameters $\boldsymbol{\beta}_c$ and functions $\tau_c(\bar{d}_t, \mathbf{w}_1)$, (b) the expectations $E(\varepsilon_t(\bar{d}_t) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t)$, and (c) the effects of treatments on the time-varying confounders $E(\mathbf{X}_t(\bar{d}_{t-1}) - \mathbf{X}_t(\bar{d}_{t-1}^*) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t)$. These issues are now addressed in turn. Identification of the derived effects depends on additional considerations, discussed later.

4.1 Model Identification

As each respondent has only one observed history, parameters of the causal models are not identified without additional assumptions. Suppose U and $\bar{V}_T$ were observed. Sequential ignorability (conditional exchangeability) extends the ignorability assumption for a one-period study (Rosenbaum and Rubin 1983) and is a standard identification assumption in the literature on longitudinal causal inference (Robins and Hernán 2009): for all t,

$$\{Y_j(\bar{d}_j)\}_{j=t}^{T} \perp\!\!\!\perp D_t \mid \bar{Y}_{t-1}, \bar{\mathbf{Z}}_t, \bar{D}_{t-1} = \bar{d}_{t-1}, \mathbf{W}_1, \bar{V}_t, U, \quad (18)$$

where $\perp\!\!\!\perp$ denotes statistical independence, $\mathbf{Z}_t = \mathbf{X}_t$ if $Y_{t-m} \notin \mathbf{X}_t$, $m = 1, \ldots, t-1$, and otherwise $\mathbf{Z}_t$ is the subvector of $\mathbf{X}_t$ obtained by deleting those elements $Y_{t-m} \in \mathbf{X}_t, m = 1, \ldots, t-1$. LM incorporate (18) into their "weak dynamic conditional independence assumption." As it is an assumption about potential outcomes and only one history per respondent is observed, (18) does not in general permit empirical refutation. To estimate treatment effects using (18), a marginal structural model (Robins 1998) estimated using inverse probability of treatment weighting (IPTW) can be used, as in Lechner (2009); this requires estimating the treatment probabilities, conditional on the past. Or a structural nested mean model might be used; this requires estimating means of the potential outcomes conditional on the past (Robins and Hernán 2009). But if there are also unmeasured confounders, as here, further assumptions would be needed to estimate such models.

Theorem 1 uses an assumption similar to (18) that includes the unobserved variables, enabling identification of the parameters of model FEC-SEI from the parameters of model FER-SEE. The proof is given in the Appendix.

Theorem 1. For all t, if

$$\{\mathbf{X}_j(\bar{d}_{j-1}), Y_j(\bar{d}_j), V_j\}_{j=t+1}^{T}, Y_t(\bar{d}_t) \perp\!\!\!\perp D_t \mid \bar{\mathbf{X}}_t, \bar{D}_{t-1} = \bar{d}_{t-1}, \mathbf{W}_1, \bar{V}_t, U, \quad (19)$$

$$E(Y_t \mid \bar{\mathbf{X}}_t, \bar{D}_t = \bar{d}_t, \mathbf{W}_1, \bar{V}_t, U) = E(Y_t(\bar{d}_t) \mid \{\mathbf{X}_m(\bar{d}_{m-1})\}_{m=1}^{t}, \mathbf{W}_1, \bar{V}_t, U). \quad (20)$$

If model FEC-SEI holds, $\bar{D}_{k(t)} = \bar{D}_{t-1}$, $B_t = \{\bar{d}_{t-1}\}$, $A = \{\mathbf{w}_1 : \mathbf{w}_1 \in C, C \subseteq \mathrm{Im}(\mathbf{W}_1)\}$, then $E(\varepsilon_t(\bar{d}_t) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t) = 0$.

Consequently, if (19) and causal model FEC-SEI hold, $\boldsymbol{\beta}_c$ and $\tau_c(\bar{d}_t, \mathbf{w}_1)$, for all $\mathbf{w}_1$, are equal, respectively, to $\boldsymbol{\beta}$ and $\tau(\bar{d}_t, \mathbf{w}_1)$ in regression model FER-SEE.

Assumption (19) differs from the usual sequential ignorability assumption in several ways. First, the conditioning set includes the unobserved variables U and V_t, $t = 1, \ldots, T$, and it may or may not include lagged wages. Second, in conjunction with the positivity assumption that the probability of receiving any treatment in period t is nonzero in all nontrivial subpopulations defined by previous treatments and confounders, (18) implies that the means $E(Y_t(\bar{d}_t))$ are identified (Robins and Hernán 2009). But it does not imply (20). In (19), current marital status is ignorable not only with respect to future potential wages but also with respect to future potential confounders. Though stronger than (18), it would be odd if treatment were ignorable with respect to future potential outcomes, but not the components of the model for these.

Although also similar to the "strong dynamic conditional independence" assumption based on potential confounders in LM, (19) also includes unobserved confounders, and the temporality and temporal consistency assumptions here replace the exogeneity conditions in that assumption.

Further, (19) holds for $\bar{D}_{t-1} = \bar{d}_{t-1}$ only, not also for values $\bar{d}^*_{t-1} \neq \bar{d}_{t-1}$, as in LM. To understand the substantive implications of the stronger assumption in LM, note that under LM, $E(\varepsilon_t(d_t) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t) = 0$ in Theorem 1 for regimes $\bar{d}^*_{t-1} \neq \bar{d}_{t-1}$. Suppose that $\bar{d}_t = (0, \ldots, 0, 1)$ and $\bar{d}^*_t = (0, \ldots, 0)$. Under Theorem 1, the average effect (13) in period t of first marrying in period t would then be identified, as would the average effect in the subpopulation of black men unmarried by period t previously considered. But now suppose $d_t = (1, \ldots, 1)$. Suppose also that treatment does not affect the time-varying regressors. Under (19) and model FEC-SEI, the average effect (12) in period t of being married versus unmarried in periods $1, \ldots, t$ in the subpopulation of black men who have not married by period t is not identified as $E(\varepsilon_t(d_t) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t)$ need not equal 0 for $B_t = (\bar{d}^*_{t-1}, 0)$; but if (19) is replaced by the stronger assumption in LM, $E(\varepsilon_t(d_t) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t) = 0$ for $B_t = (\bar{d}^*_{t-1}, 0)$ and (12) is identified.

Theorem 2 gives conditions under which model FER-STE identifies parameters of model FEC-STI. The proof is given in the Appendix.

Theorem 2. If (19) holds for all t and for $t = 1, \ldots, T-1$,

$$Y_t(\bar{d}_t) \perp\!\!\!\perp \bar{D}_{\{j>t\}} \mid \{\mathbf{X}_m(\bar{d}_{m-1})\}_{m=1}^{T}, \bar{D}_t = \bar{d}_t, \mathbf{W}_1, \bar{V}_T, U, \quad (21)$$

then

$$E(Y_t \mid \bar{\mathbf{X}}_T, \bar{D}_T = \bar{d}_T, \mathbf{W}_1, \bar{V}_T, U) = E(Y_t(\bar{d}_t) \mid \{\mathbf{X}_m(\bar{d}_{m-1})\}_{m=1}^{T}, \mathbf{W}_1, \bar{V}_T, U). \quad (22)$$

If model FEC-STI holds, $\bar{D}_{k(t)} = \bar{D}_{t-1}$, $B_t = \{\bar{d}_{t-1}\}$, $A = \{\mathbf{w}_1 : \mathbf{w}_1 \in C, C \subseteq \mathrm{Im}(\mathbf{W}_1)\}$, then $E(\varepsilon_t(\bar{d}_t) \mid \mathbf{W_1} \in A, \bar{D}_{k(t)} \in B_t) = 0$.

Consequently, if (19), (21), and causal model FEC-STI hold, $\boldsymbol{\beta}_c$ and $\tau_c(\bar{d}_t, \mathbf{w}_1)$, for all $\mathbf{w}_1$, are equal, respectively, to $\boldsymbol{\beta}$ and $\tau(\bar{d}_t, \mathbf{w}_1)$ in regression model FER-STE.

The additional condition (21) says that potential wages in period t do not predict future marital history, given baseline and prior time-varying confounders and future time-varying confounders. Equations (19) and (21) jointly imply

$$Y_t(\bar{d}_t) \perp\!\!\!\perp \bar{D}_T \mid \{\mathbf{X}_m(\bar{d}_{m-1})\}_{m=1}^{T}, \mathbf{W}_1, \bar{V}_T, U. \quad (23)$$

This is quite strong: a respondent's marital history does not predict his potential wages in any period, given baseline and potential confounders.

Theorems 1 and 2 give sufficient conditions for the identification of treatment effects that do not depend on time-varying regressors $\mathbf{X}_t(\bar{d}_{t-1})$. If either holds, period t treatment effects comparing histories identical through period $t-1$ ($\bar{d}_t = (\bar{d}_{t-1}, 1)$, $\bar{d}^*_t = (\bar{d}^*_{t-1}, 0)$, $\bar{d}_{t-1} = \bar{d}^*_{t-1}$) are identified, in any subpopulation $\mathbf{w}_1$ of $\mathbf{W}_1$ (Equation 12) and on average (Equation 13). If both $d_t = 0$ and $d_t = 1$ can follow every feasible history $\bar{d}_{t-1}$, the contemporaneous effect (16) is also identified, in any subpopulation $\mathbf{w}_1$ and on average.

4.2 Identification of Treatment Effects Dependent on Time-Varying Confounders

When histories $\bar{d}_t$ and $\bar{d}^*_t$ do not agree through period $t-1$, identifying the various effects of marriage on wages depends on identifying both $\boldsymbol{\beta}_c$, $\tau_c(\bar{d}_t, \mathbf{w}_1)$, and the effects of marriage on time-varying regressors, necessitating additional assumptions. In general, identifying effects of treatments on time-varying confounders poses challenges, now in a multivariate context, similar to those addressed in this article, namely the necessity of adjusting for confounders, some of which may be time varying, and also for unobserved variables. Nevertheless, in some instances, reasonable assumptions can be used to identify selected effects of interest.

One possibility is to assume that period t marital status, $t = 1, \ldots, T$, is independent of future potential confounders $\{\mathbf{X}_j(\bar{d}_{j-1})\}_{j=t+1}^{T}$, given prior values, marital history, and a set of measured confounders. Also $E(\mathbf{X}_t(\bar{d}_{t-1}) \mid \mathbf{W} \in A)$ is then identified. But, in general, it seems unreasonable to assume that unobserved confounders of the relationship between marital status and potential wages are not also confounding the relationship between marital status and time-varying regressors, which are components of the potential response.

Another possibility is that time-varying confounders are not affected by treatment history: for all i and $\bar{d}_{t-1} \in M_{t-1}$, $\mathbf{X}_{it}(\bar{d}_{t-1}) = \mathbf{X}_{it}$. A more plausible assumption might be that time-varying confounders are affected only by the last $r-1$ marital states: $\mathbf{X}_t(\bar{d}_{t-1}) = \mathbf{X}_t(\bar{d}^*_{t-1})$ for all $\bar{d}_{t-1} = (\bar{d}_{t-r}, \bar{d}_{\{j:t-r<j\leq t-1\}})$ and $\bar{d}^*_{t-1} = (\bar{d}^*_{t-r}, \bar{d}^*_{\{j:t-r<j\leq t-1\}})$ with $\bar{d}_{\{j:t-r<j\leq t-1\}} = \bar{d}^*_{\{j:t-r<j\leq t-1\}}$. Similarly, if $\mathbf{X}_t(\bar{d}_{t-1})$ depends only on the number of periods married, $\mathbf{X}_t(\bar{d}_{t-1}) = \mathbf{X}_t(\bar{d}^*_{t-1})$ for all $\bar{d}_{t-1}$ and $\bar{d}^*_{t-1}$ in which treatment is received j times in the first $t-1$ periods, $j = 0, \ldots, t-1$. When $\bar{d}_{t-1}$ and $\bar{d}^*_{t-1}$ satisfy either of these conditions, the basic effects (12) and (13) of $\bar{d}_t$ versus $\bar{d}^*_t$ are identified if Theorem 1 or 2 holds.

As noted before, if either $d_t = 0$ or $d_t = 1$ can follow any feasible partial regime $\bar{d}_{t-1}$, the contemporaneous effect (16) does not depend on the time-varying regressors, reducing to (17). Otherwise, (16) may depend on these, as in the example in Section 5.

The effect of treatment in period 1 (15) also depends on the "potential" probabilities $p(\bar{d}_{\{j:2\leq j\leq t\}}(d_1) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t)$, $d_1 = 0, 1$. To identify (15) from fixed effects models, further assumptions are needed.

5. WHAT DO STUDIES OF THE WAGE PREMIUM ESTIMATE?

The results above are now used to reexamine empirical studies of the marital wage premium. In one of the first studies to use a fixed effects model with panel data, KN set out a framework that many researchers follow. Attention thus focuses mainly on this article, which is treated as representative. Interest centers on average effects of marital status for all men.

KN used data from the 1976, 1978, and 1980 NLS of Young Men on 1228 white males. As is common, men who obtained additional education during the study period were excluded. First, KN estimated a pooled cross-sectional regression with the natural logarithm of the hourly wage rate Y_t as the dependent variable, with time-invariant and time-varying regressors for labor market experience, educational level, union status, birth year, region, occupation and industry, an indicator variable for the presence of nonspousal dependents, and two indicators $D_{1t} = 1$ for respondents married with spouse present in period t, 0 otherwise, and $D_{2t} = 1$ for men separated or divorced in period t, 0 otherwise. At any given level of the independent variables, in each period, married-spouse-present men earn 11% more per hour than never-married men, while divorced or separated men make 9% more than never-married men.

The model above does not adjust for unobserved confounders. To adjust for time-invariant observed and unobserved confounders, for example, birth year, ability, psychological dispositions, preexisting mental and physical health, KN then estimated an FER-STE model without period effects:

$$Y_{it} = \boldsymbol{\beta}'\vec{\mathbf{X}}_{it} + \tau_1 D_{1it} + \tau_2 D_{2it} + U_i + \varepsilon_{it}. \qquad (24)$$

Now, for any given level of the covariates $\mathbf{X}_t$ and U, married-spouse-present men make 6% more than never-married men ($\hat{\tau}_1 = 0.06$), while divorced or separated men make 4% more than never-married men ($\hat{\tau}_2 = 0.04$).

KN then stated that (24) is misspecified, claiming that it forces the effect of marriage on wages to accrue immediately upon marriage, precluding accumulation with time married. Therefore, they add terms for duration in the married and divorced/separated states; now only about 10%–20% of the wage premium appears to be due to selection on unobserved variables. From this, they conclude that the bulk of the marital wage premium is due to the effect of marriage on wages.

There are several problems with these analyses. First, men who married before period 1 cannot be exposed to the never-married state in period 1, and it can be argued that these men should therefore be dropped from the analysis. If they are nevertheless included on the grounds that one might imagine the counterfactual state in which they were never married before period 1, this will almost certainly lead to violation of Theorem 1 or 2 if marriage affects human capital formation, as per the productivity hypothesis. Let $L_t = D_{1t} + 2D_{2t}$, with values 0, 1, 2, denote period t marital status. Under Theorem 1 or 2, $Y_1(\ell_1) \perp\!\!\!\perp L_1 \mid \mathbf{X}_1, U$ implies

$$\begin{aligned} E(Y_1(0) \mid \mathbf{X}_1 = \mathbf{x}_1, L_1 = 1, U = u) \\ = E(Y_1(0) \mid \mathbf{X}_1 = \mathbf{x}_1, L_1 = 0, U = u), \end{aligned} \qquad (25)$$

that is, married-spouse-present men ($L_1 = 1$) and never-married men ($L_1 = 0$) with values $\mathbf{x}_1$ and u would have the same average potential wage $Y_1(0)$. KN reported that men who are married, with spouse present in period 1 have been married for an average of 7.7 years. If the productivity hypothesis is correct, these men, had they not married, would not have upgraded their human capital to the same extent and would therefore have values of $\mathbf{X}_1$ and/or U different from $\mathbf{x}_1$ and/or u. Their wages, had they not married, should therefore be less than those of their never-married counterparts with values $\mathbf{x}_1$ and u, contradicting (25).

In addition, $\mathbf{X_t}$ and D_t are concurrent, and if $\mathbf{X_t}$ depends on period t marital status, neither Theorem 1 nor Theorem 2 will hold. Given the period length and variables used here, it seems plausible that $\mathbf{X_t}$ depends on prior, but not period t, marital status. Nevertheless, Akerlof (1998, p. 300), who also restricted his sample to men who did not accrue education over the study period, deliberately excluded occupation, industry, and union status as time-varying regressors, arguing that "upgrading of occupational status is one of the ways in which income would be increased with marriage." However, if these variables are also relevant for marital status in the next period, neither Theorem 1 nor Theorem 2 will hold when they are dropped from the analysis. The appropriate approach in this case would be to let $\mathbf{X_t}$ consist of these variables measured in the previous period.

Further, studies in the literature fail to include time-varying wages and mental and physical health variables as regressors. Yet surely, changes in these variables affect both subsequent marital status and wage rates. This implies that (19) does not hold. Thus, Theorem 1 cannot be used to justify equating parameters of models FER-SEE and FEC-SEI and Theorem 2 cannot be used to justify equating parameters of models FER-STE and FEC-STI.

Even had KN (and/or others) excluded respondents married before period 1, included a rich enough set of time-varying confounders to make Theorem 1 more plausible, and used an FER-SEE model, they inappropriately interpreted model coefficients, arguing that τ_1 represents the cumulative effect of marriage on wages and that this effect occurs immediately after marriage. Similar remarks apply to the interpretation of the coefficient τ_2.

First, when the analysis is restricted to the subpopulation of men who do not obtain additional education during the study period, τ_1 is not in general a causal parameter, even if Theorem 1 or 2 was applicable. Let R denote respondent's education. Using the period 1 responses, $\tau_1 = E(Y_1(1) \mid \underline{X}_1 = \underline{x}, U = u, D_1 = 1, R_1 = R_2(1) = R_3(1, D_2)) - E(Y_1(0) \mid \underline{X}_1 = \underline{x},\ U = u,\ D_1 = 0, R_1 = R_2(0) = R_3(0, D_2))$. Thus, even if potential wages do not depend on D_1, unless marital status, contrary to the productivity hypothesis, does not affect education, in which case $R_1 = R_2(0) = R_3(0, D_2) = R_2(1) = R_3(1, D_2)$, τ_1 compares men in two different subpopulations. Similar remarks apply to τ_2.

To understand what the τ coefficients in an FER-SEE or FER-STE regression model would represent if the men who obtained additional schooling during the study period had been included and Theorem 1 or 2 was applicable, consider, for example, the FEC-STI causal model analogous to the FER-STE regression model:

$$Y_{it}(\bar{\ell}_t) = \boldsymbol{\beta}_c' \vec{\mathbf{X}}_{it}(\bar{\ell}_{t-1}) + \tau_{1c} d_{1t} + \tau_{2c} d_{2t} + U_i + \varepsilon_{it}(\bar{\ell}_t), \quad (26)$$

with $E(\varepsilon_{it}(\bar{\ell}_t) \mid \{\mathbf{X}_{i,j}(\bar{\ell}_{j-1})\}_{j=1}^{T}, \mathbf{W}_{1i}, \bar{V}_T, U_i) = 0$.

If Theorem 2 is applied, $\tau_1 = \tau_{1c} = E(Y_1(1) - Y_1(0)) = E(Y_2(0, 1) - Y_2(0, 0)) = E(Y_3(0, 0, 1) - Y_3(0, 0, 0))$ is the basic effect of transitioning from the never-married status to the married-spouse-present state in any period. It is also a conditional intermediate effect of marriage. And it is also the contemporaneous effect of being married versus never married in period 1.

However, τ_{1c} is not the contemporaneous effect of marriage on wages in periods 2 and 3. Once a respondent is married or divorced, he cannot become never married; thus, $d_3 = 0$ implies that $d_1 = d_2 = 0$, and $d_2 = 0$ implies that $d_1 = 0$. Consequently, the contemporaneous effect $E(Y_2(D_2 = 1) - Y_2(D_2 = 0))$ of being married spouse present versus never married in period 2 is

$$\tau_{1c} + \boldsymbol{\beta}_c' E(\vec{\mathbf{X}}_2(1) - \vec{\mathbf{X}}_2(0) \mid L_1 = 1) \Pr(L_1 = 1) + \boldsymbol{\beta}_c' E(\vec{\mathbf{X}}_2(2) - \vec{\mathbf{X}}_2(0) \mid L_1 = 2) \Pr(L_1 = 2). \quad (27)$$

Whether this is larger or smaller than τ_{1c} depends on the effects of marital status in period 1 on the time-varying regressors in period 2 and the marital status distribution in period 1. Similarly, τ_{2c} is not the contemporaneous effect on wages of being separated or divorced versus never married in either period 2 or period 3.

As noted earlier, KN interpreted the coefficients in (24) as cumulative effects that accrue immediately upon marriage. As this is not substantively plausible, they estimated additional models with marital duration variables, interpreting the duration coefficients as effects. But the interpretation mentioned earlier (and that for the additional models) ignores marriage effects on the time-varying confounders. For example, under (26), respondents marrying in period 2 receive $[E(Y_2(0, 1) - Y_2(0, 0)) = \tau_{1c}]\%$ more per hour than their never-married counterparts, while respondents who marry in period 1 (and stay married) receive $[E(Y_2(1, 1) - Y_2(0, 0)) = \tau_{1c} + \boldsymbol{\beta}_c' E(\vec{\mathbf{X}}_2(1) - \vec{\mathbf{X}}_2(0))]\%$ more than their never-married counterparts.

The brief examination here indicates that researchers using panel data to study effects of marriage on men's wages have included respondents in the analysis who should be excluded, omitted important confounders, and inappropriately conditioned on time-varying confounders such as education by excluding respondents who obtained additional schooling during the study period. Nor have researchers defined the effects they are attempting to estimate or realized that there are different effects of interest. These problems have led to misinterpretation of results and can lead to inappropriate policy recommendations. For example, in contrast to KN, researchers who found insignificant marital status coefficients argued that there is no effect of marriage on wages. But this conclusion, was Theorem 1 (Theorem 2) applicable and an FER-SEE (FER-STE) model used, would only be true for conditional intermediate effects (or the immediate effect of transitioning into marriage) and the literature does not suggest that it is these effects that are of primary interest. Perhaps these studies would have arrived at different conclusions had researchers defined the effects of interest clearly and appropriately estimated them.

It seems reasonable to conclude that little is known about the various possible effects of marriage on men's wages, much less how these come about. Further, as existing panel datasets with good wage data do not contain adequate information on important time-varying confounders such as substance abuse, depression, and physical disability, new data need to be collected before any effects of marriage on wages can be reliably estimated and any reasonable policies based on these can be formulated. In addition, to estimate basic effects that depend on the effects of marital history on time-varying confounders, new methods will need to be developed.

6. DISCUSSION

Fixed effects regression models for panel data are often used to adjust for selection on unobservables, and a number of research literatures have relied heavily on findings from these models. However, researchers using these models have not defined the effects they are attempting to estimate or understood the types of conditions that must be met in order to estimate treatment effects using these models. This can lead to incorrect conclusions and misguided policy recommendations.

To clarify these issues, a basis set of potential outcomes is defined and used to construct treatment effects of interest.

Theorems 1 and 2 give conditions under which various effects comparing regimes $\bar{d}_t = (\bar{d}_{t-1}, 1)$ and $d_t^* = (\bar{d}_{t-1}^*, 0)$, with $\bar{d}_{t-1} = \bar{d}_{t-1}^*$, are identified. If interest focuses on such effects, fixed effects regression models can be very useful.

However, identification of the basic effects (12) for regimes $\bar{d}_t = (\bar{d}_{t-1}, d_t)$ and $\bar{d}_t^* = (\bar{d}_{t-1}^*, d_t^*)$, with $\bar{d}_{t-1} \neq \bar{d}_{t-1}^*$, also depends in general on the effects of treatments on time-varying regressors, which may be difficult to estimate. When effects for such regimes are of primary interest, researchers should consider using other methods, for example, marginal structural models (Robins 1998; Hernán, Brumback, and Robins 2001; Robins and Hernán 2009). These models express the means of the potential outcomes $E(Y_t(\bar{d}_t))$ (or $E(Y_t(\bar{d}_t) \mid W \in A)$) as a function of treatment regime $\bar{d}_t$. Under an additional positivity condition, when treatment is sequentially ignorable, given a set of observed variables, these means are identified and can be estimated using IPTW. With an unobserved confounder U in addition to the observed confounders, a random effects probit model, for example, might be used to estimate the weights (Gibbons et al. 1994); of course, this will necessitate making additional assumptions (Robins, Greenland, and Hu 1999), for example, normality of the U_i.

To apply to the majority of applied work, this article focused on fixed effects models for continuous dependent variables with the form (1); the results are also readily extended to random effects models with this form. Theorems 1 and 2 give general conditions under which conditional expectations for the observed dependent variable and potential outcomes in the basis set are equal (Equations 20 and 22). Thus, identification of treatment effects for nonlinear panel data models and other types of dependent variables, for example, binary variables and counts, will require similar types of conditions.

Fixed effects models are also used with clustered data, for example, siblings or twins in families and individuals in neighborhoods. Although issues involving time-varying confounders do not arise, other problems, such as inappropriate exclusion of respondents, as in the study by Antonovics and Town (2004) previously mentioned, often arise in applications. If the SUTVA assumption is made, sufficient conditions for identifying treatment effects are more straightforward than those in panel studies. However, SUTVA is more likely to be violated. For example, suppose that older sibling A obtains more education if he marries, and suppose that his education affects that of younger sibling B, hence, presumably B's wages; thus, A's marital status affects B's wages. In this case (e.g., see Sobel 2006), typical estimates of treatment effects can be very misleading.

APPENDIX

Proof of Theorem 1. For $t = 1, \ldots, T$ and $s \leq t$, (19) implies

$$Y_t(\bar{d}_t) \perp\!\!\!\perp D_s \mid \{\mathbf{X}_m(d_{m-1})\}_{m=1}^{t}, \bar{D}_{s-1} = \bar{d}_{s-1}, \mathbf{W}_1, \bar{V}_t, U. \tag{A.1}$$

For $\bar{D}_t = \bar{d}_t$, $Y_t = Y_t(\bar{d}_t)$, $\bar{X}_t = \{\mathbf{X}_m(\bar{d}_{m-1})\}_{m=1}^{t}$, hence

$$E(Y_t \mid \bar{\mathbf{X}}_t, \bar{D}_t = \bar{d}_t, \mathbf{W}_1, \bar{V}_t, U) = E(Y_t(\bar{d}_t) \mid \{\mathbf{X}_m(\bar{d}_{m-1}\}_{m=1}^{t}, \bar{D}_t = \bar{d}_t, \mathbf{W}_1, \bar{V}_t, U), \tag{A.2}$$

and repeated application of (A.1) then gives

$$E(Y_t(\bar{d}_t) \mid \{\mathbf{X}_m(\bar{d}_{m-1}\}_{m=1}^{t}, \bar{D}_t = \bar{d}_t, \mathbf{W}_1, \bar{V}_t, U) = E(Y_t(\bar{d}_t) \mid \{\mathbf{X}_m(\bar{d}_{m-1}\}_{m=1}^{t}, \mathbf{W}_1, \bar{V}_t, U), \tag{A.3}$$

establishing the result. Under model FEC-SEI, (A.2) implies $E(\varepsilon_t(\bar{d}_t) \mid \mathbf{W} \in A, \bar{D}_{k(t)} \in B_t) = 0$ for $\bar{D}_{k(t)} = \bar{D}_{t-1}$, $B_t = \{\bar{d}_{t-1}\}$, $A = \{\mathbf{w}_1 : \mathbf{w}_1 \in C, C \subseteq \mathrm{Im}(\mathbf{W}_1)\}$.

Proof of Theorem 2. For $\bar{D}_T = \bar{d}_T$, $Y_t = Y_t(\bar{d}_t)$, $\bar{X}_T = \{\mathbf{X}_m(\bar{d}_{m-1})\}_{m=1}^{T}$, combining this with (21) gives

$$E(Y_t \mid \bar{\mathbf{X}}_T, \bar{D}_T = \bar{d}_T, \mathbf{W}_1, \bar{V}_T, U) = E(Y_t(\bar{d}_t) \mid \{\mathbf{X}_m(\bar{d}_{m-1}\}_{m=1}^{T}, \bar{D}_t = \bar{d}_t, \mathbf{W}_1, \bar{V}_T, U). \tag{A.4}$$

For $t = 1, \ldots, T$, $s \leq t$, (19) implies

$$Y_t(\bar{d}_t) \perp\!\!\!\perp D_s \mid \{\mathbf{X}_m(d_{m-1})\}_{m=1}^{T}, \bar{D}_{s-1} = \bar{d}_{s-1}, \mathbf{W}_1, \bar{V}_T, U, \tag{A.5}$$

and the result follows upon repeated application of (A.5). Under model FEC-STI, (A.2) implies $E(\varepsilon_t(\bar{d}_t) \mid \mathbf{W}, \bar{D}_{k(t)} \in B_t) = 0$ for $\bar{D}_{k(t)} = \bar{D}_{t-1}$, $B_t = \{\bar{d}_{t-1}\}$, $A = \{\mathbf{w}_1 : \mathbf{w}_1 \in C, C \subseteq \mathrm{Im}(\mathbf{W}_1)\}$.

[Received September 2010. Revised September 2011.]

REFERENCES

Abbring, J. H., and van den Berg, G. J. (2003), "The Nonparametric Identification of Treatment Effects in Duration Models," *Econometrica*, 71, 1491–1517. [523]

Akerlof, G. A. (1998), "Men Without Children," *The Economic Journal*, 108, 287–309. [521,526]

Almirall, D., Ten Have, T., and Murphy, S. A. (2010), "Structural Nested Mean Models for Assessing Time-Varying Effect Moderation," *Biometrics*, 66, 131–139. [523,524]

Antonovics, K., and Town, R. (2004), "Are All the Good Men Married? Uncovering the Sources of the Marital Wage Premium," *American Economic Review*, 94, 317–321. [521,528]

Arellano, M., and Honoré, B. (2001), "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics* (Vol. 5), eds. J. J. Heckman and E. E. Leamer, Amsterdam: Elsevier North-Holland, pp. 3231–3296. [523]

Avner, A., and Lerman, R. (2007), "How Do Marital Status, Work Effort, and Wage Rates Interact?," *Demography*, 44, 623–647. [522]

Becker, G. S. (1981), *A Treatise on the Family*, Cambridge, MA: Harvard University Press. [521]

Beegle, K., Dehejia, R. H., and Gatti, R. (2006), "Child Labor and Agricultural Shocks," *Journal of Development Economics*, 81, 80–96. [522]

Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5–46. [522]

——— (1992), "Comment: Sequential Moment Restrictions in Panel Data," *Journal of Business and Economic Statistics*, 10, 20–26. [522]

Chun, H., and Lee, I. (2001), "Why Do Married Men Earn More: Productivity or Marriage Selection?," *Economic Inquiry*, 39, 307–319. [521]

Cohen, Y., and Haberfeld, Y. (1991), "Why Do Married Men Earn More Than Unmarried Men?," *Social Science Research*, 20, 29–44. [521]

Cornwell, C., and Rupert, P. (1997), "Unobservable Individual Effects, Marriage and the Earnings of Young Men," *Economic Inquiry*, 35, 285–294. [521]

Frangakis, C., and Rubin, D. B. (2002), "Principal Stratification in Causal Inference," *Biometrics*, 58, 21–29. [523]

Gibbons, R. D., Hedeker, D., Charles, S. C., and Frisch, P. (1994), "A Random-Effects Probit Model for Predicting Medical Malpractice Claims," *Journal of the American Statistical Association*, 89, 760–767. [528]

Gray, J. S. (1997), "The Fall in Men's Returns to Marriage: Declining Productivity Effects or Changing Selection," *Journal of Human Resources*, 32, 481–504. [521]

Hernán, M. A., Brumback, B., and Robins, J. M. (2001), "Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments," *Journal of the American Statistical Association*, 96, 440–448. [523,528]

Hill, M. (1979), "The Wage Effects of Marital Status and Children," *Journal of Human Resources*, 14, 579–594. [521]

Hsiao, C. (2003), *Analysis of Panel Data* (2nd ed.), Cambridge: Cambridge University Press. [523]

Isaacson, G. (2007), "Twin Data vs. Longitudinal Data to Control for Unobserved Variables in Earnings Functions—Which Are the Differences?," *Oxford Bulletin of Economics and Statistics*, 69, 339–362. [521]

Korenman, S., and Neumark, D. (1991), "Does Marriage Really Make Men More Productive?," *Journal of Human Resources*, 26, 282–307. [521]

Krashinsky, H. (2004), "Do Marital Status and Computer Usage Really Change the Wage Structure?," *Journal of Human Resources*, 39, 774–791. [521]

Lai, T. L., Small, D. S., and Liu, J. (2008), "Statistical Inference in Dynamic Panel Data," *Journal of Statistical Planning and Inference*, 138, 2763–2776. [523]

Lechner, M. (2009), "Sequential Causal Models for the Evaluation of Labor Market Programs," *Journal of Business and Economic Statistics*, 27, 71–83. [523,525]

Lechner, M., and Miquel, R. (2005), "Identification of the Effects of Dynamic Treatments by Sequential Conditional Independence Statements," Discussion Paper 2005-17, Universität St. Gallen, Switzerland. [523]

Loh, E. S. (1996), "Productivity Differences and the Marriage Wage Premium for White Males," *Journal of Human Resources*, 31, 556–589. [521]

Mincy, R., Hill, J., and Sinkewicz, M. (2009), "Marriage: Cause or Mere Indicator of Future Earnings Growth?," *Journal of Policy Analysis and Management*, 28, 417–439. [521,522]

Neugebauer, R., and van der Laan, M. J. (2006), "Causal Effects in Longitudinal Studies: Definition and Maximum Likelihood Estimation," *Computational Statistics and Data Analysis*, 51, 1664–1675. [523]

Neumayer, E. (2003), "Good Policy Can Lower Violent Crimes: Evidence From Fixed Effects Estimation in a Cross-National Panel of Homicide Rates," *Journal of Peace Research*, 40, 619–640. [522]

Reed, W., and Harford, K. (1989), "The Marriage Premium and Compensating Wage Differentials," *Journal of Population Economics*, 2, 237–265. [521]

Robins, J. M. (1998), "Marginal Structural Models," in *1997 Proceedings of the American Statistical Association on Bayesian Statistical Science*, pp. 1–10. [525,528]

Robins, J. M., Greenland, S., and Hu, F. C. (1999), "Estimation of the Causal Effect of a Time-Varying Exposure on the Marginal Mean of a Repeated Binary Outcome" (with discussion), *Journal of the American Statistical Association*, 94, 687–712. [528]

Robins, J. M., and Hernán, M. (2009), "Estimation of the Causal Effects of Time-Varying Exposures," in *Longitudinal Data Analysis*, eds. G. Fitzmaurice, M. Davidian, G. Verbeke, G. Molenberghs, Boca Raton, FL: Chapman and Hall, pp. 553–597. [523,524,525,528]

Rodgers, W. M., III, and Stratton, L. S. (2010), "Male Marital Wage Differentials: Training, Personal Characteristics, and Fixed Effects," *Economic Inquiry*, 48, 722–742. [521]

Rosenbaum, P. R. (1984), "The Consequences of Adjustment for a Concomitant Variable That Has Been Affected by the Treatment," *Journal of the Royal Statistical Society*, Series A, 147, 656–666. [521]

Rosenbaum, P. R., and Rubin, D. B. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55. [524]

Rubin, D. B. (1980), Discussion of "Randomization Analysis of Experimental Data in the Fisher Randomization Test," by D. Basu, *Journal of the American Statistical Association*, 75, 591–593. [523]

Schoeni, R. F. (1995), "Marital Status and Earnings in Developed Countries," *Journal of Population Economics*, 8, 351–359. [521]

Sobel, M. E. (2006), "What Do Randomized Studies of Housing Mobility Demonstrate? Causal Inference in the Face of Interference," *Journal of the American Statistical Association*, 101, 1398–1406. [528]