# Identification of Marginal Effects in a Nonparametric Correlated Random Effects Model*

C. Alan Bester and Christian Hansen

First Version: August 2006  Revised: October 2007

**Abstract.** In this paper, we consider identification and estimation of average marginal effects in a correlated random effects model without imposing strong functional form assumptions on the structural likelihood or the mixing distribution. Identification is achieved through imposing that the mixing distribution depends on observed covariates only through an index function and the assumption that at least three time periods are available for each cross sectional unit. We leave the functional form of the index function unrestricted subject to smoothness conditions. We present identification results for this model and consider estimation of the marginal effects of interest. We illustrate the use of the approach through a brief empirical example which considers the relationship between insider trading activity and trading volume.

*Keywords: index, sufficient statistic, insider trading*

## 1. Introduction

The use of panel data is extremely common in empirical economics. Panel data is appealing as it allows researchers to deal flexibly with time-invariant individual specific effects that may not be independent of covariates of interest. In a linear model where unobserved individual specific heterogeneity enters as an additively separable term, the individual specific heterogeneity may flexibly be accommodated by allowing for individual specific intercepts that are treated as parameters to be estimated. This approach is widespread in economics and is appealing since one may estimate the common slope parameters of the model without imposing any structure on the individual specific effects.

Unfortunately, this approach is not readily generalizable to nonlinear, nonseparable, or dynamic models. In these more general settings, leaving the individual heterogeneity unrestricted usually

* Address correspondence to C. Hansen, Asst. Prof. of Econometrics and Statistics, The University of Chicago, Graduate School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA, chansen1@chicagogsb.edu.

results in inconsistent estimators of the common parameters due to the incidental parameters problem. That is, noise in the estimation of the individual specific effects when the time dimension is short results in inconsistent estimates of the common parameters due to the nonlinearity of the problem. See, for example, Neyman and Scott (1948) for an early discussion or Arellano and Honoré (2001).

There is a large literature in econometrics that offers approaches to identification and estimation in panel data models with individual specific heterogeneity. These approaches may broadly be classified in two categories. Both approaches assume a model for the outcome of interest that is common across individuals up to an unobserved, individual specific effect that takes values in a finite dimensional space. We refer to this as the structural model, and when parametric, to its parameters as the structural or common parameters. Fixed effect approaches treat individual specific effects as parameters to be estimated, leaving the distribution of unobserved heterogeneity relatively unrestricted at the cost of introducing a large number of nuisance parameters. Random effects approaches, on the other hand, typically assume that the distributions of individual specific effects belongs to a known parametric family indexed by a finite dimensional parameter.

Both fixed and random effects approaches have drawbacks. Fixed effects approaches will generally be subject to the incidental parameters problem. For some models, a clever transformation exists which removes the unobserved heterogeneity and allows consistent estimation of common parameters; see Wooldridge (2002) for several well-known examples in parametric models. The drawback of these approaches is that they are very model specific and rely heavily on knowledge of functional form. Recently, there have also been a number of fixed effects approaches developed which estimate the individual specific effects and then attempt to improve the performance of the resulting estimators of common parameters via bias-correction; see, for example, Hahn and Kuersteiner (2002), Hahn and Kuersteiner (2004) and Hahn and Newey (2004), and Woutersen (2005). These approaches seem quite promising, but rely on parametric assumptions for the structural model and asymptotics where the cross-sectional and time dimension both go to infinity.

Random effects approaches, on the other hand, generally bypass the incidental parameters problem, either by assuming that unobserved individual specific characteristics are independent of observed covariates, or by assuming that unobserved effects are drawn from a known parametric distribution defined by a finite dimensional parameter. For a parametric structural model, estimation may then proceed simply using any appropriate parametric estimator. However, the resulting estimates may be heavily influenced by independence assumptions or by functional form assumptions about both the structural model and the distribution of unobserved effects.

In this paper, we present an approach to estimation and inference that complements the aforementioned approaches. In many economic applications, the objects of interest are changes in the conditional expectation of the outcome of interest or other functionals of the structural model with a change in an observed covariate, holding the distribution of unobserved heterogeneity fixed. We consider identification and estimation of these marginal effects without parametric assumptions on the structural model or the distribution of unobserved effects. Without further restrictions, marginal effects will generally not be identified. This is because, with observational data and a model in which unobserved effects are correlated with observed covariates, a change in a covariate impacts the distribution of the outcome of interest through both the structural likelihood and distribution of heterogeneity.

We focus on identification of the marginal effect of a change in a particular observed covariate on a conditional expectation of interest holding unobserved heterogeneity fixed averaged against the conditional distribution of unobserved heterogeneity. Identification is achieved through three restrictions. First, we impose that the distribution of unobserved effects depends on the particular observed covariate only through an index function, which may be viewed as a sufficient statistic. We also assume that, once individual specific heterogeneity is conditioned upon, the structural model depends only on the contemporaneous value of the covariate. Finally, we assume that at least three time periods are available for each cross sectional unit. In the appendix, we consider a dynamic model and show that the second assumption may be relaxed when more than three time periods are available.

Under the first two assumptions, the three time periods allow us to use variation in conditional expectations in time periods other than the period of interest to estimate how the distribution of unobserved effects changes as the covariates change. Using this information, we are then able to construct an estimate of the effect of interest with the distribution of unobserved heterogeneity held fixed. With the exception of smoothness conditions and the two functional restrictions mentioned above, our identification results are obtained without any functional form assumptions regarding the index function, the distribution of unobserved effects conditional on the index, or the structural model. The structural model is assumed common across individuals but not necessarily across time periods. Thus, our approach identifies an interesting economic effect within a very general model.

Our approach is a random effects type approach in that we achieve identification with a fixed time dimension by imposing structure on the distribution of unobserved individual level effects. However, the approach may be regarded as intermediate to conventional random effects approaches and fixed effects approaches in that the restrictions we impose are nonparametric in nature. We also illustrate how the approach leads to a set of overidentifying restrictions that can be used to test the validity of the identifying restrictions.

The next section provides a brief overview of the model we consider and discusses related research. We develop the identification result in Section 3 and briefly discuss estimation in Section 4. Section 5 contains an empirical example in which we consider the effect of market volume on trading decisions of corporate insiders, and Section 6 concludes.

## 2. **Model and Parameter of Interest**

We consider panel models with time invariant individual specific effects. Let the observed data be $\{(y_i, x_i)\}_{i=1}^N$ where $y_i \in \mathcal{Y} \subseteq \mathbb{R}^T$ is a vector of outcomes for individual $i$ and $x_i = [x_{i1}, ..., x_{id_x}] \in \mathcal{X} \subseteq \mathbb{R}^{T \times d_x}$ is a $T \times d_x$ matrix of explanatory variables with $s^{th}$ column a $T \times 1$ vector $x_{is}$. Letting the time invariant individual specific heterogeneity be represented for each $i$ by $\alpha_i \in \mathcal{A}$, we suppose that the unknown distribution function of $y_{it}$ given $x_i$ and $\alpha_i$ can be represented as

$$y_{it}|x_i, \alpha_i \sim G_t(y_{it}|x_i, \alpha_i) = G_t(y_{it}|x_{it}, \alpha_i). \tag{2.1}$$

We refer to $G_t(y_{it}|x_{it}, \alpha_i)$ as the structural model. We note that this definition of the structural model allows for essentially arbitrary intertemporal heteroskedasticity but imposes that $x_{it}$ and $\alpha_i$ are sufficient to capture the effects of all past and future realizations of $x$ on the outcome $y$ at time $t$. We also suppose that individual specific effects are drawn from distribution function $Q(\alpha|x)$ that may in general depend on the full matrix of explanatory variables and note that without additional restrictions on $Q(\cdot)$ this imposes essentially no restrictions on the model.

We suppose that the researcher, if possible, would choose to estimate quantities of the form $\frac{\partial}{\partial x_{ist}}\mathrm{E}[m(y_{it})|x_i, \alpha_i]$; that is, the researcher would ideally like to know how certain expectations of the outcome variable, $y_{it}$, change due to changes in one of the corresponding conditioning variables, $x_{ist}$, with individual heterogeneity held constant. Of course, with $T$ finite, it is clear that such objects are generally not consistently estimable as they would require estimation of individual specific distribution functions. As a result, we suppose that a researcher would be satisfied with an estimate of the marginal effect of $x_{ist}$ on the expected value of $m(y_{it})$, with individual specific heterogeneity held constant, averaged over the distribution of individual specific effects:

$$b_{ts}(x_i) = \int_{\mathcal{A}} \left( \frac{\partial}{\partial x_{ist}} \mathrm{E}[m(y_{it})|x_i, \alpha_i] \right) dQ(\alpha_i|x_i). \tag{2.2}$$

Throughout the remainder of the paper, we define $b_{ts}(x_i)$ as the marginal effect of interest and consider a set of assumptions under which $b_{ts}(x_i)$ is identified and estimable.

Our approach begins by noting that, under fairly general conditions, expectations of the form $\mathrm{E}\left[m(y_{it})|x_i\right]$ and their derivatives are nonparametrically identified from the observed data. The derivative of this expectation with respect to a covariate $x_{ist}$, assuming it exists and under mild regularity conditions necessary to interchange the order of integration and differentiation, has the form

$$\frac{\partial}{\partial x_{ist}} \mathrm{E}\left[m(y_{it})|x_i\right] = b_{ts}(x_i) + \int_{\mathcal{A}} \mathrm{E}\left[m(y_{it})|x_i, \alpha_i\right] d\left( \frac{\partial}{\partial x_{ist}} Q(\alpha_i|x_i) \right) \tag{2.3}$$

In other words, when individual specific heterogeneity $\alpha_i$ is unobserved, changing the value of $x_{ist}$ affects the left hand side conditional expectation through two channels. The first is the quantity of interest, $b_{ts}(x_i)$, which represents the effect of a change in $x_{ist}$ under the structural model (2.1) averaged over the distribution of individual-specific heterogeneity for individuals of (observable) type $x_i$. The second term arises because changing the value of a covariate can change the distribution from which $\alpha_i$ is drawn. In general, additional restrictions on $G$, $Q$, and/or $m$ will be necessary to ensure the two right hand side terms are separately identified. For example, under the simplest random effects assumption, $Q(\alpha_i|x_i) \equiv Q(\alpha_i)$, marginal effects are trivially identified as the second right hand side term in (2.3) is zero.

In this paper, we will impose restrictions on the dependence of $Q(\alpha_i|x_i)$ on $x_i$ to achieve identification of $b_{ts}(x_i)$. In this sense, our approach is fundamentally a random effects approach, though the conditions we impose are nonparametric in nature and allow correlation between unobservables and observed covariates. In particular, our key identifying assumption will be that there exists an unknown sufficient statistic for each $s = 1, ..., d_x$ that captures the effect of $x_{is}$ on $\alpha_i$. This assumption implies that the distribution of unobserved heterogeneity may be written as

$$Q(\alpha_i|x_i) = Q(\alpha_i|h_1(x_{i1}), ..., h_{d_x}(x_{id_x})) \tag{2.4}$$

for unknown functions $h_s(x_{is})$ that map $\mathbb{R}^T$ to $\mathbb{R}$. In the following sections, we will show that this assumption with the additional smoothness conditions that the $h_s(\cdot)$ be differentiable in each of their entries and the measure over $\mathcal{A}$ induced by $Q(\alpha_i|h_1(x_{i1}), ..., h_{d_x}(x_{id_x}))$, $dQ(\alpha_i|h_1(x_{i1}), ..., h_{d_x}(x_{id_x}))$, be differentiable with respect to the $h_s(\cdot)$ will allow us to achieve identification of the marginal effect $b_{ts}(x_i)$ given in (2.2). Essentially, identification obtains because the index restriction allows us to use information from other time periods to obtain an estimate of the second right hand side term in (2.3). Note that if one is only interested in marginal effects associated with a single covariate $x_s$, $b_{ts}(x_i)$, the index and differentiability conditions need only hold with respect to $x_s$.

The index restriction itself is easily seen as defining a restricted set of types for the individual observations. For example, in the case of a single covariate, we have that any individuals, say $i$ and $j$, for which $h(x_i) = h(x_j)$ will have unobserved heterogeneity drawn from the same distribution. This restriction seems quite mild. Since we will not require that the index function be specified *a priori*, we are placing only mild restrictions on the set of $x$ values that will result in individuals being of the same type. By assuming differentiability of the $h_s(\cdot)$ with respect to their arguments, we are imposing that two populations of individuals with similar observable characteristics will have time invariant unobserved characteristics that are similarly distributed. Note that we do not assume individuals with the same $x$'s have the same unobservable characteristics, simply that these latter characteristics are drawn from a common distribution.

The approach we take has a number of drawbacks. As developed, the approach does not allow for estimation or identification of the effects of discrete covariates, though the differences in marginal effects of continuous variables across categories of a discrete variable may easily be obtained under our conditions. The approach may also need substantial modification in the case of unbalanced panels. The index restriction does restrict the distribution of heterogeneity, though it leaves the functional form of this distribution largely unspecified and definitely does not restrict this distribution to fall within a given parametric class of models.

A further drawback is that the object of interest, $b_t(x)$ as defined in (2.2), changes as the number of time periods, $T$, increases. One way to make this model coherent as $T$ increases is to imagine the model above as having been defined for a very large, but finite, number of time periods, $T^* > T$, and think of the density $Q$ in (2.2) as a marginal density with $x_{T+1}, \ldots, x_{T^*}$ integrated out. Still, it is apparent that this approach would require modification to remain coherent as a literal description of a data generating process if one wishes to seriously entertain the notion of $T$ going to infinity. We note that our model may serve as a useful building block for such approaches and that extensions in this direction seem like an interesting avenue for future research.

The benefit of our approach is that we achieve essentially nonparametric identification of an interesting effect. Beyond differentiability in $x$, we impose few restrictions on the structural model, allowing, for example, for general intertemporal heteroskedasticity. The index restriction also provides testable restrictions that we will discuss below.

### 2.1. **Related Methods**

Linear index restrictions have been employed in correlated random effects estimation of panel data models in economics since at least the early work of Mundlak (1978) and Chamberlain (1980). Linear index restrictions are also commonly employed in hierarchical modeling; see Raudenbush and Bryk (2002). For example, a common assumption in these models is $h(x_i) = \bar{x}_i$. Another restriction, considered by Wooldridge (2005) is that observable characteristics $x_i$ do not affect the mean of the unobserved efffect but do affect the variance, e.g., $\alpha_i \sim N(0, \exp\{x_i'\delta\})$. More recently, Chen and Khan (2007) and Gayle and Viauroux (2007) use index restrictions to obtain identification in semiparametric panel data models. These latter two papers are similar in spirit to ours in that both treat the functional form of the index as unknown. Our approach differs from these by allowing a fully nonparametric structural model and an unspecified mixing distribution that allows correlation between unobservables and observed covariates in any time period, in addition to leaving the functional form of the index unrestricted.

There are, of course, a variety of other interesting approaches one could pursue to identify and estimate marginal effects in the present context. When the distribution of $y_i$ given $x_i$ and $\alpha_i$ or a set of moment conditions relating $y_i$ to $x_i$ and $\alpha_i$ is specified up to a finite dimensional set of parameters $\theta$, one can use parametric random or fixed effects strategies. For the linear model, identification and estimation strategies are developed under very general conditions in Hausman and Taylor (1981). If one is willing to specify a parametric model for the distribution of $\alpha_i$ given $x_i$, one can pursue a conventional random effects approach to identify and estimate $\theta$ and the parameters of the distribution of individual heterogeneity.

Alternatively, one may leave the distribution of $\alpha_i$ given $x_i$ unrestricted and attempt to estimate $\theta$ and the $\alpha_i$ jointly. This approach is appealing in that it leaves the distribution of unobserved heterogeneity unrestricted but unfortunately is subject to the incidental parameters problem of Neyman and Scott (1948) in general. Recent approaches to addressing this problem make use of asymptotics where the number of cross-sectional observations, $N$, and $T$ go to infinity jointly and propose bias-reductions based on these asymptotics; see, for example, Hahn and Kuersteiner (2002), Hahn and Kuersteiner (2004), Hahn and Newey (2004), and Woutersen (2005). While these approaches are interesting and extremely useful in many situations, they do rely on parametric structure for the structural model, and it does seem to be useful to consider approaches that do not rely on $T \to \infty$.

In the context of models in which the structural model depends on only $\alpha_i$ and a finite dimensional parameter vector $\theta$, it is also possible to perform inference about the set of identified values for $\theta$ rather than require that $\theta$ be point identified. This approach is pursued in Honoré and Tamer (2006) and Chernozhukov, Hahn, and Newey (2004) and does not require that $T \to \infty$.

Another set of approaches relies on finding a transformation that removes the unobserved heterogeneity from the problem. These transformations are well-known for the linear model, Logit model, and Poisson model; see Wooldridge (2002). In addition, there are a variety of such approaches that apply to semiparametric panel data models under the assumption that the structural model depends on $x_{it}$ and $\alpha_i$ only through a linear index but is otherwise unknown. Examples include Manski (1987) in the static binary choice model, Honoré and Kyriazidou (2000b) in dynamic binary choice, Honoré and Kyriazidou (2000a) in censored models, Horowitz and Lee (2004) in duration models, and Kyriazidou (1997) for selection models. One drawback of this approach is that it relies heavily on the assumption that individual specific heterogeneity enters additively in a linear index with $x_{it}$, is very model specific, and only applies in a small number of cases. In addition, with the exception of the linear model, only the index coefficients on $x_{it}$ is identified; and since the $\alpha_i$ are eliminated from the problem and neither they nor their distributions are estimated, marginal effects are not identified or estimated. Our approach is similar to the semiparametric approaches listed above in that we do not place functional form restrictions on either the structural model or the distribution of unobserved heterogeneity beyond index restrictions. However, we place the index restriction in the distribution of unobserved heterogeneity rather than the structural model and allow for an unknown index rather than imposing a linear index.

Another approach is pursued in Honoré and Lewbel (2002) and Lewbel (2005). In these approaches the existence of a "special" regressor that is known to enter the structural model with non-zero coefficient and be excluded from the distribution of unobserved heterogeneity is assumed. Then, assuming that the structural model depends on $x_{it}$ and $\alpha_i$ only through a linear index, simple estimators emerge. Our approach is similar in that we also impose restrictions on the distribution of unobserved heterogeneity. The index restriction is neither more nor less general than the exclusion restriction assumed with the special regressor.

Nonparametric random effects estimators are proposed in Lin and Carroll (2000) and Ullah and Roy (1998), with general properties of these estimators established in Henderson and Ullah (2005). Note that all of these approaches require that unobservables are independent of observed covariates, and varying degrees of additive separability, neither or which is assumed in this paper. Index assumptions have recently been employed in semiparametric panel models. Chen and Khan (2007) use an index assumption to obtain identification of a censored panel regression model with time-varying factor loadings. We note that our approach would also accomodate time-varying loadings on $\alpha_i$ as the structural model is allowed to change with $t$. Gayle and Viauroux (2007) employ a different index assumption to obtain identification in a semiparametric dynamic panel sample selection model. Their assumption differs from ours in that it assumes that unobservables in the selection equation are a deterministic function of time-invariant strictly exogenous observables, but leaves unobservables in the outcome equation unrestricted. Both of these papers obtain $\sqrt{n}$-consistent estimators for a common parametric component. However, as semiparametric estimators, both rely on functional form assumptions as well as additive separability between observables and unobservables. Our approach relies only on the index assumption and the assumption that, conditional on $x_{it}$ and $\alpha_i$, $y_{it}$ is independent of $x_{is}$ for all $s \neq t$. We show in the appendix that the latter assumption may be relaxed, with the cost that more than three time periods may be required to identify the effects of interest.

Perhaps the approach that is most similar to the approach we pursue in this paper is that of Altonji and Matzkin (2005). They consider identification of average marginal effects as in (2.2). They also achieve identification in a similar manner by assuming that there is a vector $z$ that is sufficient for $x$ in the distribution of unobserved heterogeneity; that is, they assume $Q(\alpha_i|x_i, z_i) = Q(\alpha_i|z_i)$ for some known vector $z$. This is quite similar to our approach in that we also assume that there is a set of sufficient statistics for $x$ in the distribution of unobserved heterogeneity. Unlike Altonji and Matzkin (2005), we do not assume that this set of statistics is known. Rather, we assume that the set of statistics takes on the index form in (2.4). Again, this assumption is neither more nor less general than the assumption in Altonji and Matzkin (2005); their vector $z$ is known but may include interactions of covariates while our set of sufficient statistics is unknown but is restricted so that there is one sufficient statistic for each covariate.

## 3. **Identification of Marginal Effects**

In this section, we consider identification of marginal effects of continuous covariates in a short $T$ panel context. For brevity, we consider only identification in static models in the main text and offer extensions to dynamic models and expectations involving more than one time period in the appendix. For notational convenience, we also limit the discussion to the case of one explanatory variable. Given this, we drop the subscript $s$ used above. Under the index restriction (2.4), the generalization to more than one covariate is straightforward. To conserve on notation, we also drop the subscript $i$ indexing individuals.

We suppose that for our single continuous covariate, we are interested in estimating the marginal effects

$$b_t(x) = \int_{\mathcal{A}} \frac{\partial}{\partial x_t} \mathrm{E}[m(y_t)|x, \alpha] dQ(\alpha|x) \tag{3.1}$$

for some $t$. We note that $\mathrm{E}[m(y_t)|x]$ and its derivatives with respect to $x$ are identified from the data. Under the model developed in Section 2, we have

$$\mathrm{E}[m(y_t)|x] = \int_{\mathcal{A}} \int_{\mathcal{Y}_t} m(y_t) dG_t(y_t|x, \alpha) dQ(\alpha|h(x_1, ..., x_T)).$$

Thus, letting $\mathcal{D}^m_{t,\tau}(x) = \frac{\partial}{\partial x_\tau} \mathrm{E}[m(y_t)|x]$ we have

$$\begin{aligned}
\mathcal{D}^m_{t,t}(x) &= \int_{\mathcal{A}} \int_{\mathcal{Y}_t} m(y_t) \frac{\partial}{\partial x_t} dG_t(y_t|x, \alpha) dQ(\alpha|h(x_1, ..., x_T)) \\
&\qquad + \int_{\mathcal{A}} \int_{\mathcal{Y}_t} m(y_t) dG_t(y_t|x, \alpha) \frac{\partial}{\partial x_t} dQ(\alpha|h(x_1, ..., x_T)) \\
&= b_t(x) + \left( \frac{\partial}{\partial x_t} h(x_1, ..., x_T) \right) \int_{\mathcal{A}} \int_{\mathcal{Y}_t} m(y_t) dG_t(y_t|x, \alpha) \frac{\partial}{\partial h} dQ(\alpha|h(x_1, ..., x_T)) \\
&= b_t(x) + h_t(x) c_t(x) \qquad (3.2)
\end{aligned}$$

and, for $\tau \neq t$,

$$\begin{aligned}
\mathcal{D}^m_{t,\tau}(x) &= \left( \frac{\partial}{\partial x_\tau} h(x_1, ..., x_T) \right) \int_{\mathcal{A}} \int_{\mathcal{Y}_t} m(y_t) dG_t(y_t|x, \alpha) \frac{\partial}{\partial h} dQ(\alpha|h(x_1, ..., x_T)) \\
&= h_\tau(x) c_t(x) \qquad (3.3)
\end{aligned}$$

where $h_t(x) = \frac{\partial}{\partial x_t} h(x_1, ..., x_T)$ and $c_t(x) = \int_{\mathcal{A}} \int_{\mathcal{Y}_t} m(y_t) dG_t(y_t|x, \alpha) \frac{\partial}{\partial h} dQ(\alpha|h(x_1, ..., x_T))$. Like $\mathcal{D}^m_{t,\tau}$, $b_t$ and $c_t$ are real-valued functionals; their dependence on $m$ is surpressed for notational convenience.

As in (2.3), we see that the derivative with respect to $x_t$ of the conditional expectation of $m(y_t)$ given $x$ is equal to the marginal effect of interest, the derivative of the structural conditional expectation with respect to $x_t$ with the distribution of unobserved heterogeneity fixed integrated against the distribution of unobserved heterogeneity, plus an additional term that corresponds to the change in unobserved heterogeneity when $x_t$ changes. Under the assumption that only the current value of $x$ affects the current value of $y$ in the structural model $G$, we also have that the derivative of the conditional expectation of $m(y_t)$ with respect to $t \neq \tau$ only depends on how individual heterogeneity changes when $x_\tau$ changes. Due to the index restriction, the terms corresponding to changes in the distribution of individual heterogeneity consist of a common component multiplied by the change in the index due to changing the particular component of $x$. We will make use of these facts below with $T \geq 3$ to show identification of $b_t(x)$. Before considering $T \geq 3$, we briefly illustrate that with $T = 2$ we achieve identification if we are willing to assume the index function is known or satisfies certain functional restrictions.

### 3.1. $\mathbf{T = 2}$

With $T = 2$, we may use expressions (3.2) and (3.3) to define a system of four equations:

$$\mathcal{D}^m_{1,1}(x) = b_1(x) + h_1(x)c_1(x)$$
$$\mathcal{D}^m_{1,2}(x) = h_2(x)c_1(x)$$
$$\mathcal{D}^m_{2,1}(x) = h_1(x)c_2(x)$$
$$\mathcal{D}^m_{2,2}(x) = b_2(x) + h_2(x)c_2(x).$$

Inspection of this system of equations shows that there will not be a unique solution for $b_1(x)$ and $b_2(x)$ without additional restrictions.

One restriction that has been employed in practice is that the index function is known, possibly up to a finite dimensional parameter. For example, it may be assumed that $h(x) = x_1 + x_2$, or more generally that that $h(x) = x'\delta$; see Mundlak (1978) and Chamberlain (1980). With a known $h(x)$, we also know the derivatives $h_1(x)$ and $h_2(x)$. Given these we may obtain

$$b_1(x) = \frac{\partial}{\partial x_1}\mathrm{E}[m(y_1)|x] - \left(\frac{h_1(x)}{h_2(x)}\right)\frac{\partial}{\partial x_2}\mathrm{E}[m(y_1)|x]$$

and

$$b_2(x) = \frac{\partial}{\partial x_2}\mathrm{E}[m(y_2)|x] - \left(\frac{h_2(x)}{h_1(x)}\right)\frac{\partial}{\partial x_1}\mathrm{E}[m(y_2)|x].$$

We assume that both $h_1(x)$ and $h_2(x)$ are nonzero. If one of $h_1(x)$ and $h_2(x)$ is zero, then identification of both $b_1(x)$ and $b_2(x)$ will generally not be possible. If both $h_1(x)$ and $h_2(x)$ are known to be zero, identification of both effects of interest is immediate.

If one is confident in the specification of $h(x)$, then this offers a way to estimate the marginal effects of interest under quite weak restrictions on the distribution of individual specific heterogeneity and the structural function. In many cases this is not the only way to estimate $b_t$. Wooldridge

(2005) provides examples where average marginal effects are identified and can be obtained without differentiating $\mathrm{E}[y|x]$ with respect to $x$. However, Wooldridge (2005) also points out that, without additional assumptions, one cannot distinguish between omitted heterogeneity and misspecification in the structural model.

Alternatively, one may treat the index as unknown but assume it satisfies functional restrictions sufficient to determine the ratio of derivatives, $h_t/h_\tau$. For example, one could assume $h$ is symmetric, $h(x_1, x_2) = h(x_2, x_1)$, in which case we have $h_t/h_\tau \equiv 1$. In general, however, the results will depend sensitively on the chosen model for the index function. In the following, we show that with $T \geq 3$ we may identify the effects of interest while allowing for an unspecified index function.

### 3.2. $\mathbf{T \geq 3}$

With $T = 2$ and the index function $h$ known, the marginal effect $b_t(x)$ is identified because the effect of a change in the distribution of unobserved heterogeneity may be estimated using information from the other time period, $\tau \neq t$. This is done by the taking the derivative of the conditional expectation, $\frac{\partial}{\partial x_\tau}\mathrm{E}[m(y_t)|x]$, and multiplying by the ratio of derivatives of the index function, $h_t/h_\tau$, which is available because $h$ is assumed known. In this section, we show that when three (or more) time periods are available, marginal effects may be identified without assuming a known index function. This occurs because an additional time period $s \notin \{t, \tau\}$ may be used to form an estimate of the appropriate ratio of derivatives of $h$. We also note that, in addition to identifying marginal effects, our index assumption implies a set of functional restrictions that may be used in constructing a specification test.

Without loss of generality, suppose we are interested in the first period marginal effect, $b_1(x)$. Equations (3.2) and (3.3) define a system of $T^2$ equations, four of which are

$$
\begin{aligned}
\mathcal{D}^m_{1,1}(x) &= \frac{\partial}{\partial x_1}\mathrm{E}[m(y_1)|x] = b_1(x) + h_1(x)c_1(x) \\
\mathcal{D}^m_{1,2}(x) &= \frac{\partial}{\partial x_2}\mathrm{E}[m(y_1)|x] = h_2(x)c_1(x) \\
\mathcal{D}^m_{3,1}(x) &= \frac{\partial}{\partial x_1}\mathrm{E}[m(y_3)|x] = h_1(x)c_3(x) \\
\mathcal{D}^m_{3,2}(x) &= \frac{\partial}{\partial x_2}\mathrm{E}[m(y_3)|x] = h_2(x)c_3(x).
\end{aligned}
$$

Note that the left hand sides of all four equations are identified from observed $(y, x)$ outcomes. Similar to the $T = 2$ case, the derivative $\mathcal{D}_{1,2}$ is proportional to the unknown function $c_1$. The difference here is that, with an additional time period available, the ratio of derivatives $\frac{h_1}{h_2}$ may be estimated using information from the third period by taking the ratio of derivatives of the conditional expectation $\mathrm{E}[m(y_3)|x]$ with respect to $x_1$ and $x_2$. Provided that $\mathcal{D}^m_{3,2}(x) \neq 0$, the marginal effect of interest is identified and given by

$$b_1(x) = \mathcal{D}^m_{1,1}(x) - \left(\frac{\mathcal{D}^m_{3,1}(x)}{\mathcal{D}^m_{3,2}(x)}\right) \mathcal{D}^m_{1,2}(x).$$

Note that, if any $\mathcal{D}^m_{t,\tau}(x)$ for $t \neq \tau$ is zero, deciding which quantities are identified will generally require inspection of the full system of $T^2$ equations. Inspection of the system will generally suggest that either only a subset of the $T$ marginal effects of interest are identified or that the model is misspecified.

The index assumption also provides additional testable implications, even in the $T = 3$ case. To see one example, notice that the estimator of $b_1(x)$ above is not the only one that can be constructed using information from the first three time periods; we also have

$$b_1(x) = \mathcal{D}^m_{1,1}(x) - \left(\frac{\mathcal{D}^m_{2,1}(x)}{\mathcal{D}^m_{2,3}(x)}\right) \mathcal{D}^m_{1,3}(x)$$

when $\mathcal{D}^m_{2,3}(x) \neq 0$. The resulting estimator of $b_1(x)$ would be distinct from the one constructed above because it uses information from the second period to estimate the ratio of derivatives, $h_1/h_3$, and information from the third period to estimate the integral term $c_1(x) = \int_{\mathcal{A}} \mathrm{E}[m(y_1)|x]\, d\Big(\frac{\partial Q}{\partial h}\Big)$. A testable implication of our index assumption is that the two estimators should agree at any set of $x$ values where both are well-defined.

In practice, either or both estimators may be undefined when their denominators $\mathcal{D}^m_{2,3}$ or $\mathcal{D}^m_{3,2}$ are zero. However, inspection of (3.3) reveals that any function $\varphi(y)$ in the domain of the time-$s$ conditional expectation operator may be used to estimate the ratio of derivatives of the index function,

$$\frac{\mathcal{D}^{\varphi}_{s,t}(x)}{\mathcal{D}^{\varphi}_{s,\tau}(x)} = \frac{\frac{\partial}{\partial x_t}\mathrm{E}\left[\varphi(y_s)|x\right]}{\frac{\partial}{\partial x_\tau}\mathrm{E}\left[\varphi(y_s)|x\right]} = \frac{h_t \int_{\mathcal{A}} \int_{\mathcal{Y}} \varphi\, dG_s\, d\Big(\frac{\partial Q}{\partial h}\Big)}{h_\tau \int_{\mathcal{A}} \int_{\mathcal{Y}} \varphi\, dG_s\, d\Big(\frac{\partial Q}{\partial h}\Big)} = \frac{h_t}{h_\tau},$$

provided that $s \notin \{t, \tau\}$ and the mapping $x \mapsto \mathrm{E}[\varphi(y_s)|x]$ is differentiable in $x_t$ and $x_\tau$. When the support of $y$ is large (or continuous), considering other functions $\varphi(y)$ may offer considerable latitude in constructing the ratio of derivatives of the index functions. This also suggests that, for continuous $y$, identification will generally fail due to a zero denominator only when $h_\tau(x) = 0$.

The general $T \geq 3$ identification result for static models is summarized in the following proposition.

**Proposition 1.** *Let $(\Omega, \sigma(\Omega), \mathbb{P})$ be a probability space and $(\tilde{y}, \tilde{x}, \tilde{\alpha})$ a random variable taking values in $\mathcal{Y} \times \mathcal{X} \times \mathcal{A}$, where $\mathcal{Y} = \otimes_{t=1}^{T} \underline{\mathcal{Y}}$ with $\underline{\mathcal{Y}} \subseteq \mathbb{R}$, $\mathcal{X} \subseteq \mathbb{R}^T$, and $\mathcal{A} \subseteq \mathbb{R}^{d_\alpha}$. Let $t \in \{1, \ldots, T\}$, $x^* = (x_1^*, \ldots, x_T^*) \in \mathcal{X}$, and $m \in L^2(\underline{\mathcal{Y}})$ be given. Suppose that there exist*

(i) *A measurable function $G_t(y_t|x, \alpha)$ such that $\forall (y_t, x, \alpha) \in \underline{\mathcal{Y}} \times \mathcal{X} \times \mathcal{A}$, $\mathbb{P}\{\tilde{y}_t < y_t | \tilde{x} = x, \tilde{\alpha} = \alpha\} = G_t(y_t|x_t, \alpha)$, and this function is differentiable in $x_t$ at $x^*$,*

(ii) *A measurable function $Q(\alpha|x)$ such that $\forall (\alpha, x) \in \mathcal{A} \times \mathcal{X}$, $\mathbb{P}\{\tilde{\alpha} < \alpha | \tilde{x} = x\} = Q(\alpha|x)$, and this function is differentiable in $x$ at $x^*$,*

(iii) *$\tau \in \{1, \ldots, T\}$ such that $t \neq \tau$, and the operator-valued mapping $x \mapsto \mathcal{E}_t(x)$ defined by*

$$[\mathcal{E}_t(x)]\, m = E_{\mathbb{P}}\left[m(y_t)|\tilde{x} = x\right],$$

*is differentiable with respect to $x_t$ and $x_\tau$ at $x = x^*$, meaning that the function $f_t^m(x) = E_{\mathbb{P}}[m(y_t)|\tilde{x} = x]$ has well-defined partial derivatives $\left|\frac{\partial f_r^m}{\partial x_{r'}}(x)\right| < \infty$ for $r' \in \{t, \tau\}$ at $x = x^*$.*

(iv) *$s \in \{1, \ldots, T\}$ and $\varphi \in L^2(\underline{\mathcal{Y}})$, such that $t \neq \tau \neq s$, and that the function $f_s^\varphi(x) = E_{\mathbb{P}}[\varphi(y_s)|\tilde{x} = x]$ has well-defined partial derivatives $\left|\frac{\partial f_s^\varphi}{\partial x_{r'}}(x)\right| < \infty$ for $r' \in \{t, \tau\}$ at $x = x^*$, and $\frac{\partial f_s^\varphi}{\partial x_\tau}(x^*) \neq 0$. Note that $\varphi = m$ is permissible so long as $\frac{\partial f_s^m}{\partial x_\tau}(x^*) \neq 0$.*

*For a given $m \in L^2(\underline{\mathcal{Y}})$, denote by $\mathcal{D}_{r,r'}^m(x^*)$ the partial derivative of the mapping $x \mapsto [\mathcal{E}_r(x)]\, m$ with respect to $x_{r'}$ at $x = x^*$. Then the marginal effect $b_t(x^*)$ defined in (3.1) is identified and satisfies*

$$b_t(x^*) = \mathcal{D}_{t,t}^m(x^*) - \left(\frac{\mathcal{D}_{s,t}^m(x^*)}{\mathcal{D}_{s,\tau}^m(x^*)}\right) \mathcal{D}_{t,\tau}^m(x^*) \tag{3.4}$$

Conditions (i)-(iv) are quite mild and assume little beyond the existence of the object of interest. (i) and (ii) assume the existence of conditional distributions which are differentiable in their conditioning arguments. (iii) and (iv) ensure that the derivatives of the conditional expectations of interest exist.

The substantive restriction needed for identification beyond the index assumption is embedded in (iv) and assumes that a nonzero term can be found for the denominator in the expression (3.4). This assumption functions like the instrumental variables assumption that the instrument be correlated to the endogenous variable in the model and is akin to the existence of a first-stage relationship. The assumption of a nonzero denominator term makes this an essentially pointwise identification argument. In particular, unless one is willing to assume that the denominator is bounded away from zero almost everywhere or another substantive functional restriction, it seems that this condition should be checked at each point of interest.

### 3.3. **Overidentifying Restrictions**

In addition to providing sufficient structure for identification of average marginal effects, the index structure in the distribution of unobserved effects also provides a set of overidentifying restrictions that can be used for specification testing. In this section, we briefly illustrate these overidentifying restrictions. Since the index assumption is a substantial functional restriction, we believe that having a simple set of testable implications should be useful in many situations. We also note that, while the index assumption will typically generate overidentification in static models, just identification is possible in the dynamic case, as shown in the appendix.

For a value of $x$ and a $t$ of interest, there will generally be $(T-1)(T-2)$ ways to construct $b_t(x)$. To see this, note that $b_t(x) = \mathcal{D}^m_{t,t}(x) - \left(\frac{\mathcal{D}^m_{s,t}(x)}{\mathcal{D}^m_{s,\tau}(x)}\right)\mathcal{D}^m_{t,\tau}(x)$ for any $\{s, \tau\}$ with $s \neq \tau$, $s \neq t$, and $\tau \neq t$. If one fixes $s$, there are obviously $(T-2)$ values of $\tau$ that satisfy $\tau \neq s$ and $\tau \neq t$; and for any $t$, there are $(T-1)$ other potential values for $s$. For any $b_t(x)$, we then have $(T-1)(T-2)-1 = T^2-3T+1$ overidentifying restrictions since these $(T-1)(T-2)$ ways to construct $b_t(x)$ should all agree.

It is important to note that the total number of overidentifying restrictions provided by any particular $x$ variable is always $T^2-3T+1$ despite the fact that we have, in principle, $(T-1)(T-2)$ available estimators for each component of the vector $b(x) = (b_1(x), ..., b_T(x))'$. This results because once we have considered all possible estimators for a particular $t$, all other restrictions for other $t$ are functionally dependent on the set of restrictions used when considering time $t$. This dependence is easiest to see in the $T = 3$ case. With $T = 3$, there are two ways to identify each possible $b_t(x)$ which should produce the same answer. In particular, we may define $b_{11}(x) = \mathcal{D}_{1,1}(x) - \left(\frac{\mathcal{D}_{2,1}(x)}{\mathcal{D}_{2,3}(x)}\right)\mathcal{D}_{1,3}(x)$,

$b_{12}(x) = \mathcal{D}_{1,1}(x) - \left(\frac{\mathcal{D}_{3,1}(x)}{\mathcal{D}_{3,2}(x)}\right)\mathcal{D}_{1,2}(x)$, $b_{21}(x) = \mathcal{D}_{2,2}(x) - \left(\frac{\mathcal{D}_{1,2}(x)}{\mathcal{D}_{1,3}(x)}\right)\mathcal{D}_{2,3}(x)$, $b_{22}(x) = \mathcal{D}_{2,2}(x) - \left(\frac{\mathcal{D}_{3,2}(x)}{\mathcal{D}_{3,1}(x)}\right)\mathcal{D}_{2,1}(x)$, $b_{31}(x) = \mathcal{D}_{3,3}(x) - \left(\frac{\mathcal{D}_{1,3}(x)}{\mathcal{D}_{1,2}(x)}\right)\mathcal{D}_{3,2}(x)$, and $b_{32}(x) = \mathcal{D}_{3,3}(x) - \left(\frac{\mathcal{D}_{2,3}(x)}{\mathcal{D}_{2,1}(x)}\right)\mathcal{D}_{3,1}(x)$. The model then implies that $b_{11}(x) = b_{12}(x)$, $b_{21}(x) = b_{22}(x)$, and $b_{31}(x) = b_{32}(x)$. To see that these statements are functionally dependent, notice that, for example, $b_{11}(x) = b_{12}(x) \Leftrightarrow \left(\frac{\mathcal{D}_{2,1}(x)}{\mathcal{D}_{2,3}(x)}\right)\mathcal{D}_{1,3}(x) = \left(\frac{\mathcal{D}_{3,1}(x)}{\mathcal{D}_{3,2}(x)}\right)\mathcal{D}_{1,2}(x) \Leftrightarrow \mathcal{D}_{1,3}(x)\mathcal{D}_{2,1}(x)\mathcal{D}_{3,2}(x) = \mathcal{D}_{1,2}(x)\mathcal{D}_{2,3}(x)\mathcal{D}_{3,1}(x)$. Similarly, we also have that $b_{21}(x) = b_{22}(x) \Leftrightarrow \mathcal{D}_{1,3}(x)\mathcal{D}_{2,1}(x)\mathcal{D}_{3,2}(x) = \mathcal{D}_{1,2}(x)\mathcal{D}_{2,3}(x)\mathcal{D}_{3,1}(x)$ and $b_{31}(x) = b_{32}(x) \Leftrightarrow \mathcal{D}_{1,3}(x)\mathcal{D}_{2,1}(x)\mathcal{D}_{3,2}(x) = \mathcal{D}_{1,2}(x)\mathcal{D}_{2,3}(x)\mathcal{D}_{3,1}(x)$; that is, the coefficient restriction for each possible $t$ value implies the same restriction on the model.

In practice, one may work out which restrictions are redundant for any given $T$. However, it will generally be simpler to simply test the set of restrictions implied for a particular value of $t$. We note that the discussion above implicitly assumes that sufficient denominator terms are nonzero at the value of $x$ of interest for $b_t(x)$ to be identified. Clearly any test based on these overidentifying restrictions will lack power when the model is not identified. We will discuss testing of the overidentifying restrictions in the following section where we outline a simple estimator for the marginal effects averaged over the distribution of unobserved heterogeneity.

## 4. **Estimation**

The identification result in the previous section is also constructive as it suggests a simple approach to pointwise estimation of the marginal effects of interest, $b_t(x)$. In particular, the identification result shows that $b_t(x)$ may be constructed as a combination of derivatives of conditional expectations that are readily estimated using any number of nonparametric derivative estimators. It is again worth pointing out that the identification result developed above requires that the term that shows up in the denominator of the expression for $b_t(x)$ be nonzero. Without strong beliefs that the population derivatives that appear in the denominator are bounded away from zero almost everywhere, this suggests that identification and estimation may best viewed pointwise as, just as in instrumental variables models when examining the first stage relationship, one will likely want to consider the validity of the identification assumption on a point by point basis. For this reason, we present the following results for pointwise estimation of $b_t(x)$.

Let $\widehat{\mathcal{D}}^m_{t,\tau}(x)$ be an estimate of the derivative $\mathcal{D}^m_{t,\tau}(x)$. We define an estimator of $b_t(x)$ as $\widehat{b}_t(x) = \widehat{\mathcal{D}}^m_{t,t}(x) - g(\widehat{\mathcal{D}}^m_{s,\tau}(x), b_n)\widehat{\mathcal{D}}^m_{s,t}(x)\widehat{\mathcal{D}}^m_{t,\tau}(x)$ where $g(t,b)$ is a trimming function and $b_n$ is a bandwidth. For simplicity, we assume that the functional used to estimate the ratio of derivatives of the index function is the same as the functional of interest, and suppress the the superscript $m$ as well as the argument $x$ in the notation below.

Before continuing, we note that the ratio of derivatives of the index function may also be estimated by $g(\widehat{\mathcal{D}}^\varphi_{s,\tau}(x), b_n)\widehat{\mathcal{D}}^\varphi_{s,t}(x)$, where the functionals $\mathcal{D}^\varphi_{\cdot,\cdot}$ are obtained by replacing the function $m$ in equation (3.3) with another function $\varphi$ in the domain of the appropriate conditional expectation operator. When the support of $y$ is rich, this offers considerable flexibility in estimating the ratio $h_t/h_\tau$. Assuming that there are multiple functions that provide non-zero denominators, it may also be possible to improve the efficiency of the estimator by considering averages over a range of such functions. Modifying the estimation results presented below in this direction seems like an interesting technical extension, but is beyond the scope of the present paper.

To establish the pointwise properties of the estimator, we impose the following conditions:

A1. $|\mathcal{D}_{s,\tau}| > 0$.

A2. For $\mathcal{D} = (\mathcal{D}'_1, ..., \mathcal{D}'_T)'$ where $\mathcal{D}_t = (\mathcal{D}_{t,1}, ..., \mathcal{D}_{t,T})'$, $a_n(\widehat{\mathcal{D}} - \mathcal{D} - B_n) \xrightarrow{d} N(0, V_x)$.

A3. $g(u,b) = (1/u)1(|u| \geq 2b) + m(u,b)1(b \leq |u| < 2b)$ is two times continuously differentiable and $b_n \to 0$ as $n \to \infty$.

Condition A1 is necessary for identification of the model at $x$ using $\mathcal{D}_{s,\tau}$ in the denominator. All that is required for identification is that A1 is satisfied for at least one $\{s, \tau\}$ pair at the point $x$. We note that A1 is similar to the usual rank condition for identification in a linear instrumental variables model. The identification in our model comes from variation induced in the conditional expectation of $m(y_\tau)$ for $\tau \neq t$ where $t$ is the period of the effect we are trying to identify through varying $X$'s that are not included in the structural equation. If at the point of interest, changing $X$ does not induce variation in the conditional expectation at other time periods, which may be viewed analogously to the first stage in the conventional linear IV model, the model will not be identified at that $x$.

Condition A2 simply requires pointwise asymptotic normality of the derivatives of the conditional expectation. Sufficient conditions for this condition to be satisfied are well-known for a variety of nonparametric estimators. See, for example, Pagan and Ullah (2006) and Li and Racine (2006). The last condition defines a trimming function and assumes that asymptotically the trimming vanishes and is not restrictive. We introduce trimming because it is easy to do so and may induce additional regularity in the estimator. We show in the appendix that if the trimming is aggressive enough, the estimator remains well-behaved, though is not consistent for the quantity of interest, even when the model is not identified.

Under these conditions, we can verify the pointwise asymptotic normality of our estimator of the marginal effect of $x$ integrated against the conditional distribution of unobserved heterogeneity. We provide the proof in the appendix.

**Proposition 2.** *Suppose the data $\{Y, X\}$ were generated according to model (2.1) and (2.4), that conditions A1-A3 are satisfied, and define $w = e_{tt} - \frac{\mathcal{D}_{s,t}}{\mathcal{D}_{s,\tau}} e_{\tau t} - \frac{\mathcal{D}_{\tau,t}}{\mathcal{D}_{s,\tau}} e_{st} + \frac{\mathcal{D}_{s,t}\mathcal{D}_{\tau,t}}{\mathcal{D}_{s,\tau}^2} e_{s\tau}$ where $e_{ij}$ is a vector defined as $\mathcal{D}$ above with a one in the location corresponding to $\mathcal{D}_{i,j}$ and zeros elsewhere. Then $\widehat{b}_t - b_t \xrightarrow{p} 0$ and $a_n(\widehat{b}_t - b_t - w'B_n) \xrightarrow{d} N(0, w'V_x w)$.*

We note that Proposition 2 may easily be extended to provide the joint distribution between multiple estimates of $b_t(x)$ when they exist or to estimates of $b(x) = (b_1(x), ..., b_T(x))'$ by changing the definition of the vector $w$ to an appropriate matrix where each column of the matrix is defined analogously to $w$ above but picks off and correctly weights the appropriate elements from the vector of derivative estimators $\widehat{\mathcal{D}}$. We present this result in the following notationally burdensome corollary and note that the proof is the same as that of Proposition 1 with appropriate changes of notation.

**Corollary 1.** *Suppose the data $\{Y, X\}$ were generated according to model (2.1) and (2.4). Let $\beta(x)$ be a vector of estimates $b_{t,\tau,s}(x) = \mathcal{D}_{t,t}(x) - \left(\frac{\mathcal{D}_{s,t}(x)}{\mathcal{D}_{s,\tau}(x)}\right)\mathcal{D}_{t,\tau}(x)$ where we have explicitly indexed by the time period of the effect of interest and the other two time periods used to estimate the index function. Suppose $\beta(x) = (b_{t_1,\tau_1,s_1}(x), ..., b_{t_J,\tau_J,s_J}(x))'$ where $t$, $s$, and $\tau$ are $J \times 1$ vectors with $j^{th}$ elements $t_j$, $s_j$, and $\tau_j$ respectively. Suppose that conditions A1-A3 are satisfied for each element of $\beta(x)$, and define a matrix $w$ with $l^{th}$ column $w_l = e_{t_l t_l} - \frac{\mathcal{D}_{s_l,t_l}}{\mathcal{D}_{s_l,\tau_l}} e_{\tau_l t_l} - \frac{\mathcal{D}_{\tau_l,t_l}}{\mathcal{D}_{s_l,\tau_l}} e_{s_l t_l} + \frac{\mathcal{D}_{s_l,t_l}\mathcal{D}_{\tau_l,t_l}}{\mathcal{D}_{s_l,\tau_l}^2} e_{s_l \tau_l}$ where $e_{ij}$ is a vector defined as $\mathcal{D}$ above with a one in the location corresponding to $\mathcal{D}_{i,j}$ and zeros elsewhere. Then $\widehat{\beta} - \beta \xrightarrow{p} 0$ and $a_n(\widehat{\beta} - \beta - w'B_n) \xrightarrow{d} N(0, w'V_x w)$.*

Corollary 1 gives the limiting distribution of a vector of estimators $\beta(x)$ composed of elements of the form $b_{t,\tau,s}(x) = \mathcal{D}_{t,t}(x) - \left(\frac{\mathcal{D}_{s,t}(x)}{\mathcal{D}_{s,\tau}(x)}\right)\mathcal{D}_{t,\tau}(x)$. The results can be used to test joint hypotheses involving derivatives across multiple time periods or to test overidentifying restrictions. As discussed in Section 3.3, one would typically fix $t$ when forming $\beta(x)$ and consider all admissible $\{s,\tau\}$ combinations for estimating $b_t(x)$ when considering testing overidentifying restrictions. One could then use $(a_n)^2(R(\widehat{\beta} - \widehat{w'B_n}))'(R\widehat{w'V_x w}R')^{-1}(R(\widehat{\beta} - \widehat{w'B_n})) \xrightarrow{d} \chi^2_{T^2-3T+1}$ for testing the null hypothesis $R\beta = 0$ implied by the overidentifying restrictions where $R$ is an appropriately defined restriction matrix.

When considering testing overidentifying restrictions, it is worth noting that the tests may be formulated in terms involving only products, not ratios, of the raw derivatives. Note that the overidentifying restrictions may always be represented as a set of equality restrictions of the form

$$\left(\frac{\mathcal{D}^m_{s,t}(x)}{\mathcal{D}^m_{s,\tau}(x)}\right)\mathcal{D}^m_{t,\tau}(x) - \left(\frac{\mathcal{D}^m_{r,t}(x)}{\mathcal{D}^m_{r,r'}(x)}\right)\mathcal{D}^m_{t,r'}(x) = 0$$

for various $\{s,\tau\}$ and $\{r,r'\}$. We can then always multiply through by the product of the denominator terms to produce a test which does not involve ratios of derivative estimators. For example, we may test either $\left(\frac{\mathcal{D}_{2,1}(x)}{\mathcal{D}_{2,3}(x)}\right)\mathcal{D}_{1,3}(x) = \left(\frac{\mathcal{D}_{3,1}(x)}{\mathcal{D}_{3,2}(x)}\right)\mathcal{D}_{1,2}(x)$ or $\mathcal{D}_{1,3}(x)\mathcal{D}_{2,1}(x)\mathcal{D}_{3,2}(x) = \mathcal{D}_{1,2}(x)\mathcal{D}_{2,3}(x)\mathcal{D}_{3,1}(x)$ in the $T = 3$ case. The behavior of the test using the products is immediate given Assumption A2 and requires neither Assumption A1 nor A3. Of course, regardless of which formulation is used, the test will have no power when the model is not identified.

We also note that when there are multiple ways to estimate a $b_t(x)$, we may wish to combine these estimates to reduce variance. It is well known that for a set of $J$ estimates of $b_t(x)$ stacked as a $J \times 1$ vector, $\mathbf{b}_t(x)$, the variance minimizing linear combination of $\mathbf{b}_t(x)$ is given by $\mathbf{c}'\mathbf{b}_t(x)$ where $\mathbf{c} = V_{\mathbf{b}}^{-1}\iota/(\iota' V_{\mathbf{b}}^{-1}\iota)$ where $\iota$ is a $J \times 1$ vector of ones and $V_{\mathbf{b}}$ is the variance-covariance matrix of $\mathbf{b}_t(x)$. This quantity may readily be estimated, and it's asymptotic distribution follows from Corollary 1. We explore the use of this estimator in the empirical example presented below.

## 5. **Insider Trading Example**

To illustrate the approach developed in the previous sections, we present a brief application that considers the association between insider trading and trading volume of an asset as measured by turnover in a given quarter. We illustrate the index assumption by discussing it in the context of this particular economic application. We also demonstrate the use of the estimator discussed in Section 4 and the related overidentification test.

### 5.1. **Empirical Model and Assumptions**

Let $y_{it}$ denote a measure of trading activity over a quarter $t$ by firm $i$ insiders. Denote by $turn_{it}$ the turnover of firm $i$'s stock during a quarter: the total trading volume of firm $i$'s stock during quarter $t$ divided by shares outstanding. Let $\alpha_i$ be a firm-specific cost of insider trading which may be related to the observed turnover. In some of our specifications, we also consider excess returns to holding a firm's stock for six months following the quarterly earnings announcement in quarter $t$, $r_{i,t}$. These returns are included in the structural model in order to proxy for insiders' portfolio rebalancing and allow informed traders to make decisions based on future returns. We consider specifications where returns enter both the structural model and the distribution over unobserved heterogeneity in a fairly flexible, dynamic way. Specifically, in our panel with $T = 3$, we allow all returns from time -1 to time 4, $r_{i,-1}$ to $r_{i,4}$, to enter both the structural model and the distribution of unobserved heterogeneity in an unrestricted fashion. This means that the average marginal effect of returns on expected trading activity are unidentified. Since we are chiefly interested in turnover and believe that informed insiders may potentially make trading decisions based on both past and future asset returns, we believe the added flexibility outweighs the cost of nonidentification of the effects of returns.

Much of the existing empirical literature on insider trading, e.g. Meulbroek (1992) and Rozeff and Zaman (1998), has focused on the cross-sectional variation of future returns as a function of past insider activity. This analysis is complicated by endogeneity and the need for a large set of controls. In our analysis, we choose to focus on the turnover variable for simplicity. A simple prediction of models of informed trade, e.g. Kyle (1985), is that insider activity should respond to market volume. In particular, insiders trading on nonpublic information should be more likely to engage in buying or selling of own company stock when overall trading volume is high, as high market volume may conceal trading by insiders and help to mitigate the price impact of their trades. However, we expect that unobservable, firm-specific characteristics may be correlated with observed turnover, for example because boards of directors recognize that turnover affects incentives associated with insider trading and adjust firm policy accordingly due to concerns about regulatory scrutiny. Differences in insiders' trading activity across firms with different levels of turnover may therefore be due to systematic differences in firm policy as well as the direct effect on incentives. The approach we have developed in this paper provides a flexible framework within which to examine the relationship between insider trading and turnover while controlling flexibly for correlated firm-specific heterogeneity and without imposing strong functional form assumptions.

We consider a simple empirical model of the form

$$y_{it} = g_t\left(turn_{it}, \alpha_i, u_{it}\right)$$
$$\alpha_i = q\left(h(turn_{i1}, turn_{i2}, turn_{i3}), v_i\right)$$

where $u_{it}$ and $v_i$ are unobserved error terms that are independent of each other and of all observables. We note that the model allows flexibly for intertemporal heteroskedasticity by allowing the form of the function $g_t$ and $u_{it}$ to change with $t$. In some specifications, past and future returns, $r_{i,-1}$, $r_{i0}$, $r_{i1}$, $r_{i2}$, and $r_{i3}$, are also included in the structural model, $g_t$. Including additional controls would almost certainly entail explicitly adopting a semiparametric model due to dimensionality considerations, an extension beyond the scope of the present paper. Two important restrictions are evident in the reduced form: first, that the variable of interest $turn_{it}$ enter the distribution of firm-specific effects through an index $h : \mathbb{R}^3 \to \mathbb{R}$, and second, that conditional on $\alpha_i$ and possibly the return path, only the current quarter realization of turnover affects the current quarter outcome, $y_{it}$.

In this case, we can think of the index restriction as an assumption that there are types of firms in the economy and that these firm types are related to turnover in a flexible but restricted fashion. A simple index restriction that has been used elsewhere in the economics literature is that $h(turn_{i1}, turn_{i2}, turn_{i3}) = \sum_{t=1}^{3} turn_{it}$, which defines types of firms based on the total amount of turnover over a three quarter period. That is, all firms that have high total turnover over the period (firms for which $h(turn_{i1}, turn_{i2}, turn_{i3})$ is large) have unobservables that are drawn from similar distributions, while firms which have low total turnover over the period (firms for which $h(turn_{i1}, turn_{i2}, turn_{i3})$ is small) also have unobservables drawn from similar distributions, though these distributions may differ substantially across high and low turnover firms. In particular, it seems plausible that boards of directors of firms with high overall turnover might be more likely to institute restrictions on insider trading due to concerns about regulatory scrutiny and the awareness that high turnover potentially makes trading on nonpublic information more profitable.

We note that the index assumption is, of course, much more general than this, allowing for much more exotic index functions. For example, insider trading may also tend to be more tightly regulated at firms where turnover is more volatile. The substantive restriction is that we must believe that there are types of firms in the economy that are well-defined by a single function of the path of turnover over the year. It is also worth noting that the index assumption may be more plausible as we add additional controls to the model. In our analyses with returns included, firms will be of similar type only if they have similar values of $h(turn_{i1}, turn_{i2}, turn_{i3})$ over the period and similar return paths over the six period window considered.

The other substantive restriction in the model is that conditional on firm-specific heterogeneity and possibly the return path, insiders only base trading decisions during a quarter on turnover during that quarter. This should be a plausible restriction in circumstances in which insiders would always like to trade on nonpublic information but are constrained by the desire to hide the activity. In this case, they should be willing to trade any time turnover is high, with the sign of the activity depending on their current period inside information, and would not necessarily consider whether turnover will be high or low in the future when making their current trading decisions.

### 5.2. **Empirical Results**

To obtain our empirical results, we use a subset of the data considered in Roulstone (2006). The data were collected from several sources, including earnings and accounting variables from Compustat; returns, market value, and trading volume from CRSP; and insider purchases and sales from Thompson Financial Insider Trading Data Feed and the National Archives of Insider Trading Summaries. Our data consist of a quarterly panel of U.S. firms from 1999Q2-1999Q4. We use firms that have one earnings announcement in each quarter of 1999 and exclude firms with missing values for any variable. We also exclude firms with turnover greater than 1.5 (150%) in a quarter and firms for which insider trading accounts for more than 10% of quarterly trading volume, in part to exclude extreme events such as acquisitions or bankruptcy. For our outcomes, we consider a binary variable for whether there was any insider selling activity within a firm during a given quarter and a binary variable for whether there was any insider buying activity within a firm during a given quarter. As noted above, a simple model of informed trade suggests that turnover should be positively related to both the probability of insider buys and the probability of insider sales. Our final sample is a balanced panel with $N = 3804$ firms over $T = 3$ quarters. Descriptive statistics are provided in Table 1.

To obtain our estimation results, we used a local linear estimator to nonparametrically estimate the set of derivatives used in constructing our point estimates of the marginal effect of changing turnover on the probability of insider buys or sales. For all of our estimates, we used a Gaussian kernel with bandwidth matrix equal to $hCov(X)$ where $h$ is a bandwidth parameter and $Cov(X)$ is the sample covariance matrix of the explanatory variables in the model. For all specifications, we chose $h$ by using cross-validation for estimates of the conditional expectation of $E[y_t|turn_1, turn_2, turn_3]$ or $E[y_t|turn_1, turn_2, turn_3, r_{-1}, r_0, r_1, r_2, r_3]$ depending on which covariates appear in the model. While we recognize that this may not be the optimal choice of bandwidth for our problem, it seemed *a priori* reasonable and provides sensible results. We have also calculated results using .50 and 1.25 times the cross-validated bandwidth (available upon request) and obtain qualitatively similar results. Finally, for all of our results we estimate asymptotic standard errors and report bootstrapped critical values for t-statistics and for the test of the overidentifying restriction. As with the point estimates, we obtained bootstrapped critical values for our test statistics for a variety of bandwidths ranging between .50 and 1.25 times the cross-validated bandwidth. We note that the majority of these bandwidths should be undersmoothing relative to the rate optimal bandwidth and so should alleviate bias concerns in obtaining bootstrapped critical values for tests using nonparametric estimators. In all cases, the results remain qualitatively similar to what is reported, and the bootstrapped critical values are numerically similar across all bandwidths considered. All bootstrap results are based on 1000 replications.

We report estimated effects for insider buying in Table 2 and for insider selling in Table 3. All estimates are for the average marginal effect evaluated at the median of each $x$. We also calculated effects with all $x$'s evaluated at their marginal first and third quartiles. While slightly different, these effects were certainly not statistically distinguishable from the effects at the median, and are therefore not reported. The top panel in each table reports the results from the baseline model with just the turnover variables included, and the bottom panel reports results from the model with both turnover and return variables. To conserve space we report only the derivatives and marginal effects associated with turnover. In each table, the columns labeled $b_{t1}$ and $b_{t2}$ correspond to the two possible estimators of $b_t$ that can be constructed for each $t$. The columns labeled $\bar{b}_t$ give variance weighted averages of $b_{t1}$ and $b_{t2}$ as discussed at the end of Section 4. Finally, the column $\bar{b}$ gives the simple average of $\bar{b}_t$ which may be taken as an overall measure of the annual relationship between insider trading and trading volume, and the column labeled "Over" reports results from testing the model's overidentifying restriction. In the columns $b_{11}$, $b_{12}$, etc. corresponding to the raw estimates of the marginal effect at time $t$, we also report the results for the "first-stage" coefficients, the derivative term involved that appears in the denominator of the estimator of marginal effect indicated by the column label. Rows labeled "estimate" and "s.e." give respectively the estimated values and standard errors. Rows labeled $c_{80}$, $c_{90}$, and $c_{95}$ respectively give bootstrapped $80^{\text{th}}$, $90^{\text{th}}$, and $95^{\text{th}}$ percentiles for t-statistics. For the "Over" column, these rows correspond to the appropriate critical values from the overidentification test.

We focus first on Table 2, which reports estimated marginal effects of turnover on the probability of insider purchases of own company stock with turnover in each period evaluated at the median of the marginal distribution. Intuitively, insider purchases of own-company stock may be more sensitive to turnover as many corporate insiders likely have large holding of own-company stock through their compensation schemes. This suggests that purchases of own-company stock are more likely to be motivated by nonpublic information than sells of own-company stock, which should occur frequently for portfolio rebalancing purposes, and that there is more of an incentive to try to "hide" these purchases in relatively higher volume periods.

Panel A of Table 2 contains results when only turnover is included as a conditioning variable. Looking first at the bottom half (II) of Panel A which contains the "first-stage" results, we see that all of the denominator terms appear to be statistically significantly different from 0 at the 5% level, suggesting that the marginal effects of interest are identified. The overidentification test statistic is also small, and we fail to reject the hypothesis that the overidentifying restriction is satisfied. This failure to reject suggests that the index restriction may be consistent with the data. The results for the estimated marginal effects in the top half (I) of Panel A are also consistent with the hypotheses of a simple model of informed trading. All of the individual marginal effects are positive though imprecisely estimated. When we consider the average estimates $\bar{b}_1$-$\bar{b}_3$ and $\bar{b}$, we see fairly strong evidence of positive effects across the three time periods and averaged over the whole time period.

In particular, all of the average estimates are sizable positive numbers, ranging from 0.227 to 0.410, and statistically significant at the 5% level. The point estimates are also large enough to seem economically relevant. Taking the average effect over the three quarters $\bar{b}$, we see that the point estimate implies an approximate one standard deviation increase in trading volume (an increase in turnover of around .2) is associated with around a 6% increase in the probability of insider buying.

The results reported in Panel B of Table 2, which includes firm-level stock return variables as additional controls, are roughly consistent with the results that exclude the return variables. In this case, the denominator terms corresponding to $b_{12}$ and $b_{32}$ are small relative to the estimated standard errors, and we see that the resulting estimates of the marginal effects are quite variable and that bootstrapped critical values for the t-statistic are also quite large. The other four denominators and estimated effects seems to reasonably well-estimated though, and all estimated effects are positive as we would expect. Looking at the average effects $\bar{b}_1$-$\bar{b}_3$ and $\bar{b}$, we again see fairly strong evidence of positive effects across the three time periods and averaged over the whole time period. The affects are somewhat attenuated relative to those estimated without the returns variables but remain statistically significant at the 5% level. In this case, the point estimate for the average effect over the three quarters, $\bar{b}$, implies an approximate one standard deviation increase in trading volume (an increase in turnover of around .2) is associated with around a 4% increase in the probability of insider buying.

Results for insider sells are presented in Table 3. These results are considerably less precise than those for insider purchases, but still suggest a positive effect of turnover on the probability of insider sales. All but two of the individual estimates of marginal effects are positive, with the two negative estimates being quite small and imprecisely estimated. Unfortunately, there are more problems with the "first-stage" on the sell side, with half of the denominators being statistically insignificant once returns are added to the model. Once again this translates into large estimated standard errors and an apparently very heavy-tailed distribution for the associated t-statistic. In the model without returns, the average effect in period 2 is statistically different from zero while none of the other average effects are statistically different. For the model with returns included, none of the averages are statistically significant. We note that this is not terribly surprising as insiders have strong incentives, mainly for portfolio rebalancing reasons, to sell stock regularly regardless of the level of turnover in the market.

Overall, the results are consistent with a simple model of insiders making informed trades; e.g. Kyle (1985). We find statistical evidence for the hypothesis that high turnover should be associated with increased insider purchases and sells of own-company stock despite the high-dimensional nonparametric regressions involved. These results are obtained while controlling flexibly for unobserved firm-level heterogeneity and without imposing strong functional form assumptions, though there is obviously no free lunch as one must live with running and believing a high-dimensional nonparametric regression in a moderately sized sample. The results suggest that the proposed estimation procedure is feasible and may usefully complement other more parametric procedures.

## 6. **Conclusion**

In this paper, we offer a simple approach to identification in a correlated random effects model. The identification results allows for a nonparametric structural likelihood and an essentially nonparametric conditional distribution of unobserved individual level heterogeneity. The identification result also imposes no stationarity restrictions in the time dimension and thus allows for extremely general types of intertemporal heteroskedasticity. We show that in this general setting identification may be achieved under an index assumption in the conditional distribution of unobserved heterogeneity. In particular, we show that identification may be obtained if there is a sufficient statistic for $x$ in the conditional distribution of unobserved heterogeneity; that is, $q(\alpha|x_1, ..., x_T, h(x_1, ..., x_T)) = q(\alpha|h(x_1, ..., x_T))$ for some function $h(x_1, ..., x_T)$. Our identification results allow the function $h(x_1, ..., x_T)$ to be an unknown nonparametric function subject to some smoothness conditions. Because we are imposing restrictions on the distribution of unobserved effects, our approach is best thought of as a random effects approach though it is very flexible, allowing nonparametric specification of the structural likelihood, the distribution of unobserved heterogeneity conditional on the sufficient statistic $h(x_1, ..., x_T)$, and the function $h$. The sufficient statistic $h$ may also be viewed a general index function that defines a set of types.

We also present a simple estimator, discuss testing of overidentifying assumptions, and present an empirical application. The empirical application looks at the effect of turnover in a company's stock on buying and selling activity of corporate insiders. We find fairly consistent evidence that, controlling for unobserved firm level heterogeneity, corporate insiders are more likely to purchase own company stock in quarters where turnover in the stock is high as would be predicted by a simple of informed insider trading, e.g. Kyle (1985). The evidence on the sell side is weaker though consistent with insiders being more likely to sell own-company stock when turnover is high. These results should complement available results which are based on parametric models of insider trading decisions.

The chief difficulty in implementing the present approach is that it requires nonparametric derivative estimation over a high-dimensional space. This suggests that the approach will essentially be unfeasible in smaller panels. A useful direction for further research related to the present paper would be in providing models under sensible semiparametric restrictions. A useful approach which has been pursued in other contexts and seems particularly well-suited to the present application would be to assume that the statistic $h$ takes the form of a linear index. To keep treatment of the structural function and the distribution of unobserved effects parallel, one may also wish to impose that the structural function is also a nonparametric function of a linear index. We leave this extension to future research.

# 7. Appendix

## 7.1. Extensions of Identification Result

### 7.1.1. *Dynamic Case*

The identification argument in Section 3 assumes that, conditional on the unobserved effect $\alpha$, only the contemporaneous value of the covariate, $x_t$, affects the period-$t$ outcome, $y_t$. In this section, we show how our approach may be extended to the case where the $x$ and $y$ values from the previous period also enter the structural model. We introduce an initial condition $(y_0, x_0)$ which, for simplicity, is treated as exogenous but may enter the distribution of individual-specific effects. Note that the entire vector of covariates $x$ (which now includes the initial condition) is treated as known at all dates $t$. Denoting by $\mathbf{y}_{t-1} = (y_{t-1}, \ldots, y_0)$ the vector of observations for $y$ up to time $t-1$, the model is then

$$y_t|x, \mathbf{y}_{t-1}, \alpha \sim G_t(y_t|x_t, y_{t-1}, x_{t-1}, \alpha) \qquad\qquad \alpha|x, y_0 \sim Q(\alpha|x, y_0).$$

Similar to the static case, we wish to identify the marginal effects of a change in $x_t$, and now also $x_{t-1}$ and $y_{t-1}$ on the current period outcome with the distribution of unobserved heterogeneity held fixed, given by

$$b_t(x, \mathbf{y}_{t-1}) = \int_{\mathcal{A}} \left( \frac{\partial}{\partial x_t} \mathrm{E}[m(y_t)|x, \mathbf{y}_{t-1}, \alpha] \right) dQ(\alpha|h(x), y_0)$$

$$b_t^{x-}(x, \mathbf{y}_{t-1}) = \int_{\mathcal{A}} \left( \frac{\partial}{\partial x_{t-1}} \mathrm{E}[m(y_t)|x, \mathbf{y}_{t-1}, \alpha] \right) dQ(\alpha|h(x), y_0)$$

$$b_t^{y-}(x, \mathbf{y}_{t-1}) = \int_{\mathcal{A}} \left( \frac{\partial}{\partial y_{t-1}} \mathrm{E}[m(y_t)|x, \mathbf{y}_{t-1}, \alpha] \right) dQ(\alpha|h(x), y_0)$$

We first note that identification of the marginal effect of the lagged dependent variable $b_t^{y-}$ is immediate for $t \geq 2$ because the previous period outcome $y_{t-1}$ is observable and does not affect the distribution of unobserved heterogeneity. To see how our identification result changes in dynamic models, we define $\mathcal{D}_{t,\tau}^m(x; y_{t-1}, y_0) = \frac{\partial}{\partial \tau} \mathrm{E}\left[m(y_t)|x, y_{t-1}, y_0\right]$ and note that, as in (3.2),

$$\begin{aligned}
\mathcal{D}_{t,t}^m(x; y_{t-1}, y_0) = & \int_{\mathcal{A}} \int_{\mathcal{Y}_t} m(y_t)\, d\left[ \frac{\partial}{\partial x_t} G_t(y_t|x_t, y_{t-1}, x_{t-1}, \alpha) \right] dQ(\alpha|h(x_1, ..., x_T), y_0) \\
& + \left( \frac{\partial h(x)}{\partial x_t} \right) \int_{\mathcal{A}} \int_{\mathcal{Y}_t} m(y_t) dG_t(y_t|x_t, y_{t-1}, x_{t-1}, \alpha)\, d\left[ \frac{\partial}{\partial h} Q(\alpha|h(x_1, ..., x_T), y_0) \right] \\
= & \; b_t(x; y_{t-1}, y_0) + h_t(x) c_t(x; y_{t-1}, y_0).
\end{aligned}$$

In the static case, when $\tau \neq t$, all derivatives $\mathcal{D}_{t,\tau}^m$ were proportional to the integral term, $\mathcal{D}_{t,\tau}^m = h_\tau c_t$. This is still the case *except* when $\tau = t - 1$, where

$$\begin{aligned}
\mathcal{D}_{t,t-1}^m(x; y_{t-1}, y_0) = & \int_{\mathcal{A}} \int_{\mathcal{Y}_t} m(y_t)\, d\left[ \frac{\partial}{\partial x_{t-1}} G_t(y_t|x_t, y_{t-1}, x_{t-1}, \alpha) \right] dQ(\alpha|h(x_1, ..., x_T), y_0) \\
& + \left( \frac{\partial h(x)}{\partial x_{t-1}} \right) \int_{\mathcal{A}} \int_{\mathcal{Y}_t} m(y_t) dG_t(y_t|x_t, y_{t-1}, x_{t-1}, \alpha)\, d\left[ \frac{\partial}{\partial h} Q(\alpha|h(x_1, ..., x_T), y_0) \right] \\
= & \; b_t^{x-}(x; y_{t-1}, y_0) + h_{t-1}(x) c_t(x; y_{t-1}, y_0).
\end{aligned}$$

Intuitively, in static models with the index $h$ unknown, we require information from two additional time periods to identify average marginal effects of the covariate $x_t$: one, $\tau \neq t$, to identify how a change in the index $h$ affects the average and another, $s \notin \{t, \tau\}$ to identify the ratio of derivatives of the index, $h_t/h_\tau$. We can use the same approach here. Provided that $\{t, \tau, s\} \subseteq \{1, \ldots, T\}$ may be chosen with $\tau \notin \{t, t-1\}$ and $\{t, \tau\} \cap \{s-1, s\} = \emptyset$, we have

$$b_t(x; y_{t-1}, y_0) = \mathcal{D}_{t,t}^m(x; y_{t-1}, y_0) - \left( \frac{\mathcal{D}_{s,t}^\varphi(x; y_{s-1}, y_0)}{\mathcal{D}_{s,\tau}^\varphi(x; y_{s-1}, y_0)} \right) \mathcal{D}_{t,\tau}^m(x; y_{t-1}, y_0), \tag{7.1}$$

and if, in addition, $s \neq t-1$, we have

$$(7.2) \qquad b_t^{x-}(x;y_{t-1},y_0) = \mathcal{D}^m_{t,t-1}(x;y_{t-1},y_0) - \left(\frac{\mathcal{D}^\varphi_{s,t-1}(x;y_{s-1},y_0)}{\mathcal{D}^\varphi_{s,\tau}(x;y_{s-1},y_0)}\right)\mathcal{D}^m_{t,\tau}(x;y_{t-1},y_0).$$

We assume $0 \notin \{t,\tau,s\}$ as the initial condition is treated as exogenous. One may identify the first period marginal effects $b_1$ and $b_1^{x-}$, but not $b_1^{y-}$. The latter is not identified because we allowed the initial outcome $y_0$ to enter separately from the index, $Q(\alpha|h(x),y_0)$. We could also allow $x_0$ to enter $Q$ separately from $x_1,\ldots,x_T$, in which case $b_1^{x-}$ would also fail to be identified.

As before, we assume $\varphi$ can be chosen so the denominator $\mathcal{D}^\varphi_{s,\tau}$ is nonzero, with $\varphi = m$ permitted. Estimation of $b_t$ and $b_t^{x-}$ may therefore proceed exactly in static models, but with the lagged dependent variable $y_{t-1}$ and the initial condition $y_0$ included in the conditioning set when estimating conditional expectations $\mathrm{E}[m(y_t)|\cdot]$ and their derivatives $\mathcal{D}^m_{t,\cdot}$. The additional restrictions on the set of $(\tau,s)$ from which $b_t$ and $b_t^{x-}$ may be estimated ensure that each of the three $\mathcal{D}^m_{r,r'}$ that appear in the last right hand side term of (7.1) and (7.2) have the form $\mathcal{D}^m_{r,r'} = h_{r'}c_r$. For example, if $\tau = t-1$, we would have $\mathcal{D}^m_{t,\tau} = b_t^{x-} + h_\tau c_t$, and (7.1) would not hold for values in $\mathcal{X}$ where $b_t^{x-} \neq 0$.

One major difference between the dynamic and static cases is that it is possible for marginal effects of covariates in some time periods to be identified while others are not. For example, in a dynamic model when $T = 3$, the marginal effect of a contemporaneous change in $x_2$, $b_2(\cdot)$, is identified (set $\tau = 3$ and $s = 1$), while the third period marginal effect, $b_3(\cdot)$, and all marginal effects associated with lagged covariates, $b_t^{x-}$, are not. With $T = 4$, all contemporaneous marginal effects are identified, but $b_2^{x-}$ is not. Another difference is that the number of overidentifying restrictions implied by the index assumption is smaller than in the static case, and in some cases certain marginal effects will be just identified. For example, in a static model with $T = 3$, we can form two distinct estimators of $b_2(\cdot)$ based on (3.4), one by putting $\tau = 3$ and $s = 1$ and the other with $\tau = 1$ and $s = 3$. In the dynamic case, the latter estimator would, in general, be inconsistent unless $b_3^{x-} = b_2^{x-} = 0$, so the marginal effect $b_2(\cdot)$ is just identified.

Our approach may easily be extended to models where multiple lagged values of $x$ or $y$, or both, enter the structural model. Each additional lagged $x$ will further restrict the set of admissible $(\tau,s)$

for identification of contemporaneous and lagged effects for each given $t$, and in general will require a larger panel length $T$ to ensure identification of all marginal effects of interest.

### 7.1.2. *Marginal Effects Involving Outcomes in Multiple Periods*

We briefly consider estimation of marginal effects of a change in the covariate $x_t$ on a function, $M$, that involves outcomes in multiple periods. For example, we may wish to determine how a change in household income this year affects the probability a household buys a car this year *or* next year.

To fix ideas, we suppose that we are in the dynamic setting described in the previous section, where the structural model for $y_t$ contains the contemporaneous value of $x$ and the values of $x$ and $y$ from the previous period, and that we are interested in estimating marginal effects of the form

$$B_{\{t,\ldots,t+k\}}(x; y_{t-1}, y_0) = \int_{\mathcal{A}} \frac{\partial}{\partial x_t} \mathrm{E}[\, M(y_t, y_{t+1}, \ldots, y_{t+k})\, |x, \alpha] dQ(\alpha|x),$$

where $M : \mathbb{R}^{k+1} \to \mathbb{R}$ is a function that depends on outcomes in current and future periods. It is not necessary to assume that the outcomes that enter $M$ occur in the future or that they are consecutive; we make this assumption here to simplify notation and because such quantities are of interest in many economic settings. The expectation of $M$ conditional on the observed covariates $x$ and the lagged value $y_{t-1}$ can be written

$$\begin{aligned}\mathrm{E}[M(y_t, \ldots, y_{t+k})|x, y_{t-1}, y_0] = \int_{\mathcal{A}} \bigg\{ \int_{\underline{\mathcal{Y}} \times \ldots \times \underline{\mathcal{Y}}} & M(y_t, \ldots, y_{t+k})\, dG_{t+k}(y_{t+k}|x_{t+k}, y_{t+k-1}, x_{t+k-1}, \alpha) \\ & \ldots dG_t(y_t|x_t, y_{t-1}, x_{t-1}, \alpha) \bigg\} dQ(\alpha|h(x), y_0).\end{aligned}$$

As before, we differentiate this expectation with respect to the covariate $x_t$ and have

$$\begin{aligned}\mathcal{D}^M_{\{t,\ldots,t+k\},t}(x; y_{t-1}, y_0) &= \int_{\mathcal{A}} \int_{\underline{\mathcal{Y}} \times \ldots \times \underline{\mathcal{Y}}} M\, d\left[ \left( \frac{\partial G_t}{\partial x_t} G_{t+1} + G_t \frac{\partial G_{t+1}}{\partial x_t} \right) \prod_{j=t+2}^{k} G_{t+j} \right] dQ \\ &\quad + \left( \frac{\partial h(x)}{\partial x_t} \right) \int_{\mathcal{A}} \int_{\underline{\mathcal{Y}} \times \ldots \times \underline{\mathcal{Y}}} M d \left[ \prod_{j=0}^{k} G_{t+j} \right] d\left[ \frac{\partial Q}{\partial h} \right] \\ &= B_{\{t,\ldots,t+k\}}(x; y_{t-1}, y_0) \quad + \quad h_t(x)\, C_{\{t,\ldots,t_k\}}(x, y_{t-1}, y_0),\end{aligned}$$

where $\mathcal{D}^M_{\{t,\ldots,t+k\},\tau}(x; y_{t-1}, y_0) = \frac{\partial}{\partial x_\tau} \mathrm{E}[M(y_t, \ldots, y_{t+k})|x, y_{t-1}, y_0]$ and arguments in the first equality are omitted to ease notation.

Our strategy here is similar to that in section 7.1.1. Conditional on $\alpha$, the distribution of $(y_t, \ldots, y_{t+k})$ depends only on $(x_{t-1}, x_t, \ldots, x_{t+k})$. Other values of $x$, from time periods before $t-1$ or after $t+k$, affect the conditional expectation $\mathrm{E}[M|x, y_{t-1}, y_0]$ only through the distribution of unobserved effects via the index, $h$. Therefore, for time periods $\tau < t-1$ or $\tau > t+k$, its derivatives with respect to $x_\tau$ will have the form $\mathcal{D}^M_{\{t,\ldots,t+k\},\tau} = h_\tau C_{\{t,\ldots,t_k\}}$. It follows that if we can find a pair $(\tau, s)$ with $1 \leq \tau \leq T$ and $2 \leq s \leq T-k$ such that $\tau \notin \{t-1, t, \ldots, t+k\}$ and $\{t,\tau\} \cap \{s-1, s\} = \emptyset$, we can form an estimate of the marginal effect of interest using the identity

$$\begin{aligned}B_{\{t,\ldots,t+k\}}(x; y_{t-1}, y_0) = \mathcal{D}^M_{\{t,\ldots,t+k\},t}(x; y_{t-1}, y_0) \\ - \left(\frac{\mathcal{D}^\varphi_{s,t}(x; y_{s-1}, y_0)}{\mathcal{D}^\varphi_{s,\tau}(x; y_{s-1}, y_0)}\right) \mathcal{D}^M_{\{t,\ldots,t+k\},\tau}(x; y_{t-1}, y_0).\end{aligned}$$

For simplicity, we consider only functions of a single period outcome to determine the ratio of derivatives of the index, and assume that $\varphi$ can be chosen to make the denominator nonzero. As with the inclusion of lagged dependent variables in the structural model, identification of marginal effects that involve outcomes in multiple periods places further restrictions on the two additional time periods $(\tau, s)$ that can be used to estimate the effect of a change in the distribution of unobserved effects, $C_{\{\cdot\}}$, and the ratio of derivatives of the index function, $h_t/h_\tau$. In practice, the panel length $T$ necessary to identify multiperiod marginal effects can quickly become large as the number of outcomes that enter $M$, or the number of lags of $x$ that enter in the structural model, increases.

### 7.2. **Proof of Proposition 2**

The proof proceeds by linearizing the estimator $\widehat{b}_t$ and verifying the remainder after linearization is of small order.

$$\begin{aligned}\widehat{b}_t - b_t &= \widehat{\mathcal{D}}_{t,t} - g(\widehat{\mathcal{D}}_{s,\tau}, b_n)\widehat{\mathcal{D}}_{s,t}\widehat{\mathcal{D}}_{\tau,t} - \mathcal{D}_{t,t} + \frac{\mathcal{D}_{s,t}}{\mathcal{D}_{s,\tau}}\mathcal{D}_{\tau,t} \\ &= (\widehat{\mathcal{D}}_{t,t} - \mathcal{D}_{t,t}) - (g(\widehat{\mathcal{D}}_{s,\tau}, b_n)\widehat{\mathcal{D}}_{s,t}\widehat{\mathcal{D}}_{\tau,t} - g(\mathcal{D}_{s,\tau}, b_n)\mathcal{D}_{s,t}\mathcal{D}_{\tau,t}) \\ &\quad - (g(\mathcal{D}_{s,\tau}, b_n)\mathcal{D}_{s,t}\mathcal{D}_{\tau,t} - \frac{\mathcal{D}_{s,t}}{\mathcal{D}_{s,\tau}}\mathcal{D}_{\tau,t})\end{aligned} \tag{7.3}$$

From A3, we have that $a_n|g(\mathcal{D}_{s,\tau}, b_n) - 1/\mathcal{D}_{s,\tau}| = a_n|m(\mathcal{D}_{s,\tau}, b_n) - 1/\mathcal{D}_{s,\tau}|1(b_n \leq |\mathcal{D}_{s,\tau}| < 2b_n) + a_n|1/\mathcal{D}_{s,\tau}|1(|\mathcal{D}_{s,\tau}| < b_n)$. Then, from $b_n \to 0$ and $|\mathcal{D}_{s,\tau}| > 0$, there exists an $n_0$ such that for all $n > n_0$ $b_n \leq |D_{s,\tau}|/2$ which implies $a_n|g(\mathcal{D}_{s,\tau}, b_n) - 1/\mathcal{D}_{s,\tau}| \to 0$. Thus

$$a_n|g(\mathcal{D}_{s,\tau}, b_n)\mathcal{D}_{s,t}\mathcal{D}_{\tau,t} - \frac{\mathcal{D}_{s,t}}{\mathcal{D}_{s,\tau}}\mathcal{D}_{\tau,t}| \leq (a_n|g(\mathcal{D}_{s,\tau}, b_n) - 1/\mathcal{D}_{s,\tau}|)|\mathcal{D}_{s,t}\mathcal{D}_{\tau,t}| \to 0. \tag{7.4}$$

Next, we have

$$\begin{aligned} g(\widehat{\mathcal{D}}_{s,\tau}, b_n)\widehat{\mathcal{D}}_{s,t}\widehat{\mathcal{D}}_{\tau,t} - g(\mathcal{D}_{s,\tau}, b_n)\mathcal{D}_{s,t}\mathcal{D}_{\tau,t} &= (g(\widehat{\mathcal{D}}_{s,\tau}, b_n) - g(\mathcal{D}_{s,\tau}, b_n))\widehat{\mathcal{D}}_{s,t}\widehat{\mathcal{D}}_{\tau,t} \\ &\quad + g(\mathcal{D}_{s,\tau}, b_n)(\widehat{\mathcal{D}}_{s,t}\widehat{\mathcal{D}}_{\tau,t} - \mathcal{D}_{s,t}\mathcal{D}_{\tau,t}), \end{aligned} \tag{7.5}$$

and, looking first at the second term in (7.5), we have

$$\begin{aligned} g(\mathcal{D}_{s,\tau}, b_n)(\widehat{\mathcal{D}}_{s,t}\widehat{\mathcal{D}}_{\tau,t} - \mathcal{D}_{s,t}\mathcal{D}_{\tau,t}) &= g(\mathcal{D}_{s,\tau}, b_n)((\widehat{\mathcal{D}}_{s,t} - \mathcal{D}_{s,t})(\widehat{\mathcal{D}}_{\tau,t} - \mathcal{D}_{\tau,t}) \\ &\quad + \mathcal{D}_{s,t}(\widehat{\mathcal{D}}_{\tau,t} - \mathcal{D}_{\tau,t}) + \mathcal{D}_{\tau,t}(\widehat{\mathcal{D}}_{s,t} - \mathcal{D}_{s,t})) \\ &= (1/\mathcal{D}_{s,\tau})(\mathcal{D}_{s,t}(\widehat{\mathcal{D}}_{\tau,t} - \mathcal{D}_{\tau,t}) + \mathcal{D}_{\tau,t}(\widehat{\mathcal{D}}_{s,t} - \mathcal{D}_{s,t})) \\ &\quad + (1/\mathcal{D}_{s,\tau})(\widehat{\mathcal{D}}_{s,t} - \mathcal{D}_{s,t})(\widehat{\mathcal{D}}_{\tau,t} - \mathcal{D}_{\tau,t}) \\ &\quad + (g(\mathcal{D}_{s,\tau}, b_n) - (1/\mathcal{D}_{s,\tau}))((\widehat{\mathcal{D}}_{s,t} - \mathcal{D}_{s,t})(\widehat{\mathcal{D}}_{\tau,t} - \mathcal{D}_{\tau,t}) \\ &\quad + \mathcal{D}_{s,t}(\widehat{\mathcal{D}}_{\tau,t} - \mathcal{D}_{\tau,t}) + \mathcal{D}_{\tau,t}(\widehat{\mathcal{D}}_{s,t} - \mathcal{D}_{s,t})) \\ &= (1/\mathcal{D}_{s,\tau})(\mathcal{D}_{s,t}(\widehat{\mathcal{D}}_{\tau,t} - \mathcal{D}_{\tau,t}) + \mathcal{D}_{\tau,t}(\widehat{\mathcal{D}}_{s,t} - \mathcal{D}_{s,t})) + O_p(1/a_n^2) \end{aligned} \tag{7.6}$$

where the last equality follows from A2, A3, and (7.4).

Now let $g_t(t, b) = \frac{\partial g(t,b)}{\partial t}$. From $b_n \to 0$ and $\widehat{\mathcal{D}}_{s,\tau} \xrightarrow{p} \mathcal{D}_{s,\tau}$, it follows that there is a $0 < b < |\mathcal{D}_{s,\tau}|$ and an $n_0$ such that for all $n > n_0$ $2b_n \leq |\mathcal{D}_{s,\tau}| - b$ and $Pr(|\bar{\mathcal{D}}_{s,\tau}| < |\mathcal{D}_{s,\tau}| - b) < \epsilon/2$ where $\bar{\mathcal{D}}_{s,\tau}$ satisfies $|\bar{\mathcal{D}}_{s,\tau} - \mathcal{D}_{s,\tau}| < |\widehat{\mathcal{D}}_{s,\tau} - \mathcal{D}_{s,\tau}|$. Also, A2 implies that $\widehat{\mathcal{D}}^2_{s,\tau} - \mathcal{D}^2_{s,\tau} = O_p(1/a_n)$ which implies $Pr(|\bar{\mathcal{D}}^2_{s,\tau} - \mathcal{D}^2_{s,\tau}| \geq \eta\mathcal{D}^2_{s,\tau}(|\mathcal{D}_{s,\tau}| - b)^2) < (\epsilon/2)(1 - \epsilon/2)$ for some $\eta > 0$ and $b$ defined above, $\epsilon > 0$ and all $n > n_1$ for some $n_1$. Combining these results, we have that

$$\begin{aligned} Pr(|g_t(\bar{\mathcal{D}}_{s,\tau}, b_n) + 1/\mathcal{D}^2_{s,\tau}| \geq \eta) &= Pr(|-1/\bar{\mathcal{D}}^2_{s,\tau} + 1/\mathcal{D}^2_{s,\tau}| \geq \eta||\bar{\mathcal{D}}_{s,\tau}| \geq |\mathcal{D}_{s,\tau}| - b)\times \\ &\qquad Pr(|\bar{\mathcal{D}}_{s,\tau}| \geq |\mathcal{D}_{s,\tau}| - b) \\ &\quad + Pr(|g_t(\bar{\mathcal{D}}_{s,\tau}, b_n) + 1/\mathcal{D}^2_{s,\tau}| \geq \eta||\bar{\mathcal{D}}_{s,\tau}| < |\mathcal{D}_{s,\tau}| - b)\times \end{aligned}$$

$$
\begin{aligned}
&Pr(|\bar{\mathcal{D}}_{s,\tau}| < |\mathcal{D}_{s,\tau}| - b) \\
&\leq Pr(|-1/\bar{\mathcal{D}}_{s,\tau}^2 + 1/\mathcal{D}_{s,\tau}^2| \geq \eta||\bar{\mathcal{D}}_{s,\tau}| \geq |\mathcal{D}_{s,\tau}| - b) + \epsilon/2 \\
&\leq Pr(|\bar{\mathcal{D}}_{s,\tau}^2 - \mathcal{D}_{s,\tau}^2| \geq \eta\mathcal{D}_{s,\tau}^2(|\mathcal{D}_{s,\tau}| - b)^2||\bar{\mathcal{D}}_{s,\tau}| \geq |\mathcal{D}_{s,\tau}| - b) + \epsilon/2 \\
&\leq Pr(|\bar{\mathcal{D}}_{s,\tau}^2 - \mathcal{D}_{s,\tau}^2| \geq \eta\mathcal{D}_{s,\tau}^2(|\mathcal{D}_{s,\tau}| - b)^2)/(1-\epsilon/2) + \epsilon/2 \\
&\leq \epsilon
\end{aligned}
$$

for all $n > \max n_0, n_1$ which implies

$$g_t(\bar{\mathcal{D}}_{s,\tau}, b_n) + 1/\mathcal{D}_{s,\tau}^2 = o_p(1). \tag{7.7}$$

Now turning to the first term in (7.5), we have that

$$
\begin{aligned}
(g(\widehat{\mathcal{D}}_{s,\tau}, b_n) - g(\mathcal{D}_{s,\tau}, b_n))\widehat{\mathcal{D}}_{s,t}\widehat{\mathcal{D}}_{\tau,t} &= -(1/\mathcal{D}_{s,\tau}^2)\mathcal{D}_{s,t}\mathcal{D}_{\tau,t}(\widehat{\mathcal{D}}_{s,\tau} - \mathcal{D}_{s,\tau}) \\
&\quad - (1/\mathcal{D}_{s,\tau}^2)(\widehat{\mathcal{D}}_{s,\tau} - \mathcal{D}_{s,\tau})(\widehat{\mathcal{D}}_{s,t}\widehat{\mathcal{D}}_{\tau,t} - \mathcal{D}_{s,t}\mathcal{D}_{\tau,t}) \\
&\quad + (g_t(\bar{\mathcal{D}}_{s,\tau}, b_n) + 1/\mathcal{D}_{s,\tau}^2)(\widehat{\mathcal{D}}_{s,\tau} - \mathcal{D}_{s,\tau})(\widehat{\mathcal{D}}_{s,t}\widehat{\mathcal{D}}_{\tau,t} - \mathcal{D}_{s,t}\mathcal{D}_{\tau,t}) \\
&\quad + (g_t(\bar{\mathcal{D}}_{s,\tau}, b_n) + 1/\mathcal{D}_{s,\tau}^2)(\widehat{\mathcal{D}}_{s,\tau} - \mathcal{D}_{s,\tau})\mathcal{D}_{s,t}\mathcal{D}_{\tau,t} \\
&= -(1/\mathcal{D}_{s,\tau}^2)\mathcal{D}_{s,t}\mathcal{D}_{\tau,t}(\widehat{\mathcal{D}}_{s,\tau} - \mathcal{D}_{s,\tau}) + o_p(1/a_n) \qquad (7.8)
\end{aligned}
$$

where the last equality follows from (7.7) and A2.

Combining (7.1)-(7.6) then yields that $\widehat{b}_t - b_t = w'(\widehat{\mathcal{D}} - \mathcal{D}) + o_p(1/a_n)$ for $w$ defined in the statement of the proposition and the conclusion follows under A2. ■

### 7.3. **Behavior of the Estimator under Nonidentification**

In this section, we provide the properties of our estimator when the model is not identified, that is when $\mathcal{D}_{s,\tau} = 0$. In this case, we strengthen the trimming condition, A3, to

A4. $g(t, b) = (1/t)1(|t| \geq 2b) + m(t, b)1(b \leq |t| < 2b)$ is two times continuously differentiable and $a_n b_n \to \infty$ as $n \to \infty$.

Using A3' in place of A3, we may state the following result.

**Proposition 3.** *Suppose the data $\{Y, X\}$ were generated according to model (2.1) and (2.4) in the text, that conditions A2 and A4 are satisfied, and that $\mathcal{D}_{s,\tau} = 0$. Then $\widehat{b}_t - \mathcal{D}_{t,t} \xrightarrow{p} 0$ and $a_n(\widehat{b}_t - \mathcal{D}_{t,t} - e_{tt}B_n) \xrightarrow{d} N(0, e_{tt}' V_x e_{tt})$ where $e_{tt}$ is a vector defined as $\mathcal{D}$ above with a one in the location corresponding to $\mathcal{D}_{t,t}$ and zeros elsewhere.*

**Proof.** We first note that under A2 and A4, we have that for any $\epsilon > 0$ there exists an $n_\epsilon$ large enough $Pr(|\widehat{\mathcal{D}}_{s,\tau}| < b_n) = Pr(|a_n\widehat{\mathcal{D}}_{s,\tau}| < a_n b_n) \geq 1 - \epsilon$ for all $n > n_\epsilon$. We also have that for $\eta > 0$ and $\epsilon > 0$, $Pr(|g(\widehat{\mathcal{D}}_{s,\tau})/a_n| > \eta) = Pr(|g(\widehat{\mathcal{D}}_{s,\tau})/a_n| > \eta||\widehat{\mathcal{D}}_{s,\tau}| \geq b_n)Pr(|\widehat{\mathcal{D}}_{s,\tau}| \geq b_n) + Pr(0 > \eta||\widehat{\mathcal{D}}_{s,\tau}| < b_n)Pr(|\widehat{\mathcal{D}}_{s,\tau}| < b_n) \leq \epsilon$ for $n > n_\epsilon$. That is, $g(\widehat{\mathcal{D}}_{s,\tau})/a_n \xrightarrow{p} 0$.

The model also implies that whenever $\mathcal{D}_{s,\tau} = 0$ at least one of $\mathcal{D}_{s,t}$ or $\mathcal{D}_{\tau,t}$ is also zero. Thus, we may right

$$\begin{aligned}
\widehat{b}_t - \mathcal{D}_{t,t} &= \widehat{\mathcal{D}}_{t,t} - \mathcal{D}_{t,t} - g(\widehat{\mathcal{D}}_{s,\tau}, b_n)(\widehat{\mathcal{D}}_{s,t}\widehat{\mathcal{D}}_{\tau,t} - \mathcal{D}_{s,t}\mathcal{D}_{\tau,t}) \\
&= \widehat{\mathcal{D}}_{t,t} - \mathcal{D}_{t,t} + o_p(1/a_n)O_p(1/a_n)
\end{aligned}$$

where the last equality follows easily under A2 and the preceding paragraph. The conclusion is then immediate under A2. ■

## References


Altonji, J. G., and R. L. Matzkin (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73(4), 1053–1102.

Arellano, M., and B. Honoré (2001): "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics. Volume 5*, ed. by J. J. Heckman, and E. Leamer. North-Holland.

Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 47, 225–238.

Chen, S., and S. Khan (2007): "Estimation of Censored Panel Data Models with Heterogeneous Factor Loads," forthcoming, Econometric Theory.

Chernozhukov, V., J. Hahn, and W. Newey (2004): "Bound Analysis in Panel Models with Correlated Random Effects," mimeo.

Gayle, G.-L., and C. Viauroux (2007): "Root-N Consistent Semiparametric Estimators of a Dynamic Panel Sample Selection Model," *Journal of Econometrics*, 141(1), 179–212.

Hahn, J., and G. Kuersteiner (2004): "Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects," Mimeo.

Hahn, J., and G. M. Kuersteiner (2002): "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both N and T Are Large," *Econometrica*, 70(4), 1639–1657.

Hahn, J., and W. K. Newey (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72(4), 1295–1319.

Hausman, J. A., and W. E. Taylor (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49(6), 1377–1398.

Henderson, D. J., and A. Ullah (2005): "A Nonparametric Random Effects Estimator," *Economics Letters*, 88(3), 403–407.

Honoré, B. E., and E. Kyriazidou (2000a): "Estimation of Tobit-Type Models with Individual Specific Effects," *Econometric Reviews*, 19(3), 341–366.

——— (2000b): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68(4), 839–874.

Honoré, B. E., and A. Lewbel (2002): "Semiparametric Binary Choice Panel Data Models Without Strictly Exogenous Regressors," *Econometrica*, 70, 2053–2063.

Honoré, B. E., and E. Tamer (2006): "Bounds on the Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74(3), 611–632.

Horowitz, J. L., and S. Lee (2004): "Semiparametric Estimation of a Panel Data Proportional Hazard Model with Fixed Effects," *Journal of Econometrics*, 119(1), 155–198.

Kyle, A. S. (1985): "Continuous Auctions and Insider Trading," *Econometrica*, 53(6), 1315–1335.

Kyriazidou, E. (1997): "Estimation of a Panel Data Sample Selection Model," *Econometrica*, 65, 1335–1364.

Lewbel, A. (2005): "Simple Endogenous Binary Choice and Selection Panel Model Estimators," mimeo.

Li, Q., and J. S. Racine (2006): *Nonparametric Econometrics: Theory and Practice*. Princeton University Press: Princeton, NJ.

Lin, X., and R. J. Carroll (2000): "Nonparametric function estimation for clustered data when the predictor is measured without/with error," *Journal of the American Statistical Association*, 95, 520–534.

Manski, C. (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55(2), 357–362.

Meulbroek, L. K. (1992): "An Empirical Analysis of Illegal Insider Trading," *Journal of Finance*, 47(5), 1661–1699.

Mundlak, Y. (1978): "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46, 69–85.

Neyman, J., and E. L. Scott (1948): "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, 16(1), 1–32.

Pagan, A., and A. Ullah (2006): *Nonparametric Econometrics*. Cambridge University Press: Cambridge, MA.

Raudenbush, S. W., and A. S. Bryk (2002): *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks: Sage Publications, second edn.

Roulstone, D. T. (2006): "Insider Trading and the Information Content of Earnings Announcements," Chicago GSB working paper.

Rozeff, M., and M. Zaman (1998): "Overreaction and Insider Trading: Evidence from Growth and Value Portfolios," *Journal of Finance*, 53, 701–716.

Ullah, A., and N. Roy (1998): "Nonparametric and Semiparametric Econometrics of Panel Data," in *Handbook of Applied Economic Statistics*, ed. by A. Ullah, and D. E. A. Giles, vol. 1. Mercel Dekker: New York, NY.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts: The MIT Press.

——— (2005): "Unobserved Heterogeneity and Estimation of Average Partial Effects," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. W. K. Andrews, and J. H. Stock. Cambridge University Press.

Woutersen, T. (2005): "Robustness against Incidental Parameters and Mixing Distributions," Mimeo.

Table 1. Descriptive Statistics for Insider Trading Data.

| Variable | Definition | Mean | SD |
|---|---|---|---|
| $BUY1$ | Indicator for insider buying in 1999Q2 | .329 | .470 |
| $BUY2$ | Indicator for insider buying in 1999Q3 | .277 | .448 |
| $BUY3$ | Indicator for insider buying in 1999Q4 | .288 | .453 |
| $SELL1$ | Indicator for insider selling in 1999Q2 | .267 | .443 |
| $SELL2$ | Indicator for insider selling in 1999Q3 | .305 | .460 |
| $SELL3$ | Indicator for insider selling in 1999Q4 | .279 | .448 |
| $TURNOVER1$ | Total volume during 1999Q2 / shares outstanding | 0.191 | 0.207 |
| $TURNOVER2$ | Total volume during 1999Q3 / shares outstanding | 0.167 | 0.184 |
| $TURNOVER3$ | Total volume during 1999Q4 / shares outstanding | 0.178 | 0.201 |
| $r_{-1}$ | Return 6 months prior to quarterly earnings announcement 1998Q4, market adjusted | -0.124 | 0.429 |
| $r_0$ | Return 6 months prior to quarterly earnings announcement 1999Q1, market adjusted | -0.107 | 0.558 |
| $r_1$ | Return 6 months prior to quarterly earnings announcement 1999Q2, market adjusted | 0.012 | 0.414 |
| $r_2$ | Return 6 months prior to quarterly earnings announcement 1999Q3, market adjusted | 0.031 | 0.449 |
| $r_3$ | Return 6 months prior to quarterly earnings announcement 1999Q4, market adjusted | 0.079 | 0.944 |
| $r_4$ | Return 6 months prior to quarterly earnings announcement 2000Q1, market adjusted | 0.126 | 0.913 |

Summary statistics and brief definitions for variables considered in our insider trading application. Our sample consists of $N = 3804$ firms over 1999Q2-1999Q4. Our dependent variables are indicators for any insider buying during a quarter or any insider selling during a quarter.

TABLE 2. Estimated Marginal Effects of Turnover on Insider Buying of Own Company Stock

| | $b_{11}$ | $b_{12}$ | $b_{21}$ | $b_{22}$ | $b_{31}$ | $b_{32}$ | $\bar{b}_1$ | $\bar{b}_2$ | $\bar{b}_3$ | $\bar{b}$ | Over |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A. Turnover Only | | | | | | | | | | | |
| I. Marginal Effects | | | | | | | | | | | |
| estimate | 0.421 | 0.387 | 0.221 | 0.241 | 0.383 | 0.334 | 0.410 | 0.227 | 0.348 | 0.328 | |
| s.e. | 0.170 | 0.218 | 0.119 | 0.146 | 0.294 | 0.191 | 0.151 | 0.111 | 0.161 | 0.061 | |
| $c_{80}$ | 1.145 | 1.075 | 1.214 | 1.104 | 1.097 | 1.115 | 1.036 | 1.198 | 0.884 | 0.936 | |
| $c_{90}$ | 1.550 | 1.471 | 1.562 | 1.404 | 1.734 | 1.760 | 1.431 | 1.503 | 1.403 | 1.285 | |
| $c_{95}$ | 2.036 | 1.871 | 1.953 | 1.794 | 2.315 | 2.309 | 1.875 | 1.820 | 1.802 | 1.706 | |
| II. First-Stage | | | | | | | | | | | |
| estimate | -0.223 | -0.144 | -0.387 | -0.148 | -0.179 | -0.126 | | | | | 0.021 |
| s.e. | 0.069 | 0.076 | 0.101 | 0.058 | 0.120 | 0.067 | | | | | |
| $c_{80}$ | 1.265 | 1.236 | 1.220 | 1.306 | 1.314 | 1.292 | | | | | 1.438 |
| $c_{90}$ | 1.609 | 1.617 | 1.576 | 1.697 | 1.639 | 1.605 | | | | | 2.000 |
| $c_{95}$ | 1.925 | 1.921 | 1.903 | 2.001 | 2.013 | 1.929 | | | | | 2.627 |
| B. Turnover and Returns | | | | | | | | | | | |
| I. Marginal Effects | | | | | | | | | | | |
| estimate | 0.211 | 0.674 | 0.233 | 0.137 | 0.234 | 0.586 | 0.211 | 0.146 | 0.286 | 0.215 | |
| s.e. | 0.082 | 1.001 | 0.111 | 0.077 | 0.144 | 0.344 | 0.082 | 0.077 | 0.133 | 0.057 | |
| $c_{80}$ | 1.138 | 2.405 | 1.120 | 1.185 | 0.936 | 1.348 | 1.090 | 1.208 | 0.842 | 0.913 | |
| $c_{90}$ | 1.488 | 3.579 | 1.551 | 1.515 | 1.232 | 2.131 | 1.461 | 1.520 | 1.124 | 1.213 | |
| $c_{95}$ | 1.850 | 4.643 | 1.965 | 1.815 | 1.469 | 2.873 | 1.784 | 1.793 | 1.356 | 1.513 | |
| II. First-Stage | | | | | | | | | | | |
| estimate | -0.184 | -0.040 | -0.196 | -0.180 | -0.121 | -0.080 | | | | | 1.334 |
| s.e. | 0.059 | 0.069 | 0.060 | 0.055 | 0.069 | 0.057 | | | | | |
| $c_{80}$ | 1.288 | 1.283 | 1.342 | 1.224 | 1.313 | 1.310 | | | | | 1.449 |
| $c_{90}$ | 1.633 | 1.692 | 1.699 | 1.650 | 1.679 | 1.637 | | | | | 2.417 |
| $c_{95}$ | 1.953 | 2.009 | 2.027 | 2.006 | 2.013 | 2.033 | | | | | 3.379 |

Estimates of marginal effect of turnover on insider purchases of own-company stock. The dependent variable is an indicator which is one if there were any insider purchases of own company stock in a quarter. Panel A reports results from estimation of the marginal effect of turnover on the probability of insider buys; Panel B reports results from estimation of the marginal effect of turnover on the probability of insider buys controlling for returns on own company stock. $b_{t1}$ and $b_{t2}$ for $t \in \{1, 2, 3\}$ are two estimates of the marginal effect at time $t$. $\bar{b}_t$ is a variance weighted average of $b_{t1}$ and $b_{t2}$, and $\bar{b}$ is a simple average of the $\bar{b}_t$. Standard error estimates are given below point estimates in the row labeled s.e. $c_{80}$, $c_{90}$, and $c_{95}$ are bootstrap critical values for t-statistics based off 1000 bootstrap replication.

TABLE 3. Estimated Marginal Effects of Turnover on Insider Selling of Own Company Stock

| | $b_{11}$ | $b_{12}$ | $b_{21}$ | $b_{22}$ | $b_{31}$ | $b_{32}$ | $\bar{b}_1$ | $\bar{b}_2$ | $\bar{b}_3$ | $\bar{b}$ | Over |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | A. Turnover Only | | | | | | |
| | | | | | I. Marginal Effects | | | | | | |
| estimate | 0.006 | 0.261 | 0.222 | 0.343 | -0.046 | 3.612 | 0.170 | 0.311 | -0.041 | 0.147 | |
| s.e. | 0.183 | 0.141 | 0.126 | 0.104 | 0.512 | 13.931 | 0.119 | 0.100 | 0.511 | 0.194 | |
| $c_{80}$ | 1.155 | 1.062 | 1.332 | 1.159 | 1.170 | 6.820 | 1.059 | 1.167 | 1.060 | 0.822 | |
| $c_{90}$ | 1.505 | 1.415 | 1.652 | 1.467 | 1.761 | 10.330 | 1.393 | 1.528 | 1.696 | 1.125 | |
| $c_{95}$ | 1.859 | 1.724 | 1.922 | 1.720 | 2.384 | 13.806 | 1.683 | 1.834 | 2.227 | 1.517 | |
| | | | | | II. First-Stage | | | | | | |
| estimate | 0.273 | 0.223 | 0.569 | -0.226 | 0.210 | 0.020 | | | | | 1.549 |
| s.e. | 0.092 | 0.113 | 0.120 | 0.090 | 0.140 | 0.087 | | | | | |
| $c_{80}$ | 1.354 | 1.259 | 1.344 | 1.247 | 1.335 | 1.293 | | | | | 1.348 |
| $c_{90}$ | 1.810 | 1.674 | 1.734 | 1.631 | 1.694 | 1.612 | | | | | 2.287 |
| $c_{95}$ | 2.101 | 1.994 | 2.175 | 1.876 | 2.051 | 1.956 | | | | | 3.358 |
| | | | | | B. Turnover and Returns | | | | | | |
| | | | | | I. Marginal Effects | | | | | | |
| estimate | -0.008 | 0.342 | 0.086 | 0.155 | 0.297 | 1.153 | 0.007 | 0.135 | 0.336 | 0.159 | |
| s.e. | 0.119 | 0.669 | 0.106 | 0.092 | 0.340 | 1.885 | 0.116 | 0.089 | 0.328 | 0.124 | |
| $c_{80}$ | 1.154 | 1.827 | 1.240 | 1.040 | 0.746 | 3.329 | 1.014 | 1.071 | 0.667 | 0.618 | |
| $c_{90}$ | 1.468 | 2.442 | 1.567 | 1.323 | 0.943 | 4.932 | 1.360 | 1.401 | 0.882 | 0.861 | |
| $c_{95}$ | 1.781 | 2.986 | 1.795 | 1.632 | 1.205 | 6.275 | 1.624 | 1.731 | 1.035 | 1.038 | |
| | | | | | II. First-Stage | | | | | | |
| estimate | 0.200 | 0.036 | 0.301 | -0.121 | 0.089 | 0.033 | | | | | 0.646 |
| s.e. | 0.078 | 0.088 | 0.090 | 0.068 | 0.102 | 0.072 | | | | | |
| $c_{80}$ | 1.279 | 1.307 | 1.342 | 1.283 | 1.319 | 1.333 | | | | | 1.316 |
| $c_{90}$ | 1.637 | 1.685 | 1.740 | 1.638 | 1.727 | 1.678 | | | | | 1.971 |
| $c_{95}$ | 1.949 | 1.982 | 2.037 | 2.003 | 2.048 | 2.072 | | | | | 2.622 |

Estimates of marginal effect of turnover on insider sells of own-company stock. The dependent variable is an indicator which is one if there were any insider sells of own company stock in a quarter. Panel A reports results from estimation of the marginal effect of turnover on the probability of insider sells; Panel B reports results from estimation of the marginal effect of turnover on the probability of insider sells controlling for returns on own company stock. $b_{t1}$ and $b_{t2}$ for $t \in \{1, 2, 3\}$ are two estimates of the marginal effect at time $t$. $\bar{b}_t$ is a variance weighted average of $b_{t1}$ and $b_{t2}$, and $\bar{b}$ is a simple average of the $\bar{b}_t$. Standard error estimates are given below point estimates in the row labeled s.e. $c_{80}$, $c_{90}$, and $c_{95}$ are bootstrap critical values for t-statistics based off 1000 bootstrap replication.