cemmap

centre for microdata methods and practice

Nonparametric identification in panels using quantiles

Victor Chernozhukov
Ivan Fernández-Vál
Stefan Hoderlein
Hajo Holzmann
Whitney Newey

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP66/13

Nonparametric Identification in Panels using Quantiles *

Victor Chernozhukov, Ivan Fernandez-Val, Stefan Hoderlein,
Hajo Holzmann, Whitney Newey

December 17, 2013

Abstract

This paper considers identification and estimation of ceteris paribus effects of continuous regressors in nonseparable panel models with time homogeneity. The effects of interest are derivatives of the average and quantile structural functions of the model. We find that these derivatives are identified with two time periods for "stayers", i.e. for individuals with the same regressor values in two time periods. We show that the identification results carry over to models that allow location and scale time effects. We propose nonparametric series methods and a weighted bootstrap scheme to estimate and make inference on the identified effects. The bootstrap proposed allows uniform inference for function-valued parameters such as quantile effects over a region of quantiles or regressor values. An empirical application to Engel curve estimation with panel data illustrates the results.

Keywords: Panel data, nonseparable model, average effect, quantile effect, Engel curve

1 Identification for Panel Regression

A frequent object of interest is the ceteris paribus effect of x on y, when observed x is an individual choice variable partly determined by preferences or technology. Panel data holds out the hope of controlling for individual preferences or technology by using multiple observations for a single economic agent. This hope is particularly difficult to realize with discrete or other nonseparable models and/or multidimensional individual effects. These models are, by nature, not additively separable in unobserved individual effects, making them challenging to identify and estimate.

*We thank participants at Demand Estimation and Modelling Conference and Arthur Lewbel for comments. We gratefully acknowledge research support from the NSF.

A fundamental idea for using panel data to identify the ceteris paribus effect of x on y is to use changes in x over time. In order for changes over time in x to correspond to ceteris paribus effects, the distribution of variables other than x must not vary over time. This restriction is like "time being randomly assigned" or "time is an instrument." In this paper we consider identification via such time homogeneity conditions. They are also the basis of many previous panel results, including Chamberlain (1982), Manski (1987), and Honore (1992). Recently time homogeneity has been used as the basis for identification and estimation of nonseparable models by Chernozhukov, Fernandez-Val, Hahn, Newey (2013), Evdokimov (2010), Graham and Powell (2012), and Hoderlein and White (2012). Because economic data often exhibits drift over time, we also allow for some time effects, while maintaining underlying time homogeneity conditions.

Here we consider the identifying power of time homogeneity for quantile and average effects with continuous regressors. The effects of interest are derivatives of the quantile and average structural functions of the model. We find that these derivatives are identified with two time periods for "stayers", i.e. conditional on x being equal in two time periods. Time homogeneity is too strong for many econometric applications where time trends are evident in the data. We weaken homogeneity by allowing for location and scale time effects. Allowing for such time effects makes identification and estimation more complicated but more widely applicable.

Quantile identification under time homogeneity is based on differences of quantiles. It is also interesting to consider whether quantiles of differences can help identify effects of interest. We find that they can under conditions that are substantially stronger than time homogeneity. We give quantile difference identification results that restrict the distribution of individual effects conditional on x, similarly to Chamberlain (1980), Altonji and Matzkin (2005), and Bester and Hansen (2009). In our opinion these added restrictions make quantiles of differences less appealing. We therefore focus for the rest of the paper, including the application, on differences of quantiles.

To illustrate we provide an application to Engel curve estimation. The Engel curve describes how demand changes with expenditure. We use data from the 2007 and 2009 waves of the Panel Study of Income Dynamics (PSID). Endogeneity in the estimation of Engel curves arises because the decision to consume a commodity may occur simultaneously with the allocation of income between consumption and savings. In contrast with the previous cross sectional literature, we do not rely on a two-stage budgeting argument that justifies the use of labor income as an instrument for expenditure. Instead, we assume that the Engel curve relationships are time homogeneous up to location and scale time effects, which leads to identification of structural effects from panel data.

An alternative approach to identification in panel data is to impose restrictions on the conditional distribution of the individual effect given x. This approach leads to nonparametric generalizations of Chamberlain's (1980) correlated random effects model. As shown by Chamberlain (1984), Altonji and Matzkin (2005), Bester and Hansen (2009), and others, this kind of condition leads to identification of various effects. In particular, Altonji and Matzkin (2005) show identification of an average derivative conditional on the regressor equal to a specific value, an effect they call the local average response (LAR). In this paper we take a different approach, preferring to impose time homogeneity rather than restrict the relationship between observed regressors and unobserved individual effects.

Section 2 describes the model and gives an average derivative result. Section 3 gives the quantile identification result that follows from time homogeneity. Section 4 considers how quantiles of differences can be used to identify the effect of x on y. Section 5 explains how we allow for time effects. Estimation and inference are briefly discussed in Section 6, and the empirical example is given in Section 7.

2 The Model and Conditional Mean Effects

The data consist of n observations on $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iT})'$ and $\boldsymbol{X}_i = [X'_{i1}, \ldots, X'_{iT}]'$, for a dependent variable Y_{it} and a vector of regressors X_{it}. Throughout we assume that the observations $(\boldsymbol{Y}_i, \boldsymbol{X}_i)$, $(i = 1, \ldots, n)$, are independent and identically distributed. The nonparametric models we consider satisfy

Assumption 1. *There is a function ϕ and vectors of random variables A_i and V_{it} such that*

$$Y_{it} = \phi(X_{it}, A_i, V_{it}), \qquad i = 1, \ldots, n, \quad t = 1, 2, \ldots, T.$$

We focus in this paper on the two time period case, $T = 2$, though it is straightforward to extend the results to many time periods. The vector A_i consists of time invariant individual effects that often represent individual heterogeneity. The vector V_{it} represents period specific disturbances. Altonji and Matzkin (2005) considered models satisfying Assumption 1. As discussed in Chernozhukov et. al. (2013), the invariance of ϕ over time in this Assumption does not actually impose any time homogeneity. If there are no restrictions on V_{it} then t could be one of the components of V_{it}, allowing the function to vary over time in a completely general way. The next condition together with Assumption 1 imposes time homogeneity on the model.

Assumption 2. $V_{it} | \boldsymbol{X}_i, A_i \overset{d}{=} V_{i1} | \boldsymbol{X}_i, A_i$ *for all t.*

This is a static, or "strictly exogenous" time homogeneity condition, where all leads and lags of the regressor are included in the conditioning variable $\boldsymbol{X}_i$. It requires that the conditional distribution of V_{it} given $\boldsymbol{X}_i$ and A_i does not depend on t, but does allow for dependence of V_{it} over time.

Setting $U_{it} = (A_i', V_{it}')'$, an equivalent condition is

$$U_{it}|\boldsymbol{X}_i \overset{d}{=} U_{i1}|\boldsymbol{X}_i.$$

Thus, the time invariant A_i has no distinct role in this model. As further discussed in Chernozhukov et. al. (2013), this seems a basic condition that helps panel data provide information about the effect of x on y. It is like the time period being "randomly assigned" or "time is an instrument," with the distribution of factors other than x not varying over time, so that changes in x over time can help identify the effect of x on y.

Although they seem useful for nonlinear models, the time homogeneity conditions are strong. In particular they do not allow for heteroskedasticity over time, which is often thought to be important in applications. We partially address this problem below by allowing for location and scale time effect.

For notational convenience we shall drop the i subscript and let $T = 2$ in the following. Our focus in this paper is on the case where the regressors $\boldsymbol{X}$ are continuously distributed. We will be interested in several effects of $\boldsymbol{X}$ on $\boldsymbol{Y}$. For $u = (a', v')'$ we let $\phi(x, u) = \phi(x, a, v)$. We will let x or x_t denote a possible value of the regressor vector X_t and $\boldsymbol{x} = (x_1', x_2')'$ a possible value of $\boldsymbol{X} = (X_1', X_2')'$. Let $\partial_x \phi(x, u)$ denote the vector of partial derivatives of ϕ w.r.t. the coordinates of x. One effect we consider is a conditional expectation of the derivative $\partial_x \phi(X_t, U_t)$ given by

$$E\big[\partial_x \phi(x, U_t)|X_1 = X_2 = x\big].$$

This is the object considered in Hoderlein and White (2012) and is similar to the local average response considered in Altonji and Matzkin (2005). It gives the local marginal effect for individuals with regressor value x in both periods. This effect is related to the conditional average structural function (CASF):

$$m(x \mid \boldsymbol{x}) = E[\phi(x, U_t) \mid \boldsymbol{X} = \boldsymbol{x}],$$

through

$$\partial_x m(x \mid \boldsymbol{x})\Big|_{\boldsymbol{x}=(x,x)} = E\big[\partial_x \phi(x, U_t)|X_1 = X_2 = x\big],$$

under the conditions that permit interchanging the derivative and expectation.

The other effects we consider are similar to this effect except that we also condition on certain values of Y_t. One of these is given by

$$E\big[\partial_x \phi(x, U_t)|Y_t = q(\tau, x), X_1 = X_2 = x\big],$$

where $q(\tau, x)$ is the τ^{th} conditional quantile of $\phi(x, U_t)$ given $X_1 = X_2 = x$. This is a quantile derivative effect, similar to the local average structural derivative in Hoderlein and Mammen (2007). It gives the local marginal effect for individuals with regressor value x in both periods and at the quantile $q(\tau, x)$. This effect is also related to the conditional quantile structural function (CQSF), $q_\tau(x|\boldsymbol{x})$, that gives the τ-quantile of $\phi(x, U_t)$ conditional on $\boldsymbol{X} = \boldsymbol{x}$, through

$$\partial_x q_\tau(x \mid \boldsymbol{x})\Big|_{\boldsymbol{x}=(x,x)} = E\big[\partial_x \phi(x, U_t)|Y_t = q(\tau, x), X_1 = X_2 = x\big].$$

We also consider linking quantiles of arbitrary linear combinations of the dependent variables Y_1 and Y_2 to conditional expectations of the form

$$E\big(\partial_x \phi(x, U_t) \,|\, \text{linear comb of } \boldsymbol{Y}, X_1 = X_2 = x\big).$$

These are dependent variable conditioned average effects. One intended direction is to compare the derivative of the quantiles of the differences $Y_2 - Y_1$ to the differences of the derivative of the quantiles of Y_2 and Y_1 in terms of objects they identify. This note carries out that comparison.

To set the stage for the quantile results we first discuss mean identification. We first give an explanation of identification of the mean effect and then give a precise result with regularity conditions.

Consider the identified conditional mean

$$M_t(\boldsymbol{x}) = E(Y_t|\boldsymbol{X} = \boldsymbol{x}), \qquad t = 1, 2.$$

Together these conditional expectations are a nonparametric version of Chamberlain's (1982) multivariate regression model for panel data. Derivatives of them can be combined to identify the conditional mean effect. Let $f(u|\boldsymbol{x})$ denote the conditional density of U_t given $\boldsymbol{X} = \boldsymbol{x}$, that does not depend on t by Assumption 2. Assume that $\phi(x, u)$ and $f(u|\boldsymbol{x})$ are differentiable in x and $\boldsymbol{x}$ respectively and that differentiation under the integral is permitted. For $\boldsymbol{x} = (x_1', x_2')'$ we let $\partial_{x_s} M_t(\boldsymbol{x})$ and $\partial_{x_s} f(u|\boldsymbol{x})$, $s, t = 1, 2$, denote the vector of partial derivatives w.r.t. the coordinates of x_s. Then for $s, t = 1, 2$,

$$\begin{aligned}\partial_{x_s} M_t(\boldsymbol{x}) &= \partial_{x_s} E(Y_t|\boldsymbol{X} = \boldsymbol{x}] = \partial_{x_s} \int \phi(x_t, u) f(u|\boldsymbol{x}) du \\ &= 1(s = t) \int \partial_x \phi(x_t, u) f(u|\boldsymbol{x}) du + \int \phi(x_t, u) \partial_{x_s} f(u|\boldsymbol{x}) du,\end{aligned}$$

where the first term is the conditional mean effect of interest and the second term is the analog to Chamberlain's (1982) heterogeneity bias. Subtracting and using Assumption 2 gives

$$\partial_{x_2} M_2(\boldsymbol{x}) - \partial_{x_2} M_1(\boldsymbol{x}) = E\big(\partial_x \phi(x_2, U_t)|\boldsymbol{X} = \boldsymbol{x}\big) + \int \big(\phi(x_2, u) - \phi(x_1, u)\big) \partial_{x_2} f(u|\boldsymbol{x}) du.$$

Evaluating at $\boldsymbol{x} = (x', x')'$ we find that

$$E\big(\partial_x \phi(x, U_t)|X_1 = X_2 = x\big) = \partial_{x_2} M_2(x, x) - \partial_{x_2} M_1(x, x) = \partial_{x_2} \Delta M(x, x) \tag{2.1}$$

where

$$\Delta M(\boldsymbol{x}) = E(Y_2 - Y_1|\boldsymbol{X} = \boldsymbol{x}).$$

It also follows similarly that

$$\begin{aligned} E\big(\partial_x \phi(x, U_t)|X_1 = X_2 = x\big) &= -\partial_{x_1}\Delta M(x, x) \\ &= \partial_{x_1} E(Y_1 - Y_2|X_1 = x_1, X_2 = x_2)\big|_{(x_1', x_2') = (x', x')}. \end{aligned} \tag{2.2}$$

Thus, the conditional mean effect is identified from the derivative of the conditional expectation of the difference with respect to the leading time period for individuals where X_t is the same in both periods. We note here that this means the conditional mean effect is overidentified. Introducting time effects, as we do below, will lead to exact identification.

The following result makes the previous argument precise, including conditions for differentiating under integrals.

Theorem 1. *Suppose that Assumptions 1 and 2 are satisfied, $E|Y_t| < \infty$, $t = 1, 2$, and that $\phi(x, u)$ (where $u' = (a', v')$) resp. the conditional density $f(u|\boldsymbol{x})$ of $U_t = (A', V_t')'$ given $\boldsymbol{X} = \boldsymbol{x}$ are continuously differentiable in x resp. $\boldsymbol{x}$ for fixed u. Given x, suppose that for some $\varepsilon > 0$,*

$$\int \sup_{\|\delta\| \le \varepsilon, \delta = (\delta_0', \delta_1', \delta_2')'} \big\| \partial_x \phi(x + \delta_0, u)\, f(u|x + \delta_1, x + \delta_2) \big\| du < \infty,$$

$$\int \sup_{\|\delta\| \le \varepsilon, \delta = (\delta_0', \delta_1', \delta_2')'} \big\| \phi(x + \delta_0, u)\, \partial_{x_s} f(u|x + \delta_1, x + \delta_2) \big\| du < \infty, \quad s = 1, 2,$$

then (2.1) and (2.2) hold true.

Proof. It follows from the differentiability of $\phi(x, u)$ and $f(u|x_1, x_2)$ and the dominance condition that

$$\tilde{M}(x, x_1, x_2) := \int \phi(x, u) f(u|x_1, x_2) du$$

is continuously differentiable in $(x', x_1', x_2')'$ in a neighborhood of (x', x', x'), and that the order of differentiation and integration can be interchanged. Furthermore, by the structure of the model and Assumption 2, for $\boldsymbol{x} = (x_1', x_2')'$,

$$M_t(\boldsymbol{x}) = E\big(\phi(X_t, U_t)|X_1 = x_1, X_2 = x_2\big) = \tilde{M}(x_t, x_1, x_2).$$

Therefore, $M_t(\boldsymbol{x})$ is continuously differentiable in $\boldsymbol{x}$, $t = 1, 2$, in a neighborhood of $(x', x')'$, and for $s = 1, 2$,

$$\begin{aligned} \partial_{x_s} M_t(\boldsymbol{x}) &= \big(1(s = t)\partial_x \tilde{M}(x, x_1, x_2) + \partial_{x_s}\tilde{M}(x, x_1, x_2)\big)\big|_{(x, x_1, x_2) = (x_t, x_1, x_2)} \\ &= \int \big(1(s = t)\partial_x \phi(x_t, u) f(u|\boldsymbol{x}) + \phi(x_t, u)\partial_{x_s} f(u|\boldsymbol{x})\big) du \\ &= E\Big(\big(1(s = t)\partial_x \phi(x_t, U_t) + \phi(x_t, U_t) h_s(U_t|\boldsymbol{x})\big)\big|\boldsymbol{X} = \boldsymbol{x}\Big), \end{aligned} \tag{2.3}$$

where $h_s(u|x) = f(u|\boldsymbol{x})^{-1}\partial_{x_s} f(u|\boldsymbol{x})$. Subtracting and using $U_1|\boldsymbol{X} \overset{d}{=} U_2|\boldsymbol{X}$,

$$\begin{aligned}\partial_{x_2} M_2(\boldsymbol{x}) - \partial_{x_2} M_1(\boldsymbol{x}) = E\big(\partial_x \phi(x_2, U_2)\big|\boldsymbol{X} = \boldsymbol{x}\big) \\ + E\Big(\big(\phi(x_2, U_2) - \phi(x_1, U_2)\big)\, h_2(U_2|\boldsymbol{x})\big)\big|\boldsymbol{X} = \boldsymbol{x}\Big).\end{aligned}$$

Evaluating at $\boldsymbol{x} = (x', x')'$ gives (2.1), and (2.2) follows similarly by considering $\partial_{x_1} M_1(\boldsymbol{x}) - \partial_{x_1} M_2(\boldsymbol{x})$, using (2.3) and evaluating at $\boldsymbol{x} = (x', x')'$. □

This result has slightly weaker conditions than that of Hoderlein and White (2012). Here we drop their assumption that V_t is independent of X_1 conditional on A. The result given here allows for X_1 to be correlated with (V_1, V_2), as long as the marginal distribution of V_t conditional on (X_1, X_2, A) does not vary with t. We maintain these weaker conditions as we consider identification of quantile effects in the next Section.

3 Conditional Quantile Effects

Turning now to the identification of the quantile effects given above, let $Q_t(\tau \mid \boldsymbol{x})$ denote the τ^{th} conditional quantile of Y_t conditional on $\boldsymbol{X} = \boldsymbol{x} = (x_1', x_2')'$. It will be the solution to

$$\int 1(\phi(x_t, u) \leq Q_t(\tau \mid \boldsymbol{x}))f(u|\boldsymbol{x})du = \tau.$$

The pair $[Q_1(\tau \mid \boldsymbol{x}), Q_2(\tau \mid \boldsymbol{x})]$ is a quantile analog of Chamberlain's (1982) multivariate regression for panel data. We can identify a quantile analog of the Hoderlein and White (2012) average derivative effect. We first describe how these multivariate panel quantiles can be used to identify an average derivative effect then give a precise interpretation of the effect. This description helps explain the source of identification as well as the precise nature of the identified effect.

To describe how identification works, differentiate both sides of the previous identity with respect to x_s, treat the derivative of an indicator function as a dirac delta, and assume the order of differentiation and integration can be interchanged. This calculation gives

$$\begin{aligned}0 = \int_{\phi(x_t,u)=Q_t(\tau|\boldsymbol{x})} \big(\partial_{x_s} Q_t(\tau|\boldsymbol{x}) - 1(s = t)\partial_x \phi(x_t, u)\big) f(u|\boldsymbol{x})du \\ + \int 1\big(\phi(x_t, u) \leq Q_t(\tau \mid \boldsymbol{x})\big)\partial_{x_s} f(u|\boldsymbol{x})du.\end{aligned}$$

Let $g_t(\tau \mid \boldsymbol{x}) = \int_{\phi(x_t,u)=Q_t(\tau|\boldsymbol{x})} f(u|\boldsymbol{x})du$ and note that

$$g_t(\tau \mid \boldsymbol{x})^{-1} \int_{\phi(x_t,u)=Q_t(\tau|\boldsymbol{x})} \partial_x \phi(x_t, u) f(u|\boldsymbol{x})du = E\big(\partial_x \phi(x_t, U_t)|\phi(x_t, U_t) = Q_t(\tau \mid \boldsymbol{x}), \boldsymbol{X} = \boldsymbol{x}\big)$$

Solving for $\partial_{x_s} Q_t(\tau|\boldsymbol{x})$ we find that,

$$\begin{aligned}\partial_{x_s} Q_t(\tau|\boldsymbol{x}) &= 1(s=t)E\big(\partial_x \phi(x_t,U_t)|\phi(x_t,U_t) = Q_t(\tau \mid \boldsymbol{x}), \boldsymbol{X}=\boldsymbol{x}\big)\\ &\quad - g_t(\tau \mid \boldsymbol{x})^{-1}\int 1(\phi(x_t,u) \leq Q_t(\tau \mid \boldsymbol{x}))\partial_{x_s} f(u|\boldsymbol{x})du.\end{aligned}$$

Note that at $X_1 = X_2 = x$, $Q_1(\tau \mid x,x) = Q_2(\tau \mid x,x) = q(\tau,x)$ and $g_1(\tau \mid x,x) = g_2(\tau \mid x,x)$ by time homogeneity. Then differencing the conditional quantile derivatives gives

$$\begin{aligned}\partial_{x_2} Q_2(\tau|x,x) - \partial_{x_2} Q_1(\tau|x,x) &= \partial_{x_1} Q_1(\tau|x,x) - \partial_{x_1} Q_2(\tau|x,x)\\ &= E\big(\partial_x \phi(x,U_t)|\phi(x,U_t) = q(\tau,x), X_1 = X_2 = x\big),\end{aligned} \tag{3.1}$$

where the last term does not depend on t due to time homogeneity. The equation (3.1) is a panel version of the Hoderlein and Mammen (2007) identification result. It is interesting to note that, unlike in the mean case, differences of derivatives of quantiles generally differ from derivatives of quantiles of differences. Below we will consider identification from derivatives of quantiles of differences.

To make the above derivation precise we need to formulate conditions that allow differentiation under the integral. The following regularity condition is one approach to this, in particular for the dirac delta argument given above.

Assumption 3. *We can write $u = (h', e)'$ for scalar e , such that $\phi(x,u) = \phi(x,h,e)$ is continuously differentiable in x and e and there is $C > 0$ with $\partial_e \phi(x,h,e) \geq 1/C$ and $\|\partial_x \phi(x,u)\| \leq C$ everywhere. For the corresponding representation of the random vector $U_t = (H_t', E_t)$, E_t is continuously distributed given $(H_t, \boldsymbol{X})$, with conditional pdf $f_E(e|h,\boldsymbol{x})$ that is bounded and continuous in $(e,\boldsymbol{x})$, and $f(h|\boldsymbol{x})$, the conditional pdf of H given $\boldsymbol{X} = \boldsymbol{x}$, is continuous in $\boldsymbol{x}$. Moreover, given $\boldsymbol{x}$ there is a $\delta > 0$ such that*

$$\int \sup_{\|\Delta_{\boldsymbol{x}}\| \leq \delta} f(h|\boldsymbol{x} + \Delta_{\boldsymbol{x}})dh < \infty. \tag{3.2}$$

The boundedness conditions on the derivatives of $\phi(x,u)$ could further be weakened at the expense of much more complicated notation and conditions.

For fixed x let $f_{Y_x|\boldsymbol{X}}(y|\boldsymbol{x})$ denote the conditional pdf of $Y_x = \phi(x,U_t)$ given $\boldsymbol{X} = \boldsymbol{x} = (x_1', x_2')'$. The following lemma shows differentiability of $\mathbb{P}(\phi(x,U_t) \leq y|\boldsymbol{X} = \boldsymbol{x})$ with respect to x and y for given $\boldsymbol{x}$, and computes the derivatives.

Lemma 1. *If Assumption 3 is satisfied then for fixed $\boldsymbol{x}$, $\mathbb{P}(\phi(x,U_t) \leq y|\boldsymbol{X} = \boldsymbol{x})$ is differentiable in y and x with derivatives continuous in y, x and $\boldsymbol{x}$ given by*

$$\begin{aligned}&\partial_y \mathbb{P}(\phi(x,U_t) \leq y|\boldsymbol{X} = \boldsymbol{x}) = f_{Y_x|\boldsymbol{X}}(y|\boldsymbol{x}),\\ &\partial_x \mathbb{P}(\phi(x,U_t) \leq y|\boldsymbol{X} = \boldsymbol{x}) = -f_{Y_x|\boldsymbol{X}}(y|\boldsymbol{x})\, E\big(\partial_x \phi(x,U_t)|Y_x = y, \boldsymbol{X} = \boldsymbol{x}\big),\end{aligned}$$

where $Y_x = \phi(x,U_t)$.

Proof. Let $F_E(e|h,\boldsymbol{x}) = \mathbb{P}(E \le e|H = h, \boldsymbol{X} = \boldsymbol{x}) = \int_{-\infty}^{e} f_E(\epsilon|h,\boldsymbol{x})\,d\epsilon$. Then by the fundamental theorem of calculus, $F_E(e|h,\boldsymbol{x})$ is differentiable in e with derivative $f_E(e|h,\boldsymbol{x})$ that is continuous in e and $\boldsymbol{x}$. Consider

$$\begin{aligned}
\mathbb{P}(\phi(x,U_t) \le y|\boldsymbol{X} = \boldsymbol{x}) &= \int 1(\phi(x,u) \le y)\, f(u|\boldsymbol{x})\,du \\
&= \int\int 1(\phi(x,h,e) \le y)\, f_E(e|h,\boldsymbol{x})\, f(h|\boldsymbol{x}) de\, dh \\
&= \int\int 1(e \le \phi^{-1}(x,h,y))\, f_E(e|h,\boldsymbol{x})\, f(h|\boldsymbol{x}) de\, dh \\
&= \int F_E(\phi^{-1}(x,h,y)|h,\boldsymbol{x}) f(h|\boldsymbol{x})\, dh.
\end{aligned} \tag{3.3}$$

By the inverse and implicit function theorems, $\phi^{-1}(x,h,y)$ is continuously differentiable in x and y, with

$$\begin{aligned}
\partial_y \phi^{-1}(x,h,y) &= \left[\partial_e \phi(x,h,\phi^{-1}(x,h,y))\right]^{-1}, \\
\partial_x \phi^{-1}(x,h,y) &= -\frac{\partial_x \phi(x,h,\phi^{-1}(x,h,y))}{\partial_e \phi(x,h,\phi^{-1}(x,h,y))} \\
&= -\partial_x \phi(x,h,\phi^{-1}(x,h,y))\, \partial_y \phi^{-1}(x,h,y).
\end{aligned}$$

Then by Assumption 3 both $\partial_y \phi^{-1}(x,h,y)$ and $\partial_x \phi^{-1}(x,h,y)$ are continuous in y and x and bounded. Therefore,

$$\begin{aligned}
\partial_y F_E(\phi^{-1}(x,h,y)|h,\boldsymbol{x}) &= f_E(\phi^{-1}(x,h,y)|h,\boldsymbol{x})\, \partial_y \phi^{-1}(x,h,y) \\
&= f_{Y_x|\boldsymbol{X},H_t}(y|\boldsymbol{x},h), \\
\partial_x F_E(\phi^{-1}(x,h,y)|h,\boldsymbol{x}) &= f_E(\phi^{-1}(x,h,y)|h,\boldsymbol{x})\, \partial_x \phi^{-1}(x,h,y) \\
&= -f_{Y_x|\boldsymbol{X},H_t}(y|\boldsymbol{x},h)\, \partial_x \phi(x,h,\phi^{-1}(x,h,y)),
\end{aligned} \tag{3.4}$$

are both bounded and continuous in y, x and $\boldsymbol{x}$, where the last equality in each equation follows by a standard change of variables argument. From the boundedness assumptions on f_E and on $\partial_x \phi$ in Assumption 3, it follows that $\mathbb{P}(\phi(x,U_t) \le y|\boldsymbol{X} = \boldsymbol{x})$ is partially differentiable in y and x with partial derivatives continuous in $y, x, \boldsymbol{x}$, which can be computed by differentiating under the integral in (3.3). In order to establish the expressions in the lemma, insert (3.4) into the partial derivatives of (3.3) w.r.t. y and x, and note that $f_{Y_x|\boldsymbol{X},H_t}(y|\boldsymbol{x},h) f(h|\boldsymbol{x}) = f_{Y_x,H_t|\boldsymbol{X}}(y,h|\boldsymbol{x})$. The first expression is then immediate. For the second, note that given $Y_x = y$ (for a fixed x), $E_t = \phi^{-1}(x,H_t,y)$, so that

$$\begin{aligned}
& f_{Y_x|\boldsymbol{X}}(y|\boldsymbol{x})\, E\big(\partial_x \phi(x,U_t)|Y_x = y, \boldsymbol{X} = \boldsymbol{x}\big) \\
=& f_{Y_x|\boldsymbol{X}}(y|\boldsymbol{x}) \int \partial_x \phi(x,h,\phi^{-1}(x,h,y))\, f_{H_t|Y_x,\boldsymbol{X}}(h|y,\boldsymbol{x})\, dh \\
=& \int \partial_x \phi(x,h,\phi^{-1}(x,h,y))\, f_{Y_x,H_t|\boldsymbol{X}}(y,h|\boldsymbol{x})\, dh
\end{aligned}$$

□

With this result in hand we can now make precise the quantile effect sketched above.

Theorem 2. *If Assumptions 1 - 3 are satisfied, $f(u|\boldsymbol{x})$ is continuously differentiable in $\boldsymbol{x}$,*

$$\int \sup_{\|\Delta_{\boldsymbol{x}}\|\leq\delta} \|\partial_{\boldsymbol{x}} f(u|\boldsymbol{x}+\Delta_{\boldsymbol{x}})\|\, du < \infty, \tag{3.5}$$

and the conditional density of Y_t given $\boldsymbol{X}$ is positive on the interior of its support then for all $0<\tau<1$, $Q_t(\tau|\boldsymbol{x})$ exists and is continuously differentiable such that (3.1) holds true.

Proof. Let $\boldsymbol{z}=(y,x',\boldsymbol{x}')'$ and let $H(\boldsymbol{z})=\mathbb{P}(\phi(x,U_t)\leq y|\boldsymbol{X}=\boldsymbol{x})$. From Lemma 1 it follows that $H(\boldsymbol{z})$ is differentiable in y and x with partial derivatives continuous in $\boldsymbol{z}$.

From (3.5), it follows that $H(\boldsymbol{z})$ is also differentiable in $\boldsymbol{x}$ with

$$\partial_{\boldsymbol{x}} H(\boldsymbol{z}) = \int 1(\phi(x,u)\leq y)\partial_{\boldsymbol{x}} f(u|\boldsymbol{x})du$$

which is continuous in $\boldsymbol{z}$. Thus, $H(\boldsymbol{z})$ is continuously differentiable in $\boldsymbol{z}$, and the derivative w.r.t. y is strictly positive (see the expression in Lemma 1). From the implicit function theorem, there is a unique solution $Q_t(\tau|\boldsymbol{x})$, $\boldsymbol{x}=(x_1',x_2')'$, to

$$\tau = H(Q_t(\tau|\boldsymbol{x}),x_t,\boldsymbol{x}), \qquad t=1,2.$$

which is differentiable with partial derivatives

$$\partial_{x_s}Q_t(\tau|\boldsymbol{x}) = -\big(\partial_y H(Q_t(\tau|\boldsymbol{x}),x_t,\boldsymbol{x})\big)^{-1}\Big(1(s=t)\partial_x H(Q_t(\tau|\boldsymbol{x}),x_t,\boldsymbol{x}) + \partial_{x_s}H(Q_t(\tau|\boldsymbol{x}),x_t,\boldsymbol{x})\Big),$$

where $\partial_{x_s}H\big(y,x,\boldsymbol{x}\big)$ is the partial derivative w.r.t. the components of x_s in $\boldsymbol{x}=(x_1',x_2')'$, $s=1,2$. Evaluating at $x_1=x_2=x$, subtracting and plugging in the expressions for the derivatives from Lemma 1 yields (3.1). □

4 Quantiles of Transformations of the Dependent Variables

In this section we answer the question whether we can relate quantiles of the first difference of the dependent variable to causal effects. In fact, the same arguments and assumptions that are used for first differences can also be employed for arbitrary functions of the dependent variables which map the T-vector of dependent variables $\boldsymbol{Y}$ (in our case for simplicity $T=2$) into a scalar "index". However, as it turns out, if we restrict ourselves to using only two time periods of the covariates X_t, we have to strengthen the assumptions significantly to make statements about causal effects. This is related to the fact that we do not have an auxiliary equation at our disposal that allows us to correct for the heterogeneity bias that arose from the correlation of X_t and U_s.

To be more specific about the assumptions: While still considering the model specified in Assumption 1, instead of time homogeneity assumption 2, in this section we shall use independence assumptions.

Assumption 4. *1. (V_1, V_2) are independent of $(X_1, X_2)|A$,*

2. A is independent of $X_2|X_1$,

The first part of this assumption states that the transitory error component is independent of covariates, given the persistent fixed effect, which is a notion of strict exogeneity. The second part of this assumption is more restrictive as it rules out the case where A is arbitrarily correlated with the X_t process. This is a special case of the sufficient statistic type assumptions in Altonji and Matzkin (2005). To adopt a similar framework as above, we rewrite

$$\boldsymbol{U} = (V_1, V_2, A)^T ,$$

and note that the independence and strict exogeneity assumptions imply that:

Lemma 2. *Under Assumption 4, $\boldsymbol{U}$ and X_2 are independent given X_1.*

Proof. For measurable sets K_i, $i = 1, 2, 3$,

$$\begin{aligned} & P\big(V_1 \in K_1, V_2 \in K_2, A \in K_3 | X_1 = x_1, X_2 = x_2\big) \\ = & \int_{K_3} P\big(V_1 \in K_1, V_2 \in K_2 | A = a, X_1 = x_1, X_2 = x_2\big)\, P_{A|X_1,X_2}(da|x_1, x_2) \\ = & \int_{K_3} P\big(V_1 \in K_1, V_2 \in K_2 | A = a\big)\, P_{A|X_1}(da|x_1). \end{aligned}$$

Thus, the conditional distribution of $\boldsymbol{U}$ given X_1, X_2 does not depend on X_2, proving conditional independence. □

As already mentioned above, we consider now quantiles of differences and other transformations of the dependent variables. To this end, let $\psi(y_1, y_2)$ be an arbitrary (differentiable) function and note that

$$\tilde{Y} = \psi(Y_1, Y_2) = \psi\left(\phi(X_1, V_1, A), \phi(X_2, V_2, A)\right) =: g(X_1, X_2, \boldsymbol{U}), \tag{4.1}$$

so that for $\boldsymbol{u} = (v_1, v_2, a)$, we have that $g(x_1, x_2, \boldsymbol{u}) = \psi(\phi(x_1, v_1, a), \phi(x_2, v_2, a))$. Denote by $\tilde{q}(\tau, x_1, x_2)$ the conditional quantile of $\tilde{Y}$ given $X = x_1, X_2 = x_2$, so that

$$P\left[\tilde{Y} \leq \tilde{q}(\tau, x_1, x_2) | X_1 = x_1, X_2 = x_2\right] = \tau.$$

For convenience, we first formulate and prove a result along the lines of Hoderlein and Mammen (2007) for a general model of the form

$$Y = g(X_1, X_2, \boldsymbol{U}), \tag{4.2}$$

in terms of regularity assumptions similar to Assumption 3, and then specialize it to (4.1).

Assumption 5. *Suppose that in the model (4.2), we can write* $\boldsymbol{u} = (h', e)'$ *for scalar* e *, such that* $g(x_1, x_2, \boldsymbol{u}) = g(x_1, x_2, h, e)$ *is continuously differentiable in* x_2 *and* e*. Moreover, for fixed* x_1 *there is a* $C > 0$ *(possibly depending on* x_1*) with* $\partial_e g(x_1, x_2, h, e) \geq 1/C$ *and* $\|\partial_{x_2} g(x_1, x_2, \boldsymbol{u})\| \leq C$ *for all* x_2 *and* $\boldsymbol{u}$*. For the corresponding representation of the random vector* $\boldsymbol{U} = (H, E)$*,* E *is absolutely continuously distributed given* (H, X_1)*, with conditional pdf* $f_E(e|h, x_1)$ *that is bounded and continuous in* e*, and the conditional distribution of* H *given* X_1 *is absolutely continuous with pdf* $f(h|x_1)$*.*

These assumptions are by and large regularity conditions, akin to those employed in Hoderlein and Mammen (2007), e.g., differentiability conditions. They do not restrict the model significantly, and we therefore do not discuss them at length. Together with the independence condition, they allow us to establish an extension to the Hoderlein and Mammen (2007) result:

Proposition 1. *Suppose that in the model (4.2),* X_2 *is conditionally independent of* $\boldsymbol{U}$ *given* X_1*, that Assumption 5 is satisfied and that the conditional pdf of* Y *given* X_1 *and* X_2 *is positive in the interior of its support. Then for every* $0 < \tau < 1$*, the conditional quantile* $q(\tau, x_1, x_2)$ *of* Y *given* $X_1 = x_1$*,* $X_2 = x_2$ *exists and is continuously differentiable with*

$$\partial_{x_2} q(\tau, x_1, x_2) = E\left[\partial_{x_2} g(X_1, X_2, \boldsymbol{U}) | Y = q(\tau, x_1, x_2), X_1 = x_1, X_2 = x_2\right]. \tag{4.3}$$

Proof. Fix x_1, and set

$$H(y, x_2) = P(Y \leq y | X_1 = x_1, X_2 = x_2) = P(g(x_1, x_2, \boldsymbol{U}) \leq y | X_1 = x_1)$$

by the form of the model and the conditional independence assumption. Below we show that from Assumption 5, $H(y, x_2)$ is continuously partially differentiable with derivatives

$$\begin{aligned} \partial_y H(y, x_2) &= f_{Y|X_1,X_2}(y|x_1, x_2), \\ \partial_{x_2} H(y, x_2) &= -f_{Y|X_1,X_2}(y|x_1, x_2)\, E\big(\partial_{x_2} g(X_1, X_2, \boldsymbol{U}) | Y = y, X_1 = x_1, X_2 = x_2\big). \end{aligned} \tag{4.4}$$

From the positivity of the conditional density of Y given X_1, X_2 and the implicit function theorem, the conditional quantile $q(\tau, x_1, x_2)$ given by $H\big(q(\tau, x_1, x_2), x_2\big) = \tau$ exists and is differentiable in x_2 with derivative

$$\begin{aligned} \partial_{x_2} q(\tau, x_1, x_2) &= -\Big(\partial_y H\big(q(\tau, x_1, x_2), x_2\big)\Big)^{-1} \partial_{x_2} H\big(q(\tau, x_1, x_2), x_2\big) \\ &= E\big(\partial_{x_2} g(X_1, X_2, \boldsymbol{U}) | Y = q(\tau, x_1, x_2), X_1 = x_1, X_2 = x_2\big). \end{aligned}$$

where we used (4.4), thus proving the theorem.

It remains to prove (4.4), which is analogous to Lemma 1. Let $F_E(e|h,\boldsymbol{x}) = \mathbb{P}(E \le e|H = h, X_1 = x_1) = \int_{-\infty}^{e} f_E(e|h,x_1)\,de$, so that $F_E(e|h,x_1)$ is differentiable in e with derivative $f_E(e|h,x_1)$ that is continuous in e. Consider

$$\begin{aligned}
&\mathbb{P}(g(x_1,x_2,\boldsymbol{U}) \le y|X_1 = x_1) = \int 1(g(x_1,x_2,\boldsymbol{u}) \le y)\, f(\boldsymbol{u}|x_1)\,du \\
&= \int\int 1(g(x_1,x_2,h,e) \le y)\, f_E(e|h,x_1)\, f(h|x_1)de\,dh \\
&= \int\int 1(e \le g^{-1}(x_1,x_2,h,y))\, f_E(e|h,x_1)\, f(h|x_1)de\,dh \\
&= \int F_E(g^{-1}(x_1,x_2,h,y)|h,x_1) f(h|x_1)\,dh.
\end{aligned} \tag{4.5}$$

By the inverse and implicit function theorems, $g^{-1}(x_1,x_2,h,y)$ is continuously differentiable in x and y, with

$$\begin{aligned}
\partial_y\big(g^{-1}(x_1,x_2,h,y)\big) &= \big(\partial_e g(x_1,x_2,h,g^{-1}(x_1,x_2,h,y))\big)^{-1}, \\
\partial_{x_2}\big(g^{-1}(x_1,x_2,h,y)\big) &= -\frac{\partial_{x_2} g(x_1,x_2,h,g^{-1}(x_1,x_2,h,y))}{\partial_e g(x_1,x_2,h,g^{-1}(x_1,x_2,h,y))} \\
&= -\partial_{x_2} g(x_1,x_2,h,g^{-1}(x_1,x_2,h,y))\,\partial_y\big(g^{-1}(x_1,x_2,h,y)\big).
\end{aligned}$$

Then by Assumption 5 both $\partial_y\big(g^{-1}(x_1,x_2,h,y)\big)$ and $\partial_{x_2}\big(g^{-1}(x_1,x_2,h,y)\big)$ are continuous in y and x_2 and bounded. Therefore,

$$\begin{aligned}
\partial_y\big(F_E(g^{-1}(x_1,x_2,h,y)|h,x_1)\big) &= f_E(g^{-1}(x_1,x_2,h,y)|h,x_1)\,\partial_y\big(g^{-1}(x_1,x_2,h,y)\big) \\
&= f_{Y|X_1,X_2,H}(y|x_1,x_2,h), \\
\partial_{x_2}\big(F_E(g^{-1}(x_1,x_2,h,y)|h,\boldsymbol{x})\big) &= f_E(g^{-1}(x_1,x_2,h,y)|h,\boldsymbol{x})\,\partial_{x_2}\big(g^{-1}(x_1,x_2,h,y)\big) \\
&= -f_{Y|X_1,X_2,H}(y|x_1,x_2,h)\,\partial_{x_2} g(x_1,x_2,h,g^{-1}(x_1,x_2,h,y)),
\end{aligned} \tag{4.6}$$

are both bounded and continuous in y, x and $\boldsymbol{x}$, where the last equality in each equation follows by a change of variables argument together with conditional independence of X_2 and (H,E) given X_1. From the boundedness assumptions on f_E and on $\partial_{x_2} g$ in Assumption 5, it follows that $H(y,x_2)$ is partially differentiable with continuous partial derivatives which can be computed by differentiating under the integral in (4.5). Now insert (4.6) into (4.5) and note that $f_{Y|X_1,X_2,H}(y|x_1,x_2,h) f(h|x_1) = f_{Y,H|X_1,X_2}(y,h|x_1,x_2)$ by conditional independence. The first expression is then immediate. For the second, note that given $Y = y$ (for a fixed x_1, x_2), $E = g^{-1}(x_1,x_2,H,y)$, so that

$$\begin{aligned}
&f_{Y|X_1,X_2}(y|x_1,x_2)\, E\big(\partial_{x_2} g(x_1,x_2,\boldsymbol{U})|Y = y, X_1 = x_1, X_2 = x_2\big) \\
=&f_{Y|X_1,X_2}(y|x_1,x_2) \int \partial_{x_2} g(x_1,x_2,h,g^{-1}(x_1,x_2,h,y))\, f_{H|Y,X_1,X_2}(h|y,x_1,x_2)\,dh \\
=& \int \partial_{x_2} g(x_1,x_2,h,g^{-1}(x_1,x_2,h,y))\, f_{Y,H|X_1,X_2}(y,h|x_1,x_2)\,dh.
\end{aligned}$$

□

We now specialize this general result to the setup of this paper, and discuss it below in this specialized setup. To this end, we modify the regularity conditions accordingly:

Assumption 6. *Suppose that in the model (4.2), $\psi(y_1, y_2)$ is continuously partially differentiable in y_2 with $1/K \leq \partial_{y_2}\psi(y_1, y_2) \leq K$ for all y_1, y_2 for some $K > 0$. Further, assume that we can write $v = (\tilde{h}', e)'$ for scalar e , such that $\phi(x, v, a) = \phi(x, \tilde{h}, e, a)$ is continuously differentiable in x and e and such that there is a $C > 0$ with $\partial_e\phi(x, h, e, a) \geq 1/C$ and $|\partial_x\phi(x, v, a)| \leq C$ for all x, v, a. For the corresponding representation of the random vector $V_2 = (H, E)$, E is absolutely continuously distributed given (X_1, H, A, V_1), with conditional pdf that is bounded and continuous in e, and the conditional distribution of (H, A, V_1) given X_1 is absolutely continuous.*

These preliminaries lead to the expected corollary:

Corollary 3. *Suppose that in (4.1), Assumptions 1, 4 and 6 are satisfied, and that the conditional density of $\tilde{Y}$ given $X_1 = x_1, X_2 = x_2$ is positive in the interior of its support. Then (4.3) holds true.*

This result is very similar in spirit to the results in the previous section, again an LAR for a subpopulation (or a derivative for an ASF) is identified. The advantage, however, is now that we can look at subpopulations that are characterized by arbitrary combinations of Y_1 and Y_2. If we confine ourselves to linear combinations, i.e., $\tilde{Y} = \lambda Y_1 + \pi Y_2$, we can consider conditioning on arbitrary weights λ, π. Since we can vary λ, π freely, this means that we can use the entire joint distribution in the sense of the Cramer-Wold device, by looking at any linear combination, and hence use multivariate information through repeated use of one regular regression quantiles. It allows to construct subpopulations where we put different weights on the outcome in different periods. For instance, if X is schooling, and Y_t is labor income in different periods, we may think of $\tilde{Y}$ as some long run or average income. And when computing this long run income, we could either discount future income stronger or emphasize it more when characterizing the subpopulations, depending on the intention of the researcher. Of course, one should always remember that the strength in statements we can make always comes at the expense of the structure we impose on the dependence between A and X_t.

This result covers important special cases:

1. The difference: $\psi(y_1, y_2) = y_2 - y_1 = \Delta y$. Then $\tilde{q}(\tau, x_1, x_2)$ is the conditional quantile of the difference, and

$$\partial_{x_2} q^{\Delta Y}(\tau, x_1, x_2) = E\left[\partial_x\phi(x_2, V_2, A)|X_1 = x_1, X_2 = x_2, \Delta Y = q^{\Delta Y}(\tau, x_1, x_2)\right].$$

2. Y_1: Here $\psi(y_1, y_2) = y_1$, so that $\partial_{y_2}\psi = 0$, yielding

$$\partial_{x_2} q^{Y_1}(\tau, x_1, x_2) = 0.$$

This is evident, since Y_1 is conditionally independent of X_2 given X_1. An analogous argument applies to Y_2. This second result is similar in spirit to Altonji and Matzkin (2005), just replacing means by quantiles.

Note that the first special case answers one of the questions posed in the introduction: should we consider the difference of the quantiles or the quantiles of the differences, when talking about causal effects in panels. In terms of the strength of the assumptions, the verdict has to be clearly differences of quantiles. However, it also happens to be the case that under the additional structure on the dependence the quantiles of the difference yield a new effect that we could not have obtained through differences in quantiles.

5 Time Effects

The time homogeneity assumption is a strong one that often seems not to hold in applications. In this section we consider one way to weaken it, by allowing for additive location effects and multiplicative scale effects. Allowing for such time effects leads to effects of interest being exactly identified, unlike the overidentification we found in Sections 2 and 3.

We allow for time effects by replacing Assumption 1 with the following condition.

Assumption 7. *There are functions ϕ, μ_t, and σ_t, and a vector of random variables U_t such that*

$$Y_t = \mu_t(X_t) + \sigma_t(X_t)\phi(X_t, U_t), \quad (t = 1, 2).$$

The time effects μ_t and σ_t are not separately identifiable from ϕ without location and scale normalizations because

$$\mu_t(x) + \sigma_t(x)\phi(x, u) = \tilde{\mu}_t(x) + \tilde{\sigma}_t(x)\tilde{\phi}(x, u),$$

for $\tilde{\mu}_t(x) = \mu_t(x) + \sigma_t(x)\Delta_\mu(x)$, $\tilde{\sigma}_t(x) = \Delta_\sigma(x)\sigma_t(x)$, $\tilde{\phi}(x, u) = [\phi(x, u) - \Delta_\mu(x)]/\Delta_\sigma(x)$, and $\Delta_\sigma(x) \neq 0$.

In this model the effects of interest vary with time. We consider the time-averaged conditional mean effect:

$$\partial_x\bar{\mu}(x) + \partial_x\bar{\sigma}(x)E[\phi(x, U_t) \mid X_1 = X_2 = x] + \bar{\sigma}(x)E[\partial_x\phi(x, U_t) \mid X_1 = X_2 = x],$$

and the time-averaged conditional quantile effect:

$$\partial_x \bar{\mu}(x) + \partial_x \bar{\sigma}(x) q(\tau, x) + \bar{\sigma}(x) E[\partial_x \phi(x, U_t) \mid \phi(x, U_t) = q(\tau, x), X_1 = X_2 = x],$$

where $\bar{\mu}(x) = [\mu_1(x) + \mu_2(x)]/2$, $\bar{\sigma}(x) = [\sigma_1(x) + \sigma_2(x)]/2$, and $q(\tau, x)$ is the τ^{th} conditional quantile of $\phi(x, U_t)$ given $X_1 = X_2 = x$.

The conditional mean effect is related to the time-averaged CASF:

$$\bar{m}(x \mid \boldsymbol{x}) = \bar{\mu}(x) + \bar{\sigma}(x) E[\phi(x, U_t) \mid \boldsymbol{X} = \boldsymbol{x}],$$

through

$$\partial_x \bar{m}(x \mid \boldsymbol{x})\Big|_{\boldsymbol{x}=(x,x)} = \partial_x \bar{\mu}(x) + \partial_x \bar{\sigma}(x) E[\phi(x, U_t) \mid X_1 = X_2 = x] + \bar{\sigma}(x) E[\partial_x \phi(x, U_t) \mid X_1 = X_2 = x],$$

under the conditions that permit interchanging the derivative and expectation. Similarly, the conditional quantile effect is related to the time-averaged CQSF, $\bar{q}_\tau(x \mid \boldsymbol{X} = \boldsymbol{x})$, that gives the τ-quantile of $\bar{\mu}(x) + \bar{\sigma}(x)\phi(x, U_t)$ conditional on $\boldsymbol{X} = \boldsymbol{x}$, through

$$\partial_x \bar{q}_\tau(x \mid \boldsymbol{x})\Big|_{\boldsymbol{x}=(x,x)} = \partial_x \bar{\mu}(x) + \partial_x \bar{\sigma}(x) q(\tau, x) + \bar{\sigma}(x) E[\partial_x \phi(x, U_t) \mid \phi(x, U_t) = q(\tau, x), X_1 = X_2 = x].$$

Let $V_t(\boldsymbol{x}) = \text{Var}[Y_t \mid \boldsymbol{X} = \boldsymbol{x}]$, and $\sigma(x) = \sigma_2(x)/\sigma_1(x)$.

Theorem 4. *Suppose that Assumptions 2 and 7 are satisfied,* $E[Y_t^2] < \infty$, $(t = 1, 2)$, $V_t(x, x) > 0$, $(t = 1, 2)$, $\phi(x, u)$, $\mu_t(x)$, *and* $\sigma_t(x)$, $(t = 1, 2)$, *are continuously differentiable in* x, *and the conditional density of* U_t *given* $\boldsymbol{X} = \boldsymbol{x}$, $f(u|\boldsymbol{x})$, *is continuously differentiable in* $\boldsymbol{x}$. *Given* x, *suppose that for some* $\varepsilon > 0$,

$$\int \sup_{\|\delta\| \leq \varepsilon, \delta = (\delta_0', \delta_1', \delta_2')'} \left\| \partial_x \phi(x + \delta_0, u)\, f(u | x + \delta_1, x + \delta_2) \right\| du < \infty,$$
$$\int \sup_{\|\delta\| \leq \varepsilon, \delta = (\delta_0', \delta_1', \delta_2')'} \left\| \phi(x + \delta_0, u)\, \partial_{x_s} f(u | x + \delta_1, x + \delta_2) \right\| du < \infty, \quad s = 1, 2.$$

Then, $\sigma^2(x) = V_2(x, x)/V_1(x, x)$, $\mu_2(x) - \mu_1(x)\sigma(x) = E[Y_2 - \sigma(x) Y_1 \mid X_1 = X_2 = x]$, *and*

$$\begin{aligned} &\partial_x \bar{\mu}(x) + \partial_x \bar{\sigma}(x) E[\phi(x, U_t) \mid X_1 = X_2 = x] + \bar{\sigma}(x) E[\partial_x \phi(x, U_t) \mid X_1 = X_2 = x] \\ &\qquad = [\partial_{x_1} M_1(x, x) - \partial_{x_1} M_2(x, x)/\sigma(x)]/2 + [\partial_{x_2} M_2(x, x) - \sigma(x) \partial_{x_2} M_1(x, x)]/2. \end{aligned}$$

This theorem shows that the time effects are identified up to location and scale normalizations. For example, if we set $\mu_1(x) = 0$ and $\sigma_1(x) = 1$, then $\sigma_2^2(x) = V_2(x, x)/V_1(x, x)$ and $\mu_2(x) = E[Y_2 - \sigma_2(x) Y_1 \mid X_1 = X_2 = x]$. The identification of the conditional mean effect does not require any normalization. Note that we now have just one equation for identifying the conditional mean effect.

We find a similar result for quantiles.

Theorem 5. *Suppose that Assumptions 2 , 3 , and 7 are satisfied,* $\mu_t(x)$*, and* $\sigma_t(x)$*,* $(t = 1, 2)$*, are continuously differentiable in* x*,* $f(u|\boldsymbol{x})$ *is continuously differentiable in* $\boldsymbol{x}$*,*

$$\int \sup_{\|\Delta_{\boldsymbol{x}}\| \leq \delta} \|\partial_{\boldsymbol{x}} f(u|\boldsymbol{x} + \Delta_{\boldsymbol{x}})\| \, du < \infty, \tag{5.1}$$

and the conditional density of Y_t *given* $\boldsymbol{X}$ *is positive on the interior of its support. Then for all* $0 < \tau < 1$*,* $Q_t(\tau|\boldsymbol{x})$ *exists and is continuously differentiable at* $\boldsymbol{x} = (x', x')'$ *such that*

$$\begin{aligned}
&\partial_x \bar{\mu}(x) + \partial_x \bar{\sigma}(x) q(\tau, x) + \bar{\sigma}(x) E[\partial_x \phi(x, U_t) \mid \phi(x, U_t) = q(\tau, x), X_1 = X_2 = x] \\
&\quad = [\partial_{x_1} Q_1(\tau \mid x, x) - \partial_{x_1} Q_2(\tau \mid x, x)/\sigma(x)]/2 + [\partial_{x_2} Q_2(\tau \mid x, x) - \sigma(x) \partial_{x_2} Q_1(\tau \mid x, x)]/2,
\end{aligned}$$

$\sigma(x) = [Q_2(\tau_1 \mid x, x) - Q_2(\tau_2 \mid x, x)]/[Q_1(\tau_1 \mid x, x) - Q_1(\tau_2 \mid x, x)]$*, and* $\mu_2(x) - \sigma(x)\mu_1(x) = Q_2(\tau_3 \mid x, x) - \sigma(x) Q_1(\tau_3 \mid x, x)$*, for any* $0 < \tau_3 < 1$ *and* $0 < \tau_2 < \tau_1 < 1$ *such that* $[Q_1(\tau_1 \mid x, x) - Q_1(\tau_2 \mid x, x)] > 0$.

As in Theorem 4, the time effects are identified up to location and scale normalizations, whereas the conditional quantile effects are identified without any normalization. Here, however, instead of conditional mean and variance restrictions, we use quantile restrictions to identify the time effects up to the normalizations. These effects are over identified by many possible quantiles τ_1, τ_2 and τ_3. For example, for $\tau_1 = .9$, $\tau_2 = .1$ and $\tau_3 = .5$, the scale is identified by a ratio of conditional interdecile ranges across time and the location is identified by a difference of conditional medians across time.

6 Estimation and inference

The conditional mean and quantile effects of interest are identified by special cases of the functionals:

$$\theta_m(x) = h_m(\{M_t(x, x), V_t(x, x) : t = 1, 2\}), \quad x \in \mathcal{X},$$

and

$$\theta_q(w) = h_q(\{Q_t(\tau \mid x, x) : t = 1, 2\}), \quad w = (x, \tau) \in \mathcal{W},$$

respectively, where h_m and h_q are known smooth functions, $\mathcal{X}$ is a region of regressor values of interest, and $\mathcal{W}$ is a region of regressor values and quantiles of interest. We consider the estimators of θ_m and θ_q based on the plug-in rule:

$$\widehat{\theta}_m(x) = h_m(\{\widehat{M}_t(x, x), \widehat{V}_t(x, x) : t = 1, 2\}), \quad x \in \mathcal{X},$$

and

$$\widehat{\theta}_q(w) = h_q(\{\widehat{Q}_t(\tau \mid x, x) : t = 1, 2\}), \quad w = (x, \tau) \in \mathcal{W},$$

where $\widehat{M}_t(x, x)$, $\widehat{Q}_t(\tau \mid x, x)$, and $\widehat{V}_t(x, x)$ are nonparametric series estimators of $M_t(x, x)$, $Q_t(\tau \mid x, x)$, and $V_t(x, x)$.

To describe the series estimators, let $P^K(\boldsymbol{x}) = (p_{1K}(\boldsymbol{x}), \ldots, p_{KK}(\boldsymbol{x}))'$ denote a $K \times 1$ vector of approximating functions, such as power series or splines, and let $\boldsymbol{P}_i = P^K(\boldsymbol{X}_i)$. Then,

$$\widehat{M}_t(x, x) = P^K(x, x)' \left(\sum_{i=1}^n \boldsymbol{P}_i \boldsymbol{P}_i' \right)^- \sum_{i=1}^n \boldsymbol{P}_i Y_{it},$$

where A^- denotes any generalized inverse inverse of the matrix A;

$$\widehat{V}_t(x, x) = P^K(x, x)' \left(\sum_{i=1}^n \boldsymbol{P}_i \boldsymbol{P}_i' \right)^- \sum_{i=1}^n \boldsymbol{P}_i [Y_{it} - \widehat{M}_t(\boldsymbol{X}_i)]^2$$

is a series version of the (kernel) conditional variance estimator of Fan and Yao (1998); and $\widehat{Q}_t(\tau \mid x, x) = P^K(x, x)' \widehat{\beta}_t(\tau)$, where $\widehat{\beta}_t(\tau)$ is the Koenker and Bassett (1978) quantile regression estimator

$$\widehat{\beta}_t(\tau) \in \arg\min_{b \in \mathbb{R}^K} \sum_{i=1}^n [\tau - 1\{Y_{it} \leq \boldsymbol{P}_i' b\}][Y_{it} - \boldsymbol{P}_i' b].$$

We use weighted bootstrap for inference. To describe this method, let $(w_1, \ldots, w_n)$ be an i.i.d. sequence of nonnegative random variables from a distribution with mean and variance equal to one (e.g., the standard exponential distribution), independent of the data. The weighted bootstrap uses the components of $(w_1, \ldots, w_n)$ as random sampling weights in the construction of the bootstrap version of the series estimators. Thus, the bootstrap versions of $\widehat{\theta}_m(w)$ and $\widehat{\theta}_q(w)$ are

$$\widehat{\theta}^*_m(x) = h(\{\widehat{M}^*_t(x, x), \widehat{V}^*_t(x, x) : t = 1, 2\}), \quad x \in \mathcal{X},$$

and

$$\widehat{\theta}^*_q(w) = h(\{\widehat{Q}^*_t(\tau \mid x, x) : t = 1, 2\}), \quad w = (x, \tau) \in \mathcal{W},$$

where

$$\widehat{M}^*_t(x, x) = P^K(x, x)' \left(\sum_{i=1}^n w_i \boldsymbol{P}_i \boldsymbol{P}_i' \right)^- \sum_{i=1}^n w_i \boldsymbol{P}_i Y_{it}$$

is the bootstrap version of $\widehat{M}_t(x, x)$,

$$\widehat{V}^*_t(x, x) = P^K(x, x)' \left(\sum_{i=1}^n w_i \boldsymbol{P}_i \boldsymbol{P}_i' \right)^- \sum_{i=1}^n w_i \boldsymbol{P}_i [Y_{it} - \widehat{M}^*_t(\boldsymbol{X}_i)]^2$$

is the bootstrap version of $\widehat{V}^*_t(x, x)$, and $\widehat{Q}^*_t(\tau \mid x, x) = P^K(x, x)' \widehat{\beta}^*_t(\tau)$ is the bootstrap version of $\widehat{Q}_t(\tau \mid x, x)$, with

$$\widehat{\beta}^*_t(\tau) = \arg\min_{b \in \mathbb{R}^K} \sum_{i=1}^n w_i [\tau - 1\{Y_{it} \leq \boldsymbol{P}_i' b\}][Y_{it} - \boldsymbol{P}_i' b].$$

Chernozhukov, Lee, and Rosen (2013) and Belloni, Chernozhukov, and Fernandez-Val (2011) developed functional distributional theory and bootstrap consistency results for series estimators of functionals of the conditional mean and quantile functions. We can use these results to construct analytical or bootstrap confidence bands for the effects that have uniform asymptotic coverage over regressor values and quantiles. For example, the end-point functions of a $1-\alpha$ confidence band for θ_q have the form

$$\widehat{\theta}_q^{\pm}(w) = \widehat{\theta}_q(w) \pm \widehat{t}_{q,1-\alpha}\widehat{\Sigma}_q(w)^{1/2}/\sqrt{n}, \tag{6.1}$$

where $\widehat{\Sigma}_q(w)$ and $\widehat{t}_{q,1-\alpha}$ are consistent estimators of the asymptotic variance function of $\sqrt{n}[\widehat{\theta}_q(w) - \theta_q(w)]$ and the $1-\alpha$ quantile of the Kolmogorov-Smirnov maximal t-statistic

$$t_q = \sup_{w \in \mathcal{W}} \widehat{\Sigma}_q(w)^{-1/2}\sqrt{n}|\widehat{\theta}_q(w) - \theta_q(w)|.$$

The following algorithm describes how to obtain uniform bands for quantile effects using weighted bootstrap:

Algorithm 1 (Uniform inference)**.** *(i) Draw* $\{\widehat{Z}_{q,b}^* : 1 \leq b \leq B\}$ *as i.i.d. realizations of* $\widehat{Z}_q^*(w) = \sqrt{n}[\widehat{\theta}_q^*(w) - \widehat{\theta}_q(w)]$*, for* $w \in \mathcal{W}$*, conditional on the data. (ii) Compute a bootstrap estimate of* $\Sigma_q(w)^{1/2}$ *such as the bootstrap standard deviation:* $\widehat{\Sigma}_q(w)^{1/2} = \{\sum_{b=1}^B [\widehat{Z}_{q,b}^*(w) - \overline{Z}_q^*(w)]^2/B\}^{1/2}$ *for* $w \in \mathcal{W}$*, where* $\overline{Z}_q^*(w) = \sum_{b=1}^B \widehat{Z}_{q,b}^*(w)/B$*; or the bootstrap interquartile range of* $\widehat{Z}_q^*(w)$ *rescaled with the normal distribution:* $\widehat{\Sigma}_q(w)^{1/2} = [\widehat{Z}_{q,0.75}^*(w) - \widehat{Z}_{q,0.25}^*(w)]/1.349$ *for* $w \in \mathcal{W}$*, where* $\widehat{Z}_{q,p}^*(w)$ *is the p-sample quantile of* $\{\widehat{Z}_{q,b}^*(w) : 1 \leq b \leq B\}$*. (3) Compute realizations of the bootstrap version of the maximal t-statistic* $\widehat{t}_{q,b}^* = \sup_{w \in \mathcal{W}} \widehat{\Sigma}_q(w)^{-1/2}|\widehat{Z}_{q,b}^*(w)|$ *for* $1 \leq b \leq B$*. (iii) Form a* $(1-\alpha)$*-confidence band for* $\{\theta(w)_q : w \in \mathcal{W}\}$ *using (6.1) setting* $\widehat{t}_{q,1-\alpha}$ *to the* $(1-\alpha)$*-sample quantile of* $\{\widehat{t}_{q,b}^* : 1 \leq b \leq B\}$*.* □

We can construct uniform bands for the conditional mean effects using a similar algorithm replacing $\theta_q(w)$ by $\theta_m(x)$ and adjusting all the steps accordingly.

7 Engel Curves in Panel Data

In this section, we illustrate the results with an empirical application on estimation of Engel curves with panel data. The Engel curve relationship describes how a household's demand for a commodity changes as the household's expenditure increases. Lewbel (2006) provides a recent survey of the extensive literature on Engel curve estimation. We use data from the 2007 and 2009 waves of the Panel Study of Income Dynamics (PSID). Since 2005, the PSID gathers information on household expenditure for different categories of commodities. The PSID does not collect information on total expenditure. We construct the total expenditure on nondurable goods and services by adding all the expenses in housing, utilities, phone, child care, food at home, food out from home, car, transportation, schooling, clothing, leisure, and health. We exclude expenses in mortgage, home insurance, car insurance, and health insurance because these categories have many missing values. Our sample contains 968 households formed by couples without children, where the head of the household was 20 to 65 year-old in 2009, and that provided information about all the relevant categories of expenditure in 2007 and 2009. We focus on the commodities food at home and leisure for comparability with recent studies (e.g., Blundell, Chen, and Kristensen, 2007, and Imbens and Newey, 2009). The expenditure share on a commodity is constructed by dividing the expenditure in this commodity by the total expenditure in nondurable goods and services.

Endogeneity in the estimation of Engel curves arises because the decision to consume a commodity may occur simultaneously with the allocation of income between consumption and savings. In contrast with the previous cross sectional literature, we do not rely on a two-stage budgeting argument that justifies the use of labor income as an instrument for expenditure. Instead, we assume that the Engel curve relationships are time homogeneous up to location and scale time effects, and rely on the availability of panel data. Specifically, we estimate

$$Y_{it} = \mu_t(X_{it}) + \sigma_t(X_{it})\phi(X_{it}, U_{it}), \quad i = 1, \ldots, 968, \; t = 1, 2,$$

where Y is the observed share of total expenditure on food at home or leisure, X is the logarithm of total expenditure, $\mu_t(X)$ and $\sigma_t(X)$ are location and scale time effects, U is a vector of unobserved household heterogeneity that satisfies time homogeneity and captures both differences in preferences and idiosyncratic household shocks, $t = 1$ corresponds to 2007, and $t = 2$ corresponds to 2009. The inclusion of time effects might be important to account for temporal changes in preferences and relative prices across commodities. For example, the price index of nondurable goods increased by 7% between 2007 and 2009, whereas the price indexes for food and leisure increased by 10% and 6% during the same period.[1] We allow these time effects to vary with total expenditure, what gives flexibility to the model.

Table 1 reports descriptive statistics for the variables used in the analysis. Both total expenditure and expenditure shares display within and between household variation, with means and standard deviations that remain stable between 2007 and 2009. The low percentage of within variation in expenditure indicates that there might be a substantial number of households with zero or little change in expenditure across years. Figure 1 plots histogram and kernel estimates of the density of the change in expenditure between 2007 and 2009. The kernel estimates are obtained using a Gassian kernel with Silverman's rule of thumb for the bandwidth. The esti-

[1] Source: Tables 2.4.4.U and 2.4.5 of Bureau of Economic Analysis.

mates confirm that there is a high density of households with zero change in expenditure. Our methods will identify mean and quantile effects for these households with $X_{i1} = X_{i2}$.

Table 1: Descriptive Statistics

	Pooled sample			2007 sample		2009 sample	
Variable	Mean	Std. Dev.	Within (%)	Mean	Std. Dev.	Mean	Std. Dev.
Log expenditure	10.13	0.56	21	10.19	0.57	10.07	0.55
Food share	0.19	0.10	28	0.19	0.10	0.20	0.10
Leisure Share	0.10	0.10	26	0.10	0.09	0.10	0.10

Note: the source of the data is the PSID. The number of observations is 968 observations for each year.

We estimate the location time effects, scale time effects, conditional mean effects, and conditional quantile effects using sample analogs of the expressions in Theorems 4 and 5. In particular, we estimate the conditional expectation, variance, and quantile functions by the nonparametric series methods described in Section 6. We consider two different specifications for the series basis in all the estimators: a quadratic orthogonal polynomial and a cubic B-spline with three knots at the minimum, median and maximum of total log-expenditure in the data set. Both specifications are additively separable in the total log-expenditures of 2007 and 2009.[2] We also compute cross sectional estimates that do not account for endogeneity. They are obtained by averaging the nonparametric series estimates in 2007 and 2009 that use the same specification of series basis as the panel estimates but only condition on contemporaneous expenditure. For inference, we construct 90% confidence bands around the estimates by weighted bootstrap with exponential weights and 499 repetitions. These bands are uniform in that they cover the entire function of interest with 90% probability asymptotically.

Figures 2 and 3 show the estimates and confidence bands for the location and time effects functions:

$$x \mapsto \mu_2(x) - \mu_1(x)\sigma(x), \quad x \mapsto \sigma(x) = \sigma_2(x)/\sigma_1(x), \quad x \in \mathcal{X},$$

based on Theorem 4, where $\mathcal{X}$ is the interval of values between the 0.10 and 0.90 sample quantiles of log-total expenditure. Here, we find that we cannot reject the hypothesis that there are no location and scale time effects for food at home, whereas we find significant evidence of time effects for leisure with both series specifications. Figure 4 plots the estimates

[2]We select these specifications by under smoothing with respect to the specification selected by cross validation applied to the estimators of the conditional expectation function.

and confidence bands for the time-averaged conditional mean effects or CASF derivatives:

$$x \mapsto \partial_x \bar{\mu}(x) + \partial_x \bar{\sigma}(x) E[\phi(x, U_{it}) \mid X_{i1} = X_{i2} = x] + \bar{\sigma}(x) E[\partial_x \phi(x, U_{it}) \mid X_{i1} = X_{i2} = x], \quad x \in \mathcal{X}.$$

As in Blundell, Chen and Kristensen (2007), the conditional ASF is decreasing in expenditure for food at home, whereas it is increasing for leisure. We find that the curve is convex for food at home and concave for leisure. Note, however, that we should interpret the shape of our panel estimates with caution because they formally correspond to multiple conditional ASFs as the conditioning set $X_1 = X_2 = x$ changes with x along the curve. The cross sectional estimates plotted in dashed lines lie outside the confidence band for leisure, indicating significant evidence of endogeneity. We do not find such evidence for food at home.

The estimates and confidence bands for the location and scale time effects in figures 5 and 6 are based on Theorem 5 with $\tau_1 = .9$, $\tau_2 = .1$, and $\tau_3 = .5$. They are similar to the ones based on conditional means and variances in figs. 2 and 3. Figure 7 plots the estimates and confidence bands for the time-averaged conditional quantile effects or CQSF derivates integrated over the values of x:

$$\tau \mapsto \int \{\bar{\mu}(x) + \partial_x \bar{\sigma}(x) q(\tau, x) + \bar{\sigma}(x) E[\partial_x \phi(x, U_{it}) \mid \phi(x, U_{it}) = q(\tau, x), X_{i1} = X_{i2} = x]\} \, \mu(dx),$$

for $\tau \in \mathcal{T}$, where μ is the empirical measure of log-expenditure, and $\mathcal{T} = [0.1, 0.9]$. Here we find heterogeneity in the Engel curve relationship across the distribution. The pattern of the effect is increasing with the quantile index for both food at home and leisure, although the estimates are not sufficiently precise to distinguish these patterns from sampling noise. As for the CASF, we find evidence of endogeneity for leisure, but not for food at home. In figures 8 and 9, we show that the panel estimates of the CQSF as a function of expenditure are decreasing for food at home and increasing for leisure at low values of expenditure. Imbens and Newey (2009) found similar patterns in their estimates of the QSF.

Overall, the empirical results show that our panel estimates of the Engel curves are similar to previous cross sectional estimates based on IV methods to deal with endogeneity. Thus, the Engel curve relationship is decreasing for food at home and increasing for leisure. Moreover, we find evidence of the presence of time effects and endogeneity for leisure, but not for food at home. These finding are consistent with consumer preferences where food at home is a necessity good with little effect on the marginal allocation of income between consumption and savings. Leisure, on the other hand, is a superior good that affects the marginal allocation of income between consumption and savings. The Engel curve relationship is stable over time for food at home, whereas it is sensitive to changes over time in preferences and relative prices for leisure.

References

[1] Altonji, J., and R. Matzkin (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors", *Econometrica* 73, 1053-1102.

[2] Belloni, A., Chernozhukov, V., and I Fernandez-Val (2011), "Conditional quantile processes based on series or many regressors," *arXiv:1105.6154*.

[3] Bester, A.C., and C. Hansen (2009), "Identification of Marginal Effects in a Nonparametric Correlated Random Effects Model," *Journal of Business & Economic Statistics* 27(2), pp. 235–250.

[4] Browning, M. and J. Carro (2007), "Heterogeneity and Microeconometrics Modeling," in Blundell, R., W.K. Newey, T. Persson (eds.), *Advances in Theory and Econometrics, Vol. 3*, Cambridge: Cambridge University Press.

[5] Blundell, R., Chen, X., and D. Kristensen (2007) "Semi-nonparametric IV Estimation of Shape-Invariant Engel Curves." *Econometrica* 75(6). pp. 1613-1669.

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

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

[8] Chamberlain, G. (1984), "Panel Data," in Z. Griliches and M. Intriligator eds *Handbook of Econometrics*. Amsterdam: North-Holland.

[9] Chernozhukov, V., Fernandez-Val, I., Hahn, J., and W. K. Newey (2013), "Average and Quantile Effects in Nonseparable Panel Models," *Econometrica*, 81(2), pp. 535–580.

[10] Chernozhukov, V., Lee, S., and A.M. Rosen (2013), "Intersection bounds: estimation and inference," *Econometrica*, 81 (2), pp. 667–737.

[11] Evdokimov, K. (2010), "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," unpublished manuscript, Princeton University.

[12] Fan, J. and Q. Yao (1998), "Efficient estimation of conditional variance functions in stochastic regression," *Biometrika*, 85, pp. 645–660.

[13] Graham, B.W. and J.L. Powell (2012), "Identification and Estimation of Average Partial Effects in "Irregular" Correlated Random Coefficient Panel Data Models," *Econometrica* 80 (5), pp. 2105–2152.

[14] Hoderlein, S., and E. Mammen (2007): "Identification of Marginal Effects in Nonseparable Models without Monotonicity," *Econometrica*, 75, 1513 - 1519.

[15] Hoderlein, S. and H. White, (2012), 'Nonparametric identi cation in nonseparable panel data models with generalized fixed effects', *Journal of Econometrics* 168, 300-314.

[16] Honore, B.E. (1992): "Trimmed Lad and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica* 60, 533-565

[17] Imbens, G. and W.K. Newey (2009), "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica* 77, 1481-1512.

[18] Koenker R. and G. Bassett, (1978): "Regression Quantiles," *Econometrica* 46, pp. 33–50.

[19] Lewbel, A. (2006) "Entry for the New Palgrave Dictionary of Economics, 2nd Edition." Boston College. .

[20] Manski, C. (1987): "Semiparametric Analysis of Random Effects Linear Models From Binary Response Data," *Econometrica* 55, 357-362.

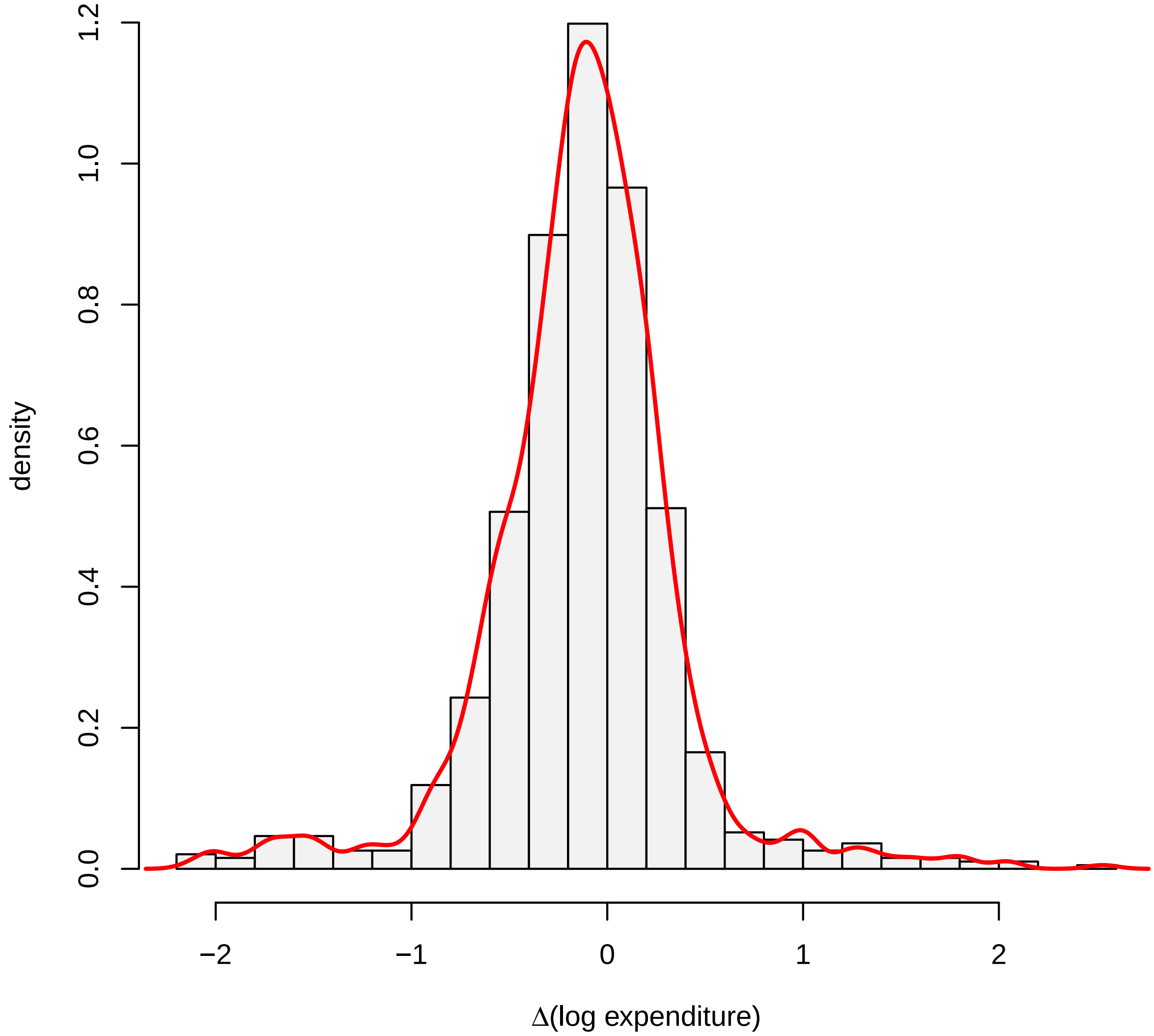

Figure 1: Density of the change in log-expenditure between 2007 and 2009.

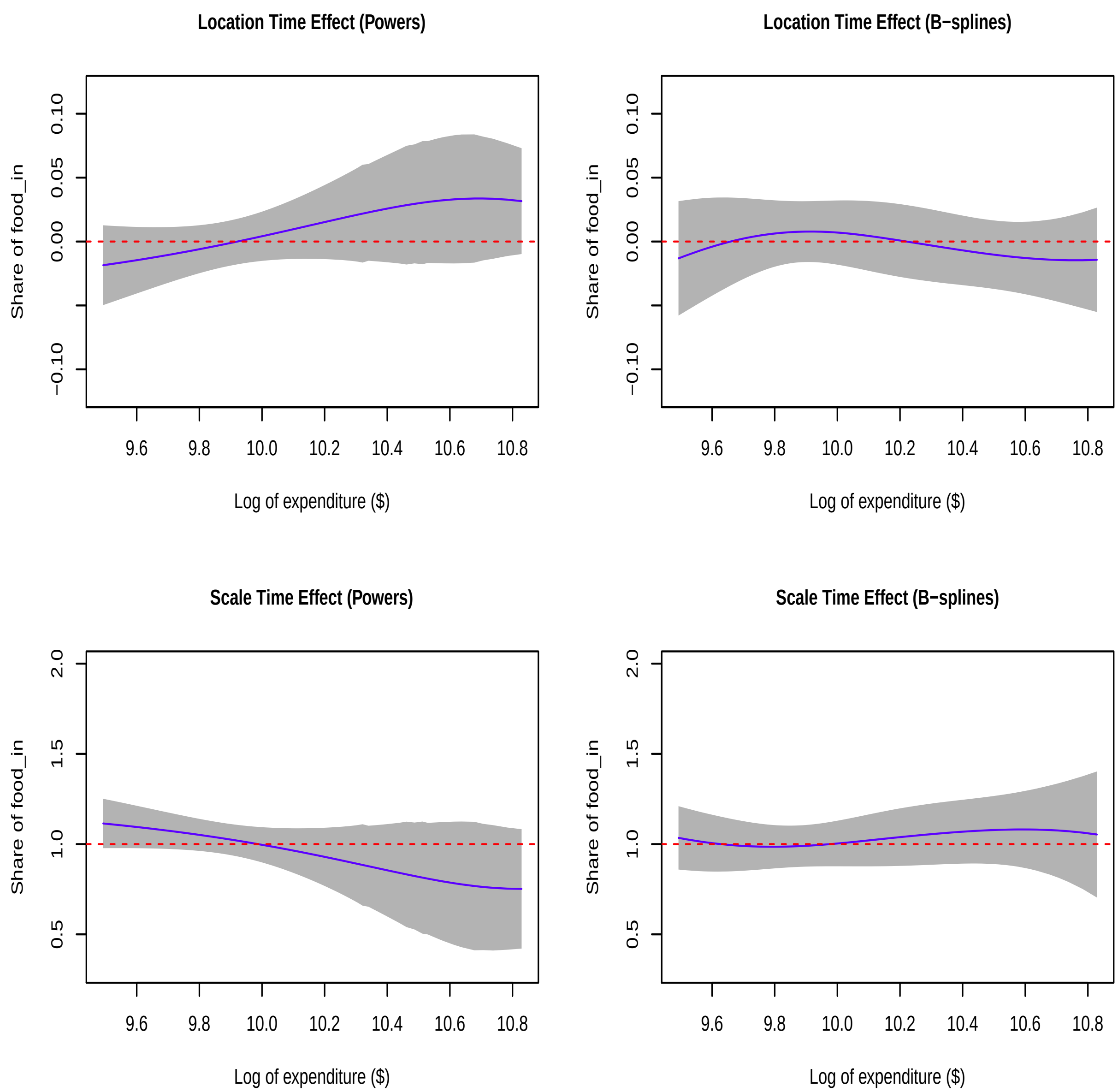

Figure 2: Location and scale time effects for food at home share: estimates from conditional means and variances.

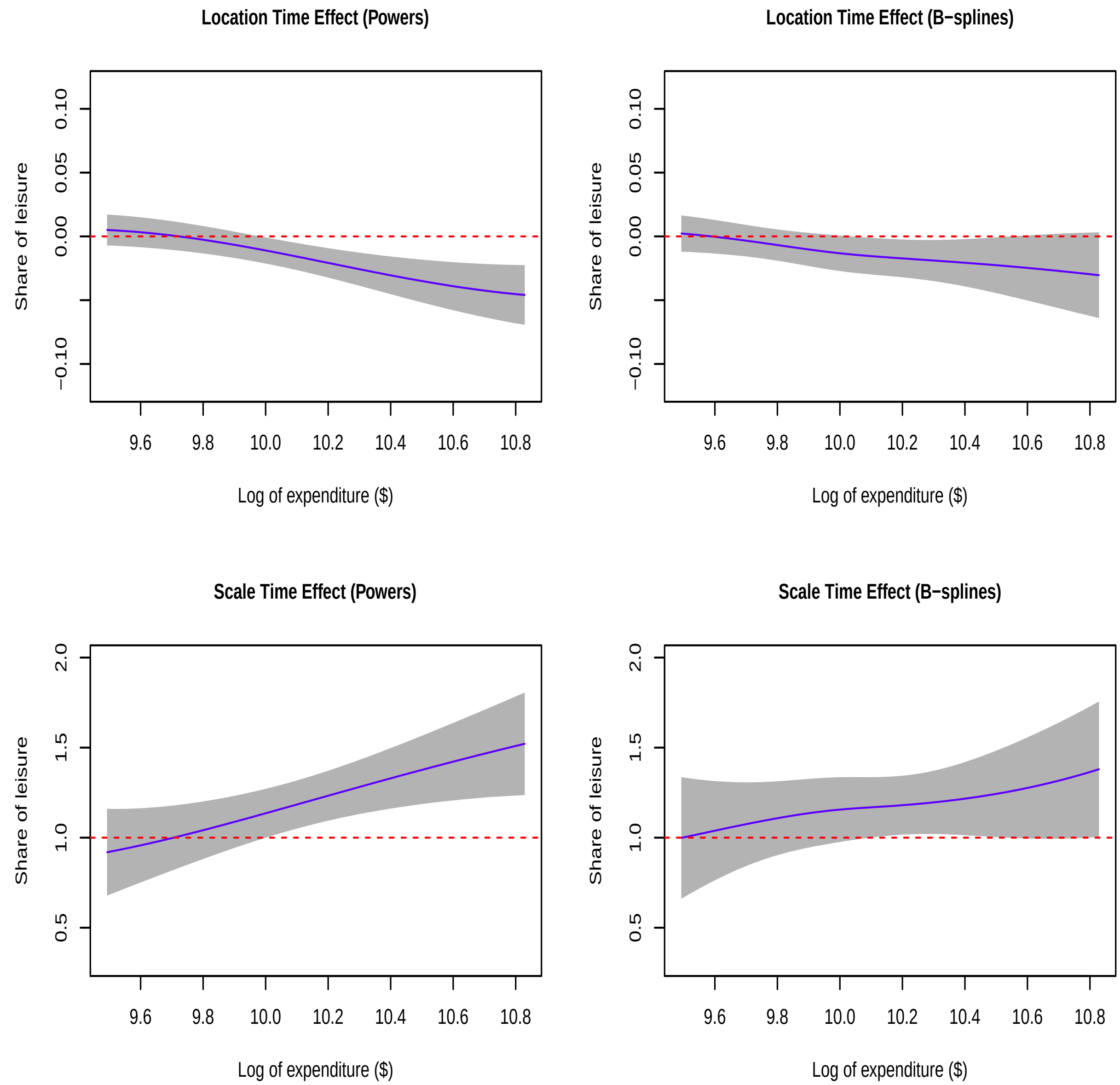

Figure 3: Location and scale time effects for leisure share: estimates from conditional means and variances.

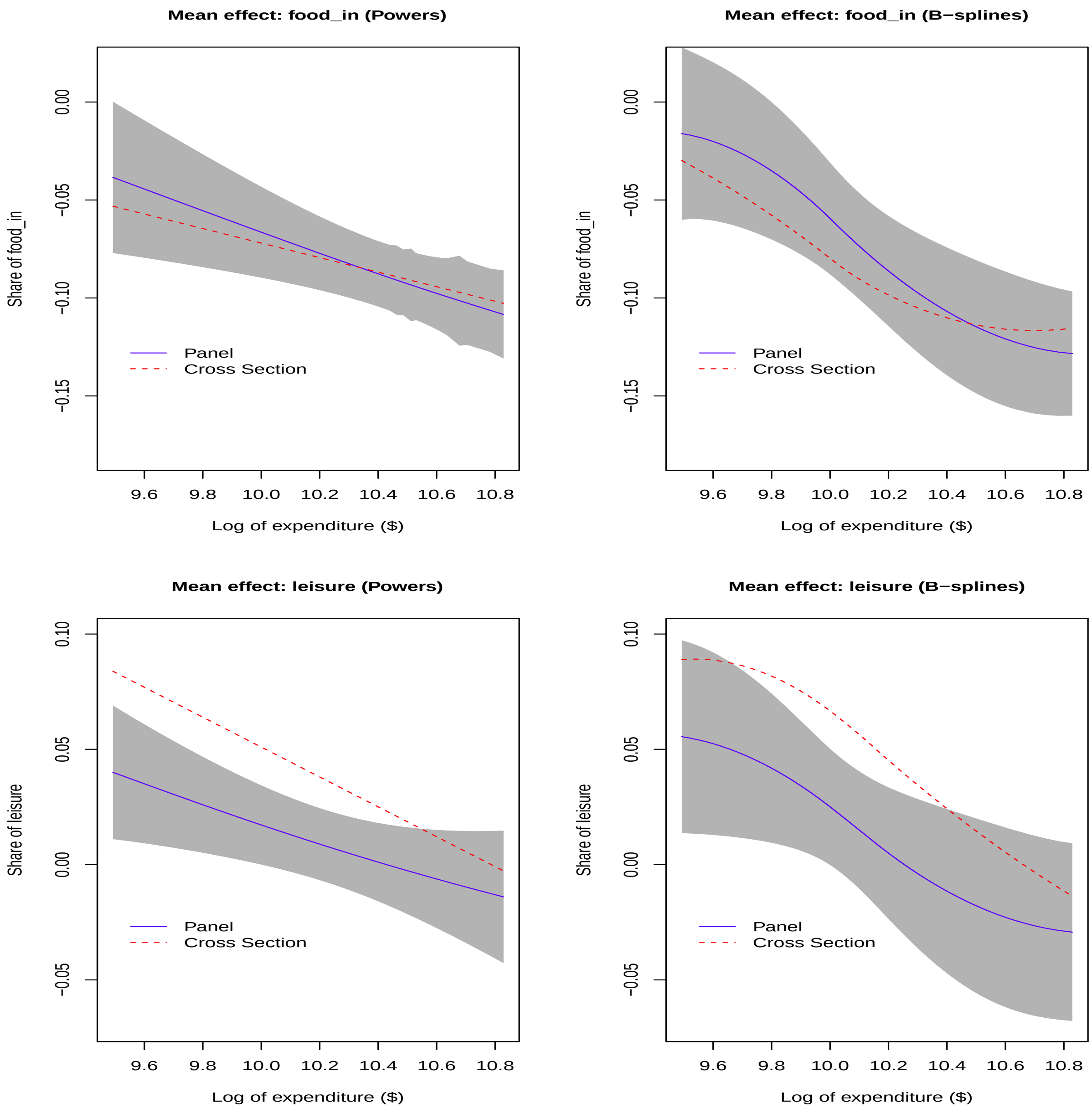

Figure 4: Conditional mean effects of log total expenditure.

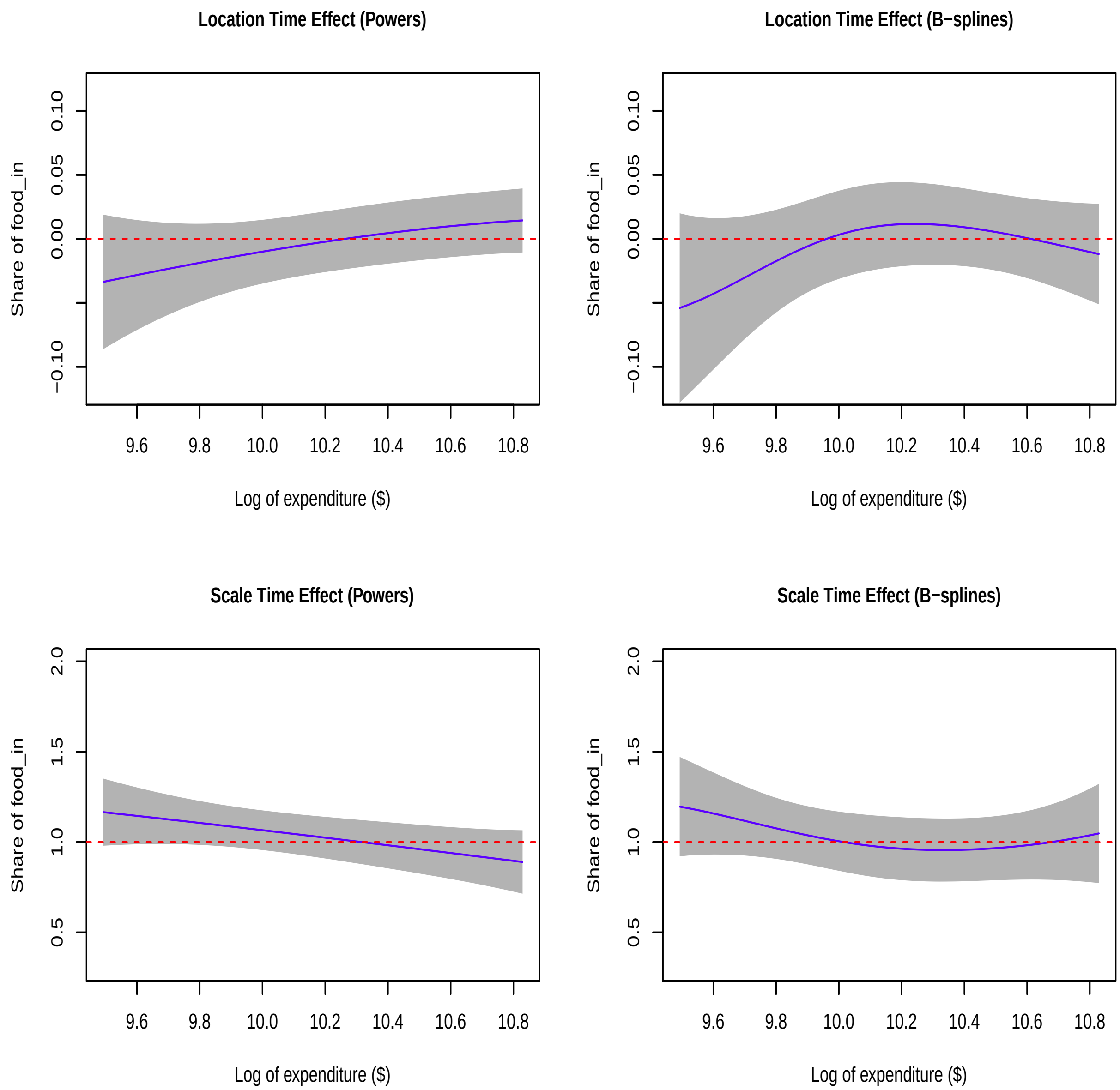

Figure 5: Location and scale time effects for food at home share: estimates from conditional quantiles.

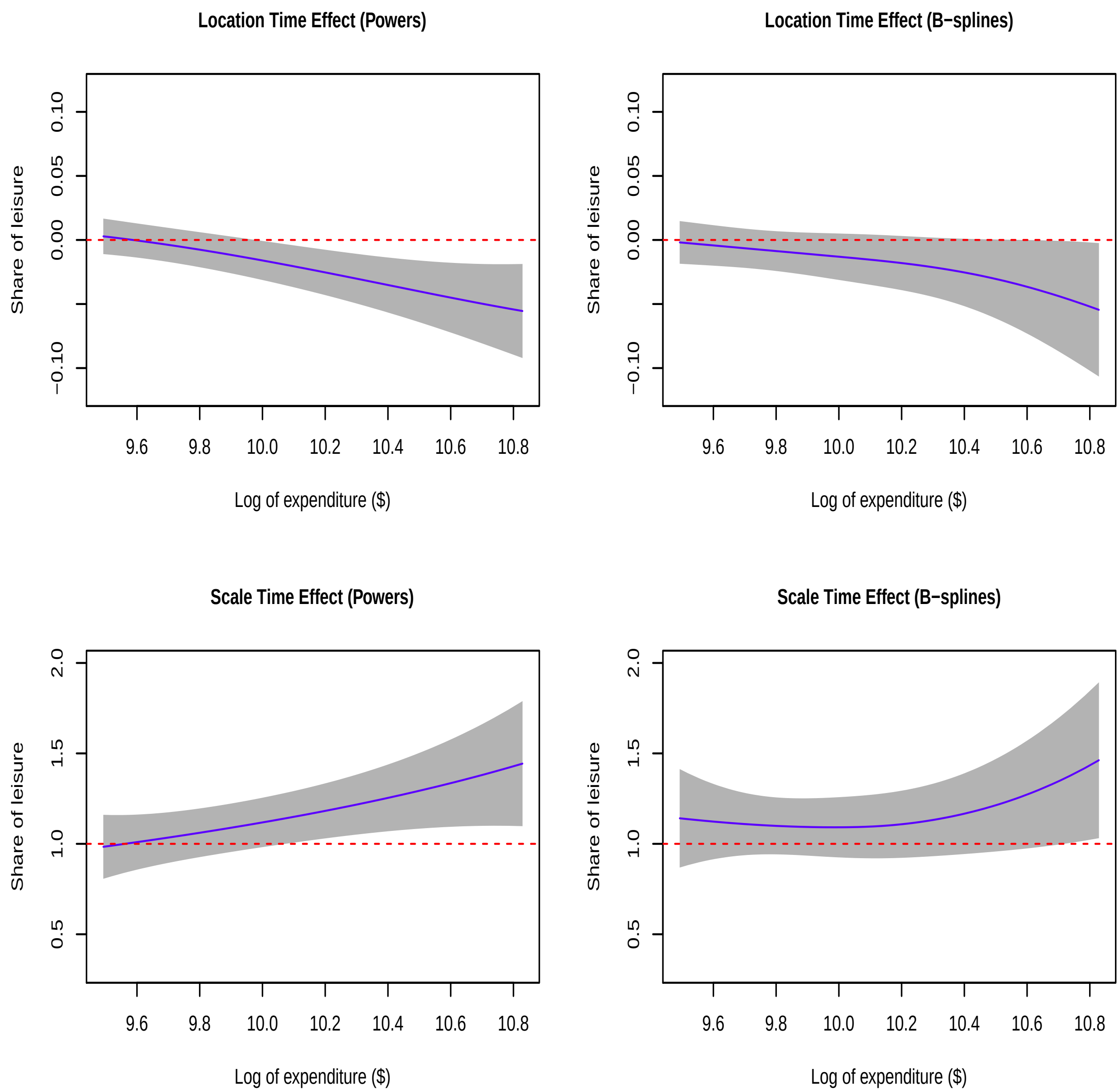

Figure 6: Location and scale time effects for leisure share: estimates from conditional quantiles.

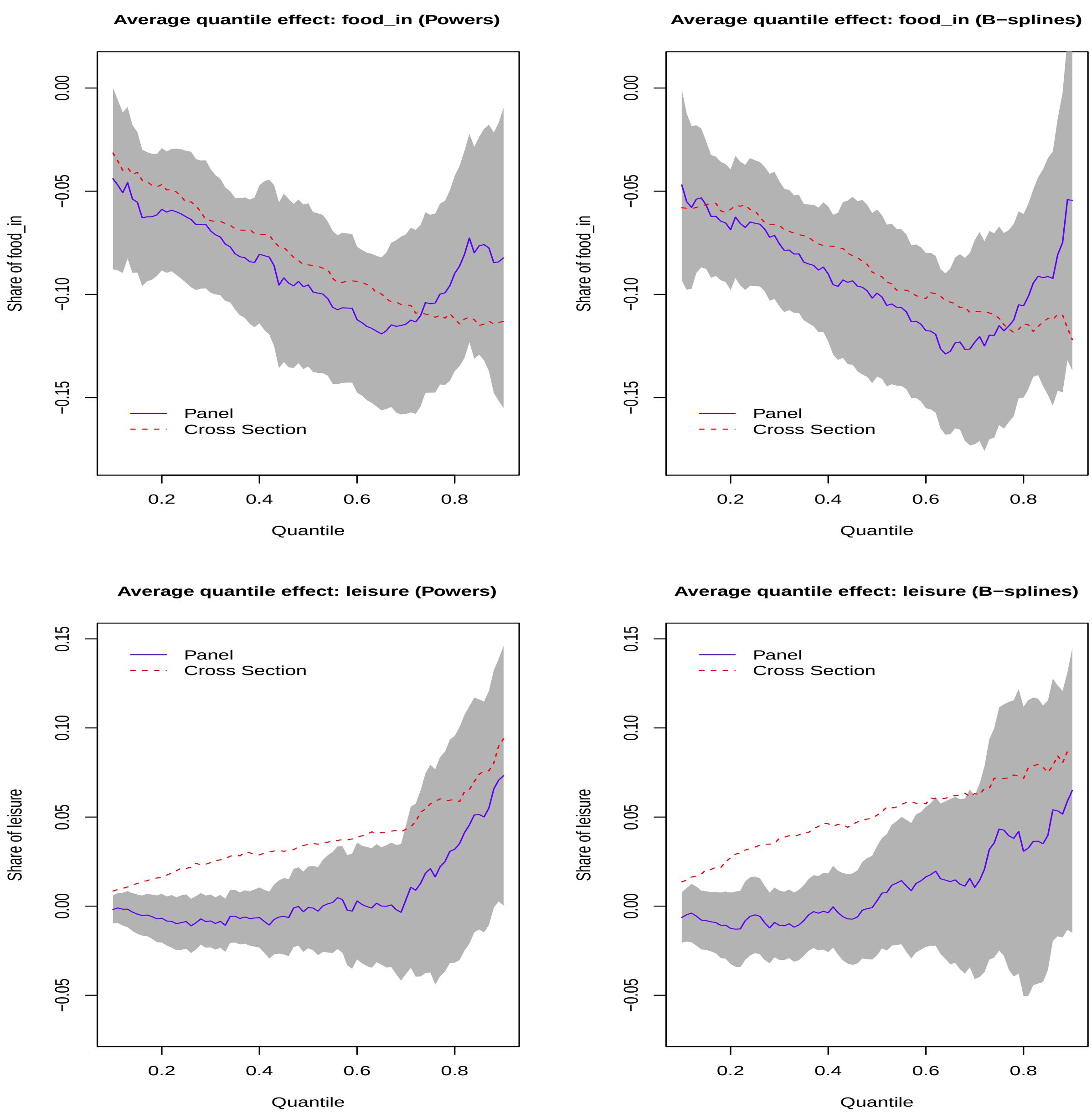

Figure 7: Average conditional quantile effects of log total expenditure.

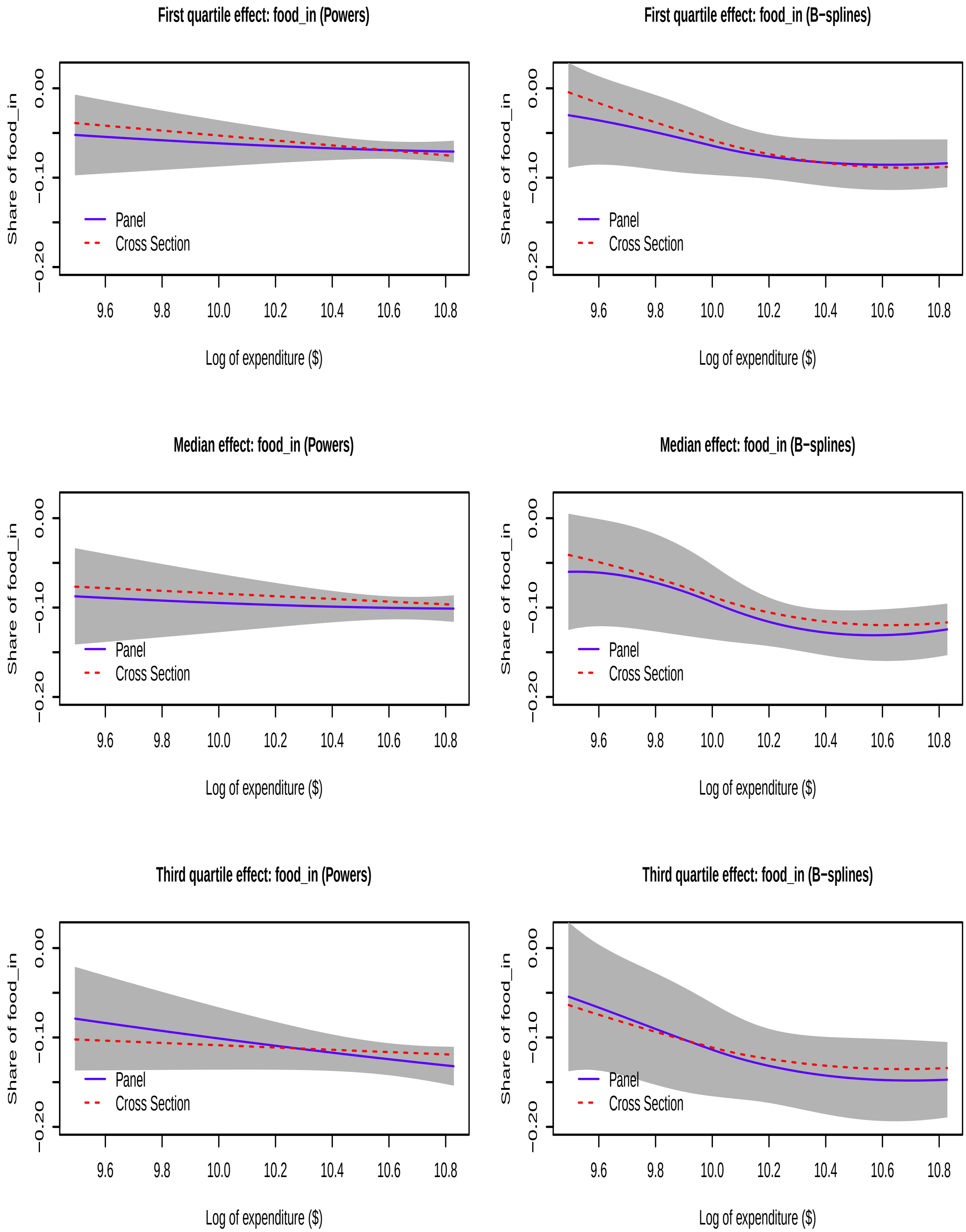

Figure 8: Conditional quartile effects of log total expenditure on food at home share.

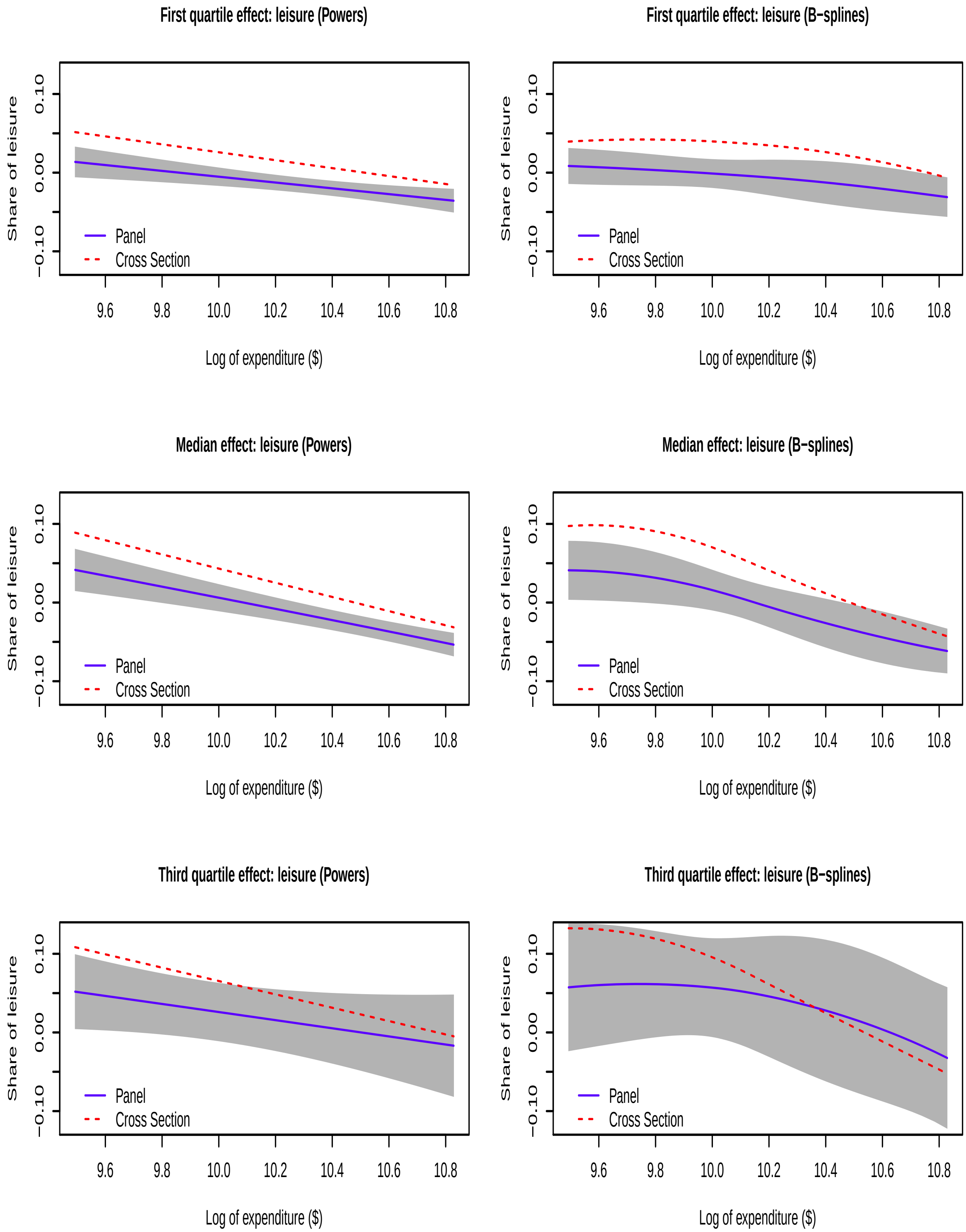

Figure 9: Conditional quartile effects of log total expenditure on leisure share.