

Estimation of a Transformation Model with Truncation, Interval Observation and Time-Varying Covariates*

Bo E. Honoré[†] LuoJia Hu[‡]

First complete version: December 2007.

This Version: November 2008.

Abstract

Abrevaya (1999b) considered estimation of a transformation model in the presence of left-truncation. This paper observes that a cross-sectional version of the statistical model considered in Frederiksen, Honoré, and Hu (2007) is a generalization of the model considered by Abrevaya (1999b) and the generalized model can be estimated by a pairwise comparison version of one of the estimators in Frederiksen, Honoré, and Hu (2007). Specifically, our generalization will allow for discretized observations of the dependent variable and for piecewise constant time-varying explanatory variables.

*This research was supported by the National Science Foundation, the Gregory C. Chow Econometric Research Program at Princeton University, and the Danish National Research Foundation (through CAM at the University of Copenhagen). We thank seminar participants and Rice University and members of Princeton's Microeconomic Reading Group for comments.

[†]Mailing Address: Department of Economics, Princeton University, Princeton, NJ 08544-1021. Email: honore@Princeton.edu.

[‡]Mailing Address: Economic Research Department, Federal Reserve Bank of Chicago, 230 S. La Salle Street, Chicago, IL 60604. Email: lhu@frbchi.org.

1 Introduction

The transformation model

$$h(T_i^*) = g(X_i'\beta) + \varepsilon_i \quad (1)$$

is often used to model durations. In models like this it is important to allow for right censoring and sometimes also for left-truncation because the samples used in many applications include spells that are in progress at the start of the sample period. See Abrevaya (1999b).

It is also sometimes desirable to allow for the dependent variable to be discretized so that one only observes whether it falls in a particular interval, so the observed duration, T_i , would be t if $T_i^* \in (t-1, t]$. See Prentice and Gloeckler (1978) or Meyer (1990). Moreover, in duration models it is often interesting to allow for time-varying covariates, which are not easily directly incorporated into the transformation model (Flinn and Heckman (1982)).

The contribution of this paper is to specify a statistical model that allows for interval observations and time-varying covariates but which simplifies to a model with interval observations from (1) when the covariates are time invariant. We then propose an estimator for the parameters of the model. The estimator can be interpreted as a generalization of the one proposed in Abrevaya (1999b).

Consider first the transformation model, (1), with strictly increasing $h(\cdot)$ and $g(\cdot)$ and with ε_i independent of X_i and continuously distributed with full support. In this model

$$\begin{aligned} P(T_i^* > t | X_i) &= P(h(T_i^*) > h(t) | X_i) \\ &= P(g(X_i'\beta) + \varepsilon_i > h(t) | X_i) \\ &= 1 - F(h(t) - g(X_i'\beta)) \end{aligned}$$

where F is the CDF for ε_i . This gives

$$P(T_i^* > t | X_i, T_i^* > t-1) = \frac{1 - F(h(t) - g(X_i'\beta))}{1 - F(h(t-1) - g(X_i'\beta))}$$

where the assumption that ε_i has full support guarantees that the denominator is not 0.

When $1 - F(\cdot)$ is log-concave (which is implied by the density of ε_i being log-concave; see Heckman and Honoré (1990)), the right-hand side is an increasing function of $g(X_i'\beta)$ and

hence of $X'_i\beta$. See the Appendix. This means that one can write the event $T_i^* > t | X_i, T_i^* > t - 1$ in the form $1\{X'_i\beta > \eta_{it}\}$ for some (possibly infinite) random variable η_{it} which is independent of X_i and has CDF $\frac{1-F(h(t)-\cdot)}{1-F(h(t-1)-\cdot)}$. Therefore, if we define

$$Y_{it} \equiv 1\{T_i^* \in (t-1, t]\} = 1\{T_i = t\}$$

then we can write

$$Y_{it} = 1\{X'_i\beta - \eta_{it} \leq 0\} \text{ for } t \text{ such that } \sum_{l \leq t-1} Y_{il} = 0 \quad (2)$$

In other words, a transformation model with discretized observations of the dependent variable and log-concave errors is a special case of the model

$$Y_{it} = 1\{X'_{it}\beta - \eta_{it} \leq 0\} \text{ for } t \text{ such that } \sum_{l \leq t-1} Y_{il} = 0 \quad (3)$$

where the difference between (2) and (3) is that the latter allows for time-varying covariates. Note that this line of argument is valid even if T_i is left truncated.

It is interesting to note that Abrevaya (1999b) also assumes log-concavity of $1 - F(\cdot)$. Note that although log-concavity implies an increasing hazard for $h(T^*)$, it does not impose such a restriction on T^{*1} .

Equation (3) is a cross-sectional version of the model considered by Frederiksen, Honoré, and Hu (2007). It is well understood that estimators of panel data models can be turned into estimators of cross-sectional models by considering all pairs of observations as units in a panel. See, for example, Honoré and Powell (1994). The insights in Frederiksen, Honoré, and Hu (2007) can therefore be used to construct an estimator of β . We pursue this in section 3 after formally defining the model in section 2.

¹Let $\lambda(\cdot)$ denote the hazard for ε . The hazard for $h(T^*)$ is then $-\frac{\partial \log(P(h(T^*) > t | X))}{\partial t} = -\frac{\partial \log(1-F(t-g(X'\beta)))}{\partial t} = \lambda(t-g(X'\beta))$. When $1-F(\cdot)$ is log-concave this is an increasing function of t . On the other hand, the hazard for T^* is $-\frac{\partial \log(P(T^* > t | X))}{\partial t} = -\frac{\partial \log(P(h(T^*) > h(t) | X))}{\partial t} = \lambda(h(t)-g(X'\beta))h'(t)$. The derivative of this with respect to t is $\lambda'(h(t)-g(X'\beta))h'(t)^2 + \lambda(h(t)-g(X'\beta))h''(t)$, which can be of either sign.

2 The model

Consider a spell with integer-valued duration, T_i , that starts at (integer-valued) time $-V_i \leq 0$. Following the discussion above, we model the event that the spell lasts at most s periods conditional on lasting at least $s - 1$, by the qualitative response model

$$Y_{is} = 1 \{X'_{is}\beta - \eta_{is} \leq 0\}, \quad (4)$$

where η_{is} is independent of X_{is} and the distribution of η_{is} is allowed to change over time.

When there is left truncation, one must distinguish between duration time and calendar time. We will index the observables, Y and X , by calendar time, and the unobservable η by duration time. At first sight, this difference is confusing, but it is necessitated by the fact that the discussion in the previous section implied that one should allow the distribution of η_{is} to vary by duration time. On the other hand, it seems natural to denote the first observation for an individual by $t = 1$.

With this notation, we assume that we observe (Y_{it}, X_{it}) starting at $t = 1$, where

$$Y_{it} = 1 \{X'_{it}\beta - \eta_{i,t+V_i} \leq 0\}. \quad (5)$$

With this notation, η 's with the same time subscript will have the same distribution under the class of models discussed above. We will let T_i denote the first time that Y_{it} equals 1. Since Y_{it} is not defined after the end of the spell, and since we want to allow for random right-censoring, we assume that we observe Y_{it} from $t = 1$ until, and including, T_i or until a random censoring time $C_i - 1$ (whichever comes first). In other words, we observe (Y_{it}, X_{it}) for $t = 1, 2, \dots, \bar{T}_i$ where $\bar{T}_i = \min \{T_i, C_i - 1\}$. So when an observation is censored, C_i will be the first time period in which individual i is not observed. We also assume that we observe the presample duration, V_i , for each observation.

The statistical assumption on the errors in (5) is that conditional on V_i , $\eta_{i,t+V_i}$ is independent of $(C_i, \{\eta_{i,s+V_i}\}_{s < t}, \{X_{is}\}_{s \leq t})$. This does not require the covariates to be strictly exogenous, and it allows for censoring to be covariate-dependent.

As discussed in the introduction, this is consistent with an underlying transformation model for T_i^* , where we observe whether a spell that started at time $-V_i$ and was in progress at time $t - 1$ is still in progress at time t .

In the next section we will apply the insight of Frederiksen, Honoré, and Hu (2007) to construct an estimator for β under these assumptions when the researcher has access to a random sample of individuals.

3 The estimator

The key insight for the construction of the estimator can be easily illustrated if we ignore censoring first (so $\bar{T}_i = T_i$ for all i).

Let t_1 and t_2 be arbitrary. Consider the two events $A = \{T_i = t_1, T_j > t_2\}$ and $B = \{T_i > t_1, T_j = t_2\}$ where $t_1 + V_i = t_2 + V_j$. Then under Assumptions 1-3, it follows immediately from Lemma 1 of Frederiksen, Honoré, and Hu (2007) that

$$P(A|A \cup B, X_{it_1}, X_{jt_2}, V_i, V_j) \begin{cases} > \frac{1}{2} & \text{if } (X_{it_1} - X_{jt_2})' \beta > 0, \\ = \frac{1}{2} & \text{if } (X_{it_1} - X_{jt_2})' \beta = 0, \\ < \frac{1}{2} & \text{if } (X_{it_1} - X_{jt_2})' \beta < 0. \end{cases}$$

This suggests estimating β by maximizing

$$\sum_{i < j} \sum_{t_1=1}^{T_i} \sum_{t_2=1}^{T_j} 1 \{t_1 + V_i = t_2 + V_j\} \cdot (1 \{T_i = t_1, T_j > t_2\} \cdot 1 \{(X_{it_1} - X_{jt_2})' b > 0\} + 1 \{T_i > t_1, T_j = t_2\} \cdot 1 \{(X_{it_1} - X_{jt_2})' b < 0\}) \quad (6)$$

(6) is the same as one of the objective functions in Frederiksen, Honoré, and Hu (2007), except that that paper considers a panel data situation.

It is convenient to rewrite (6) as

$$\sum_{i < j} 1 \{T_j + V_j > T_i + V_i > V_j\} \cdot 1 \{(X_{iT_i} - X_{j, T_i + V_i - V_j})' b > 0\} + 1 \{V_i < T_j + V_j < T_i + V_i\} \cdot 1 \{(X_{iT_j + V_j - V_i} - X_{j, T_j})' b < 0\} \quad (7)$$

This has the same structure as Han (1987)'s maximum rank correlation estimator.

When there is censoring, (6) can be modified to

$$\begin{aligned} & \sum_{i < j} \sum_{t_1=1}^{\bar{T}_i} \sum_{t_2=1}^{\bar{T}_j} 1 \{t_1 + V_i = t_2 + V_j, t_1 < C_i, t_2 < C_j\} \\ & \cdot (1 \{T_i = t_1, T_j > t_2\} \cdot 1 \{(X_{it_1} - X_{jt_2})' b > 0\} + 1 \{T_i > t_1, T_j = t_2\} \cdot 1 \{(X_{it_1} - X_{jt_2})' b < 0\}). \end{aligned} \quad (8)$$

And equation (7) can be rewritten as

$$\begin{aligned} & \sum_{i < j} 1 \{\bar{T}_j + V_j > T_i + V_i > V_j, T_i < C_i\} \cdot 1 \{(X_{iT_i} - X_{j, T_i + V_i - V_j})' b > 0\} \\ & + 1 \{V_i < T_j + V_j < \bar{T}_i + V_i, T_j < C_j\} \cdot 1 \{(X_{iT_j + V_j - V_i} - X_{j, T_j})' b < 0\}. \end{aligned} \quad (9)$$

The intuition for the estimator is essentially based on pairwise comparisons. Specifically, we compare an individual i who was observed to fail at time T_i (and thus had a complete duration $T_i + V_i$) to all other observations j that survived up to the same duration. At the true parameter value β , if the index for individual i at the time he/she failed, $X'_{iT_i}\beta$, is larger than the index for the comparable individual j at the time that corresponds to the same duration, $X'_{j, T_i + V_i - V_j}\beta$, then individual j is likely to survive longer than individual i (that is, $\bar{T}_j + V_j > T_i + V_i$). Note the set of comparison observations j includes censored spells provided the censoring time occurs after $T_i + V_i - V_j$. The additional inequality in the indicator function, $T_i + V_i > V_j$, ensures that the time at which j is being compared to i is within the sample period (i.e., not truncated).

Again this estimator has the same structure as Han (1987)'s maximum rank correlation estimator and the asymptotic distribution is therefore the one given in Sherman (1993) under the regularity conditions stated there.

4 Asymptotic properties

Consistency and asymptotic normality can be established as in Sherman (1993), Abrevaya (1999b) or Khan and Tamer (2007).

First note that some normalization of the parameter is needed, since the parameter vector is only identified up to scale. For example, we can normalize the last component of β to be 1.

The two key assumptions for consistency of the estimator are (1) at least one component of the explanatory variables X is continuously distributed with full support, and (2) the error η has full support. Without the first assumption, the parameter is not identified, since a small change in the parameter value could leave the ranking of the index unchanged. The second assumption on the error guarantees that the set of effective observations that make a nonzero contribution to the objective function is not empty. Both assumptions are standard in the semi-parametric estimation literature.

To establish asymptotic normality results, we need some additional notations. Denote $D_i = 1 \{T_i < C_i\}$, which is an observable variable indicating a complete (uncensored) spell.

The objective function can be rewritten as

$$\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} D_i \cdot 1 \{T_i + V_i > V_j, T_j + V_j > V_i, \bar{T}_j + V_j > T_i + V_i\} \cdot 1 \{X'_{iT_i} b > X'_{j, T_i + V_i - V_j} b\}.$$

Define the function

$$\begin{aligned} & \tau((t, \bar{t}, d, v, \{x_s\}_{s \leq \bar{t}}), b) \\ \equiv & E [D_i \cdot 1 \{T_i + V_i > v, t + v > V_i, \bar{t} + v > T_i + V_i\} \cdot 1 \{X'_{iT_i} b > x'_{T_i + V_i - v} b\}] \\ & + E [d \cdot 1 \{t + v > V_j, T_j + V_j > v, \bar{T}_j + V_j > t + v\} \cdot 1 \{x'_t b > X'_{j, t + v - V_j} b\}] \\ = & E [D_i \cdot 1 \{T_i + V_i > v, t + v > V_i, \bar{t} + v > T_i + V_i\} \cdot 1 \{X'_{iT_i} b > x'_{T_i + V_i - v} b\}] \\ & + E [d \cdot 1 \{t + v > V_i, T_i + V_i > v, \bar{T}_i + V_i > t + v\} \cdot 1 \{x'_t b > X'_{i, t + v - V_i} b\}]. \end{aligned}$$

Following Theorem 4 of Sherman (1993), we have

$$\sqrt{n} (\hat{\beta} - \beta) \xrightarrow{\sim} N(0, 4\Gamma^{-1} \Delta \Gamma^{-1}) \quad (10)$$

where

$$\begin{aligned}\Gamma &= E \left[\nabla_2 \tau \left((T_i, \bar{T}_i, D_i, V_i, \{X'_{is}\}) , \beta \right) \right] \\ \Delta &= E \left[\nabla_1 \tau \left((T_i, \bar{T}_i, D_i, V_i, \{X'_{is}\}) , \beta \right) \nabla_1 \tau \left((T_i, \bar{T}_i, d_i, V_i, \{X'_{is}\}) , \beta \right)' \right]\end{aligned}$$

with ∇_1 and ∇_2 denoting the first- and second-derivative operator, respectively.

Following Sherman (1993), we can further express the variance-covariance matrix in terms of “model primitives.” Specifically,

$$\Delta = V_X \left[\sum_{s_1, s_2} (\mathcal{X}_{s_2} - \mu_{s_1}(X'_{s_2}\beta)) S(T, \bar{T}, D, V, s_1, s_2, X'_{s_2}\beta) g_{X'_{s_1}\beta}(X'_{s_2}\beta) \right]$$

and

$$\Gamma = E_X \left[\sum_{s_1, s_2} (\mathcal{X}_{s_2} - \mu_{s_1}(X'_{s_2}\beta)) (\mathcal{X}_{s_2} - \mu_{s_1}(X'_{s_2}\beta))' \right. \\ \left. S_7(T, \bar{T}, D, V, s_1, s_2, X'_{s_2}\beta) \cdot g_{X'_{s_1}\beta}(X'_{s_2}\beta) \right]$$

where \mathcal{X}_{s_2} is composed of the first $K - 1$ coordinates of X'_{s_2} (the ones corresponding to the piece of β that is not normalized to 1), $g_{X'_{s_1}\beta}(\lambda)$ is the marginal density of $X'_{s_1}\beta$,

$$\mu_{s_1}(\lambda) = E \left[\mathcal{X}_{s_1} | X'_{s_1}\beta = \lambda \right]$$

and

$$S((t, \bar{t}, d, v, s_1, s_2), \lambda) = E \left[A_i(t, \bar{t}, d, v, s_1, s_2) | X'_{is_1}\beta = \lambda \right]$$

and

$$\begin{aligned} & A_i(t, \bar{t}, d, v, s_1, s_2) \\ &= D_i \cdot 1 \{ T_i + V_i > v, t + v > V_i, \bar{t} + v > T_i + V_i, T_i = s_1, T_i + V_i - v = s_2 \} \\ & \quad - d \cdot 1 \{ t + v > V_i, T_i + V_i > v, \bar{T}_i + V_i > t + v, t = s_2, t + v - V_i = s_1 \} \end{aligned}$$

The asymptotic variance matrix can be estimated by plugging in the estimator $\hat{\beta}$ and calculating sample analogs of Γ and Δ using numerical derivatives based on a smoothed version of τ . See Section 6 for more discussion.

5 Relationship to other estimators

The estimator proposed in the previous section is related to a number of existing estimators, and it coincides with some of them in special situations. For example, when $C_i = 2$ for all i , and with no left-truncation (so $V_i = 0$ for all i), (5) is a standard discrete choice model, and in that case the objective function in (9) becomes

$$\sum_{i < j}^n 1 \{Y_i > Y_j\} \cdot 1 \{(X_i - X_j)' b > 0\} + 1 \{Y_i < Y_j\} \cdot 1 \{(X_i - X_j)' b < 0\}$$

which is the objective function for Han (1987)'s maximum rank correlation estimator.

When there is left-truncation, and the covariates are time invariant and there is no censoring, the estimator defined by maximizing (9) is the same as that in Abrevaya (1999b).

Khan and Tamer (2007) consider a model with left-*censoring*, whereas we consider left-truncation. Which of the two is more interesting obviously depends on the specific application. Left-truncation will, for example, be relevant if the duration of interest is the length of employment on a given job, and one has information² on a sample of workers observed between two fixed points in time. In this case the durations are left truncated, because spells that ended before the start of the sampling will not appear in the data. When there is neither left-censoring nor left-truncation, and when the covariates are time invariant, the estimator defined by maximizing (9) coincides with the estimator proposed by Khan and Tamer (2007), except that ours applies to discretized durations and theirs to exactly measured durations.

The framework here is also closely related to standard statistical duration models with discretized observations. The proportional hazard model can be written as

$$Z(T) = -x'\beta + \varepsilon,$$

where Z is the log integrated baseline hazard and ε has an extreme value distribution. Prentice and Gloeckler (1978) and Meyer (1990) study a version of this model with interval observations. Meyer (1990) also allows for time-varying explanatory variables and for ε to be

²Including the duration of employment in the current job at the start of the sampling.

a sum of an extreme value distributed random variable and a random variable that captures unobserved heterogeneity.

6 Monte Carlo experiment

In this section, we conduct a small scale Monte Carlo study to illustrate the proposed estimation method and investigate its finite sample performance. We also demonstrate how to conduct inference and examine how good an approximation the asymptotic distribution provides for finite samples.

The designs are based on the following :

- All the designs have two explanatory variables.
- $\beta = (\beta_1, \beta_2)' = (1, 2)'$. The parameter of interest is $\theta = \beta_2/\beta_1$. The fact that this is one dimensional greatly simplifies the computations.
- Time-varying intercept $\beta_0 = -4 + (s/10)^{1.2}$. This introduces duration dependence beyond the duration dependence introduced by the shape of F and by the choice of $h(\cdot)$.
- The time between the start of a spell and the first period of observation is uniformly distributed on the integers between 1 and 5.
- The censoring time is generated as the minimum of ten periods from the start of the spell and Q periods from the start of the sample, where Q is uniformly distributed on the integers between 1 and 8.

We consider a number of designs within this framework.

Design 1: Dynamic Probit.

The two explanatory variables are generated by *i.i.d.* draws from $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\right)$.

In this design, $\eta_{i,t}$ is *i.i.d* $N(0, 4)$.

Design 2: Transformation Model Hazard.

This design is set up as a generalization of a transformation model. Specifically, using the notation of Section 1, we assume that $h(u) = \log(u)$, $g(u) = u$, and $\varepsilon \sim N(0, 1)$. Using the derivation in Section 1, this yields

$$P(T_i = t) = P(T_i^* < t | \{X_{is}\}_{s \leq t}, T_i^* > t - 1) = 1 - \frac{1 - \Phi(\log(t) - X'_{it}\beta)}{1 - \Phi(\log(t-1) - X'_{it}\beta)}.$$

As in Design 1, the two explanatory variables are generated by *i.i.d.* draws from $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\right)$.

Design 3: Feedback.

Recall that our model does not require the explanatory variables to be strictly exogenous. In this design we therefore allow for feedback from the error η to future values of X . Specifically, we follow Design 1 except that the explanatory variables are defined by

- $X_{2t} = \eta_{t-1}$ for $t > 1$ and standard normal for $t = 1$.
- $X_{1t} = X_{2t} + N(0, 1)$.

Design 4: Covariate-Dependent Censoring and Truncation.

Our model allows censoring and truncation to be correlated with explanatory variables. In this design, we follow the basic structure of Design 1 but let censoring be defined by the outcome of a probit with explanatory variable X_{1t} .

Design 5: Dynamic Probit 2.

This design is like Design 1 except that

- $X_{2s} \sim N(0.5 - 0.2s, 1)$.
- $X_{1s} = 1 \{X_{2s} + N(0, 1) > 0\}$
- $\eta_{i,s} \sim N(0, 0.1 + (0.15s))$

The summary statistics for the five designs are reported in Table 1. For each design, 100,000 observations are drawn from the underlying data generating process. We then compute the fraction of the sample that is censored, the fraction that is truncated, and the mean and standard deviation of the underlying duration.

Below, we report Monte Carlo results for the point estimates of β as well as for the performance of tests statistics based on the asymptotic distribution in section 4. To do this, we estimate the components of the variance of the estimators by sample analogs of smoothed versions of the components.³

Recall that

$$\Gamma = E \left[\nabla_2 \tau \left((T_i, \bar{T}_i, D_i, V_i, \{X'_{is}\}), \beta \right) \right]$$

and

$$\Delta = E \left[\nabla_1 \tau \left((T_i, \bar{T}_i, D_i, V_i, \{X'_{is}\}), \beta \right) \nabla_1 \tau \left((T_i, \bar{T}_i, D_i, V_i, \{X'_{is}\}), \beta \right)' \right]$$

where

$$\begin{aligned} & \tau \left((t, \bar{t}, d, v, \{x_s\}_{s \leq \bar{t}}), b \right) \\ = & E \left[D_i \cdot 1 \{T_i + V_i > v, t + v > V_i, \bar{t} + v > T_i + V_i\} \cdot 1 \{X'_{iT_i} b > x'_{T_i+V_i-v} b\} \right] \\ & + E \left[d \cdot 1 \{t + v > V_i, T_i + V_i > v, \bar{T}_i + V_i > t + v\} \cdot 1 \{x'_t b > X'_{i,t+v-V_i} b\} \right]. \end{aligned}$$

We then estimate τ by the smoothed version

$$\begin{aligned} & \hat{\tau} \left((t, \bar{t}, d, v, \{x_s\}_{s \leq \bar{t}}), b \right) \\ = & \frac{1}{n} \sum_{i=1}^n D_i \cdot 1 \{T_i + V_i > v, t + v > V_i, \bar{t} + v > T_i + V_i\} \cdot \Phi \left(\frac{X'_{iT_i} b - x'_{T_i+V_i-v} b}{h} \right) \\ & + \frac{1}{n} \sum_{i=1}^n d \cdot 1 \{t + v > V_i, T_i + V_i > v, \bar{T}_i + V_i > t + v\} \cdot \Phi \left(\frac{x'_t b - X'_{i,t+v-V_i} b}{h} \right). \end{aligned}$$

³A recent paper by Subbotin (2007) has shown that the nonparametric bootstrap can be used to estimate the quantiles and variance of various maximum rank correlation estimators. The structure of our estimator is essentially the same as that of the maximum rank correlation estimators that he considers. We therefore conjecture that the bootstrap could have been used to estimate the variance in our case as well, although this would increase the computational burden.

Then

$$\begin{aligned}
& \widehat{\nabla_1 \tau} \left((t, \bar{t}, d, v, \{x_s\}_{s \leq \bar{t}}), b \right) \\
&= \frac{1}{nh} \sum_{i=1}^n D_i \cdot 1 \{T_i + V_i > v, t + v > V_i, \bar{t} + v > T_i + V_i\} \\
&\quad \phi \left(\frac{X'_{iT_i} b - x'_{T_i+V_i-v} b}{h} \right) \left(\tilde{X}'_{iT_i} - \tilde{x}'_{T_i+V_i-v} \right) \\
&\quad + d \cdot 1 \{t + v > V_i, T_i + V_i > v, \bar{T}_i + V_i > t + v\} \\
&\quad \phi \left(\frac{x'_t b - X'_{i,t+v-V_i} b}{h} \right) \left(\tilde{x}'_t - \tilde{X}'_{i,t+v-V_i} \right),
\end{aligned}$$

and

$$\begin{aligned}
& \widehat{\nabla_2 \tau} \left((t, \bar{t}, d, v, \{x_s\}_{s \leq \bar{t}}), b \right) \\
&= \frac{-1}{nh^3} \sum_{i=1}^n D_i \cdot 1 \{T_i + V_i > v, t + v > V_i, \bar{t} + v > T_i + V_i\} \\
&\quad (X'_{iT_i} b - x'_{T_i+V_i-v} b) \phi \left(\frac{X'_{iT_i} b - x'_{T_i+V_i-v} b}{h} \right) \\
&\quad \left(\tilde{X}'_{iT_i} - \tilde{x}'_{T_i+V_i-v} \right) \left(\tilde{X}'_{iT_i} - \tilde{x}'_{T_i+V_i-v} \right)' \\
&\quad + d \cdot 1 \{t + v > V_i, T_i + V_i > v, \bar{T}_i + V_i > t + v\} \\
&\quad (X'_{iT_i} b - x'_{T_i+V_i-v} b) \phi \left(\frac{x'_t b - X'_{i,t+v-V_i} b}{h} \right) \\
&\quad \left(\tilde{x}'_t - \tilde{X}'_{i,t+v-V_i} \right) \left(\tilde{x}'_t - \tilde{X}'_{i,t+v-V_i} \right)'.
\end{aligned}$$

And therefore

$$\widehat{\Gamma} = \frac{1}{n} \sum_{i=1}^n \widehat{\nabla_2 \tau} \left((T_i, \bar{T}_i, D_i, V_i, \{X'_{is}\}), \widehat{\beta} \right)$$

and

$$\widehat{\Delta} = \frac{1}{n} \sum_{i=1}^n \widehat{\nabla_1 \tau} \left((T_i, \bar{T}_i, D_i, V_i, \{X'_{is}\}), \widehat{\beta} \right) \widehat{\nabla_1 \tau} \left((T_i, \bar{T}_i, D_i, V_i, \{X'_{is}\}), \widehat{\beta} \right)'.$$

As mentioned earlier, β is only identified up to scale. One possibility is to normalize one of the coefficients to 1, and hence essentially focus on β_2/β_1 or β_1/β_2 . Unfortunately, this normalization will lead to different MAE and RMSE depending on which of the coefficients is normalized. So if one were to compare different estimators, one might reach different

conclusions depending on a seemingly innocent normalization. This is unsatisfactory in models where there is only one parameter. For this reason, it is likely to be better to consider $\theta = \log(\beta_2/\beta_1) = \log(\beta_2) - \log(\beta_1)$ the parameter of interest. This means that the true parameter is $\log(2) \approx 0.693$ for all the designs. We estimate θ by a grid search over the interval between $-\log(6)$ and $\log(6)$ with equal grids of size $\frac{1}{200}$. Since the parameter of interest is one dimensional, the line search is feasible despite the fact that the calculation of the objective function requires $O(n^2)$ operations. When it is of higher dimension, it would be beneficial to use the method described in Abrevaya (1999a) to calculate the objective function in $O(n \cdot \log(n))$ operations. We calculate the variance of θ by applying the so-called δ -method to (10).

For each design, the Monte Carlo experiment is conducted with 5,000 replications for each of the 5 sample sizes: 100, 200, 400, 800 and 1,600. The results are reported in Tables 2–6. Overall, the results across the designs are broadly consistent with predictions from the asymptotic theory. Some additional remarks are in order.

First, the results illustrate the consistency of the estimator, since both the median absolute error (MAE) and root mean squared error (RMSE) decrease as sample size increases. Moreover, the estimator is close to median unbiased even for small sample sizes.

Second, the theory predicts that the estimator converges to the true parameter value at the rate \sqrt{n} . This is borne out in the simulation as the MAE and RMSE decrease toward zero at a rate of approximately $\sqrt{2}$ when the sample size is doubled. For example, a regression of the log of the median absolute error on the log of the sample size (and design dummies) yields a coefficient of -0.543 with a standard error of 0.006.

Third, to examine the normality prediction from the asymptotic theory, we estimate the density for $\hat{\theta} - \theta$ and plot the kernel estimate in Figures 1-5. The left-hand side of each figure gives the estimated density of the estimator of (β_2/β_1) centered at the true value. They show severe asymmetry in the distribution of the estimator: it tends to be skewed to the left, especially in small samples. As mentioned, this is expected due to the somewhat

unnatural normalization. The right-hand side of the figures shows the estimated density of the estimator of $\log(\beta_2/\beta_1)$, again centered at the true value. There one can see that the distribution becomes more symmetric and closer to normal as sample size increases.

Finally, the asymptotic theory suggests that we can conduct inferences using t-test. Under the null, the test statistic should follow a standard normal distribution. In the simulation, we compute a t-statistic for each of the 5,000 estimates $\hat{\theta}$ and calculate the fraction of times that the null is rejected at the 20% level. We focus on tests with (nominal) size of 20% rather than the conventional 5% because the results for the latter are likely to be more erratic for a finite number of simulations. The results are reported for various bandwidths that are used in the estimation of the asymptotic variance-covariance matrix of the estimator.

In general, the rejection rate is closer to the nominal size of the test when the bandwidth is smaller and sample size is larger. For example, for the bandwidth⁴ (0.05, 0.20) and sample size 1,600, the reject rate is 21.8%, 22.7% , 20.8% and 21.6% for Designs 1 through 4. These are close to being statistically indistinguishable from the nominal size. The performance of the test under some other combinations of bandwidth and sample size is less encouraging. The test also performs less well under Design 5. We speculate that this is because of the discreteness of x_{i1} .

The last row reports the reject rates computed using the average of the variance-covariance matrix estimated over all the bandwidth choices. Overall, the t-test tends to over-reject the null.

It is interesting to compare our results to a standard logit or probit estimation of (5) where one uses x_1 , x_2 and time dummies as explanatory variables. Since we expect them to perform comparably, we focus on the logit maximum likelihood estimator.⁵ Designs 1, 3 and 4 are all correctly specified probit models, so one would expect the logit estimator to do

⁴Different bandwidths are used in estimating the matrix Δ and Γ . The latter is based on a second derivative, and one would therefore expect it to require a larger bandwidth than the former.

⁵Since our estimator of θ was calculated by a grid search over the interval between $-\log(6)$ and $\log(6)$, we censored the logit maximum likelihood estimator of β_2/β_1 to be in the interval between $\frac{1}{6}$ and 6.

well for this design. This is confirmed in panels one, three and four of Table 7. The bias is small and the MAE and RMSE fall at a rate close to $\text{root-}n$. It is less clear what to expect for Designs 2 and 5. Panel 2 of Table 7 shows that the logit estimator does well for Design 2. It appears to be close to unbiased and its MAE and RMSE fall at a rate close to $\text{root-}n$. One potential explanation for this is that misspecified maximum likelihood estimators often do well when the explanatory variables are jointly normally distributed. See, for example, Ruud (1983). Design 5 shows a situation where the logit estimator does relatively poorly. The bias is quite high, and as a result, the MAE and RMSE do not fall rapidly as the sample size increases.

7 Conclusion

In this paper we propose a generalization of the transformation model that is appropriate for studying duration outcome with truncation, censoring, interval observations of the dependent variable, and time-varying covariates. We develop an estimator for this model, discuss its asymptotic properties and investigate its finite sample performance via a Monte Carlo study. Overall, the results suggest that the estimator performs well in finite samples, and the asymptotic theory provides a reasonably good approximation to the distribution. We also investigate test-statistics for the estimator. Those require estimation of the asymptotic variance of the estimator. This is somewhat sensitive to different choices of bandwidth. Investigating the optimal bandwidth choice in this case could be an interesting future research topic.

References

- ABREVAYA, J. (1999a): “Computation of The Maximum Rank Correlation Estimator,” *Economics Letters*, 62, 279–285(7).

- (1999b): “Rank Estimation of a Transformation Model With Observed Truncation,” *Econometrics Journal*, 2(2), 292–305, available at <http://ideas.repec.org/a/ect/emjrn/v2y1999i2p292-305.html>.
- FLINN, C. J., AND J. J. HECKMAN (1982): “Models for the Analysis of Labor Force Dynamics,” in *Advances in Econometrics*, ed. by R. Basman, and G. Rhodes. JAI, Greenwich, Conn.
- FREDERIKSEN, A., B. E. HONORÉ, AND L. HU (2007): “Discrete Time Duration Models with Group-Level Heterogeneity,” *Journal of Econometrics*, 141, 1014–1043.
- HAN, A. (1987): “Nonparametric Analysis of a Generalized Regression Model,” *Journal of Econometrics*, 35, 303–316.
- HECKMAN, J. J., AND B. E. HONORÉ (1990): “The Empirical Content of the Roy Model,” *Econometrica*, 58, 1121–1149.
- HONORÉ, B. E., AND J. L. POWELL (1994): “Pairwise Difference Estimators of Censored and Truncated Regression Models,” *Journal of Econometrics*, 64, 241–78.
- KHAN, S., AND E. TAMER (2007): “Partial Rank Estimation of Duration Models with General Forms of Censoring,” *Journal of Econometrics*, 127(1), 251–280.
- MEYER, B. D. (1990): “Unemployment Insurance and Unemployment Spells,” *Econometrica*, 58, 757–782.
- PRENTICE, R. L., AND L. A. GLOECKLER (1978): “Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data,” *Biometrics*, 34, 57–67.
- RUUD, P. A. (1983): “Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinomial Discrete Choice Models,” *Econometrica*, 51(1), 225–228.

SHERMAN, R. T. (1993): “The Limiting Distribution of the Maximum Rank Correlation Estimator,” *Econometrica*, 61, 123–137.

SUBBOTIN, V. (2007): “Asymptotic and Bootstrap Properties of Rank Regressions,” *SSRN eLibrary*.

8 Appendix

Assume that $H(\cdot)$ is a log-concave function and let

$$f(w) = \frac{H(a_2 - w)}{H(a_1 - w)}$$

where $a_2 > a_1$. Let $w_1 < w_2$ and

$$\Delta a = a_2 - a_1, \quad \Delta w = w_2 - w_1 \quad \text{and} \quad \lambda = \frac{\Delta a}{\Delta a + \Delta w}$$

then

$$a_2 - w_2 = \lambda(a_2 - w_1) + (1 - \lambda)(a_1 - w_2)$$

so by concavity of $\ln(H(\cdot))$,

$$\ln(H(a_2 - w_2)) > \lambda \ln(H(a_2 - w_1)) + (1 - \lambda) \ln(H(a_1 - w_2)). \quad (11)$$

Also

$$a_1 - w_1 = (1 - \lambda)(a_2 - w_1) + \lambda(a_1 - w_2)$$

so

$$\ln(H(a_1 - w_1)) > (1 - \lambda) \ln(H(a_2 - w_1)) + \lambda \ln(H(a_1 - w_2)) \quad (12)$$

Adding (11) and (12) yields

$$\ln(H(a_2 - w_2)) + \ln(H(a_1 - w_1)) > \ln(H(a_2 - w_1)) + \ln(H(a_1 - w_2))$$

and

$$\begin{aligned} \ln(f(w_2)) - \ln(f(w_1)) &= (\ln(H(a_2 - w_2)) - \ln(H(a_1 - w_2))) \\ &\quad - (\ln(H(a_2 - w_1)) - \ln(H(a_1 - w_1))) \\ &> 0 \end{aligned}$$

Hence f is an increasing function.

TABLE 1: Summary Statistics for the Designs

	Design 1	Design 2	Design 3	Design 4	Design 5
Fraction Truncated	0.260	0.349	0.316	0.260	0.183
Fraction Censored	0.297	0.100	0.190	0.425	0.543
Mean Duration	5.317	3.592	4.369	5.317	7.270
Standard Deviation of Duration	3.422	2.023	2.918	3.422	3.937

TABLE 2: Results for Design 1

Performance of Estimator					
	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
Median	0.721	0.695	0.702	0.693	0.693
MAE	0.331	0.215	0.155	0.105	0.070
Mean	0.727	0.706	0.702	0.698	0.694
RMSE	0.488	0.331	0.227	0.155	0.106
Significance when testing at 20% level					
	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
0.05, 0.20	0.332	0.240	0.212	0.207	0.218
0.05, 0.40	0.163	0.152	0.187	0.241	0.271
0.10, 0.20	0.421	0.323	0.280	0.256	0.246
0.10, 0.40	0.243	0.232	0.261	0.296	0.302
0.20, 0.20	0.494	0.382	0.323	0.285	0.260
0.20, 0.40	0.324	0.301	0.305	0.326	0.319
0.40, 0.20	0.536	0.423	0.350	0.303	0.269
0.40, 0.40	0.390	0.347	0.337	0.343	0.329
average	0.289	0.259	0.261	0.274	0.273

TABLE 3: Results for Design 2

Performance of Estimator					
	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
Median	0.718	0.713	0.698	0.693	0.692
MAE	0.540	0.370	0.255	0.175	0.125
Mean	0.656	0.712	0.705	0.697	0.695
RMSE	0.752	0.553	0.392	0.274	0.186

Significance when testing at 20% level					
	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
0.05, 0.20	0.423	0.350	0.268	0.236	0.227
0.05, 0.40	0.164	0.175	0.180	0.224	0.265
0.10, 0.20	0.515	0.422	0.335	0.280	0.253
0.10, 0.40	0.269	0.252	0.245	0.273	0.296
0.20, 0.20	0.578	0.477	0.372	0.302	0.269
0.20, 0.40	0.359	0.317	0.293	0.303	0.313
0.40, 0.20	0.620	0.514	0.394	0.317	0.279
0.40, 0.40	0.426	0.363	0.322	0.319	0.323
average	0.333	0.299	0.274	0.273	0.275

TABLE 4: Results for Design 3

Performance of Estimator					
	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
Median	0.683	0.699	0.693	0.696	0.688
MAE	0.400	0.263	0.174	0.120	0.085
Mean	0.691	0.716	0.707	0.700	0.694
RMSE	0.569	0.405	0.271	0.185	0.124

Significance when testing at 20% level					
	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
0.05, 0.20	0.389	0.289	0.214	0.201	0.208
0.05, 0.40	0.216	0.150	0.143	0.202	0.255
0.10, 0.20	0.468	0.372	0.288	0.264	0.247
0.10, 0.40	0.280	0.226	0.225	0.273	0.302
0.20, 0.20	0.551	0.455	0.345	0.308	0.273
0.20, 0.40	0.368	0.307	0.287	0.316	0.327
0.40, 0.20	0.608	0.503	0.384	0.330	0.287
0.40, 0.40	0.438	0.370	0.325	0.344	0.342
average	0.326	0.268	0.247	0.266	0.273

TABLE 5: Results for Design 4

Performance of Estimator					
	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
Median	0.706	0.693	0.703	0.698	0.693
MAE	0.385	0.260	0.178	0.118	0.085
Mean	0.704	0.703	0.704	0.701	0.694
RMSE	0.557	0.397	0.273	0.182	0.124

Significance when testing at 20% level					
	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
0.05, 0.20	0.355	0.280	0.235	0.195	0.216
0.05, 0.40	0.176	0.161	0.186	0.219	0.270
0.10, 0.20	0.437	0.369	0.307	0.250	0.249
0.10, 0.40	0.250	0.235	0.255	0.269	0.300
0.20, 0.20	0.515	0.428	0.356	0.280	0.268
0.20, 0.40	0.335	0.305	0.307	0.305	0.320
0.40, 0.20	0.568	0.472	0.383	0.296	0.280
0.40, 0.40	0.402	0.353	0.340	0.324	0.333
average	0.300	0.272	0.270	0.259	0.276

TABLE 6: Results for Design 5

Performance of Estimator					
	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
Median	0.711	0.704	0.682	0.698	0.693
MAE	0.567	0.409	0.285	0.195	0.135
Mean	0.646	0.710	0.720	0.718	0.703
RMSE	0.789	0.612	0.438	0.305	0.207

Significance when testing at 20% level					
	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
0.05, 0.20	0.518	0.414	0.365	0.355	0.372
0.05, 0.40	0.506	0.464	0.567	0.637	0.661
0.10, 0.20	0.548	0.485	0.429	0.399	0.401
0.10, 0.40	0.528	0.557	0.638	0.666	0.677
0.20, 0.20	0.607	0.541	0.469	0.425	0.411
0.20, 0.40	0.584	0.618	0.672	0.681	0.683
0.40, 0.20	0.631	0.566	0.480	0.431	0.412
0.40, 0.40	0.627	0.648	0.682	0.684	0.683
average	0.511	0.503	0.505	0.508	0.513

TABLE 7: Performance of the Logit MLE

Design 1					
	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
Median	0.699	0.699	0.692	0.694	0.693
MAE	0.240	0.165	0.114	0.084	0.059
Mean	0.715	0.702	0.696	0.695	0.693
RMSE	0.369	0.252	0.173	0.123	0.085

Design 2					
	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
Median	0.696	0.703	0.692	0.689	0.695
MAE	0.413	0.282	0.199	0.139	0.100
Mean	0.691	0.720	0.701	0.694	0.695
RMSE	0.639	0.443	0.303	0.214	0.149

Design 3					
	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
Median	0.691	0.692	0.695	0.692	0.694
MAE	0.302	0.202	0.137	0.098	0.069
Mean	0.711	0.710	0.701	0.696	0.695
RMSE	0.454	0.307	0.209	0.144	0.103

Design 4					
	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
Median	0.687	0.695	0.696	0.696	0.692
MAE	0.288	0.201	0.138	0.097	0.067
Mean	0.706	0.702	0.701	0.699	0.693
RMSE	0.438	0.302	0.209	0.146	0.100

Design 5					
	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
Median	0.884	1.008	1.046	1.055	1.063
MAE	0.521	0.406	0.361	0.362	0.370
Mean	0.761	1.012	1.080	1.082	1.073
RMSE	0.948	0.685	0.540	0.476	0.428

Figure 1: Density of Estimation Error for Estimator and Its Log for Design 1

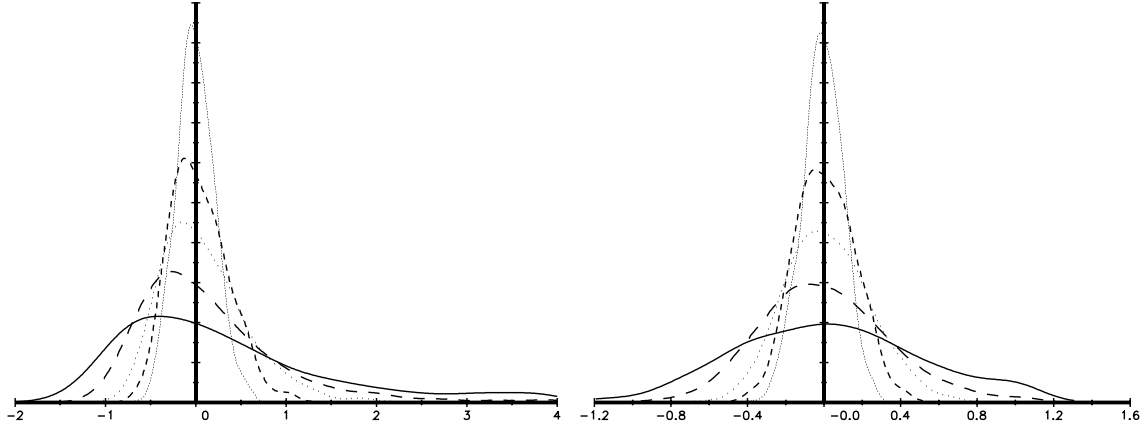


Figure 2: Density of Estimation Error for Estimator and Its Log for Design 2

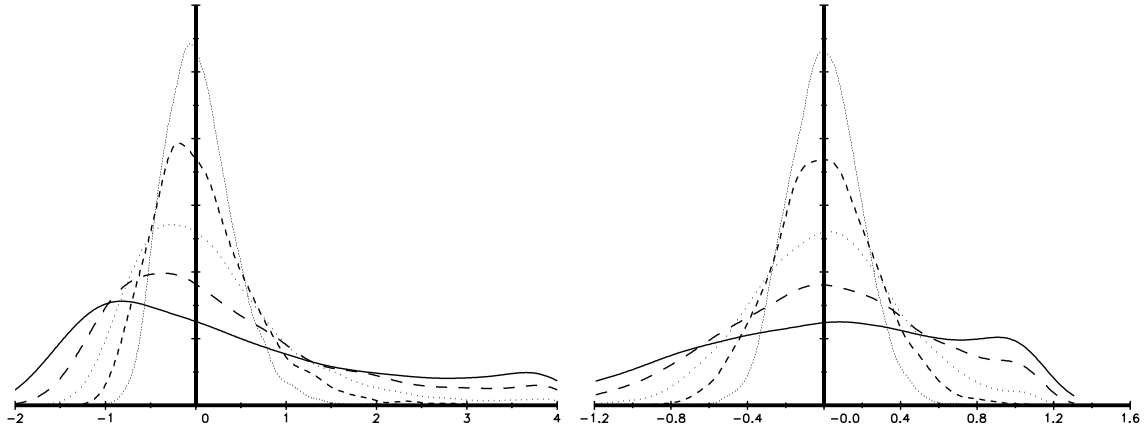


Figure 3: Density of Estimation Error for Estimator and Its Log for Design 3

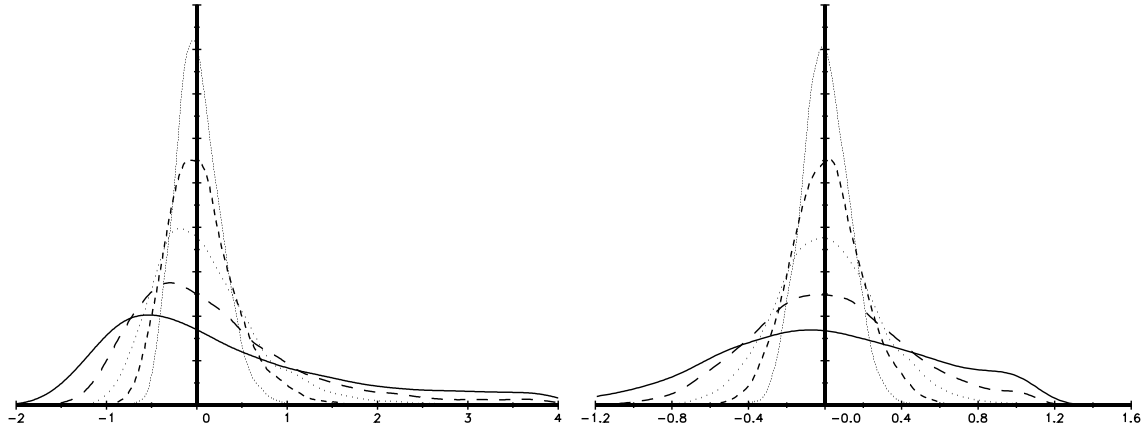


Figure 4: Density of Estimation Error for Estimator and Its Log for Design 4

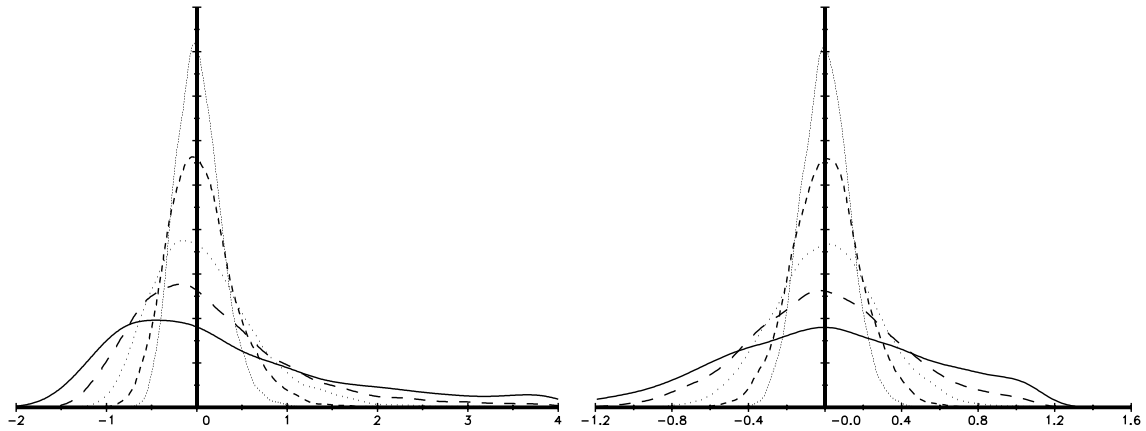


Figure 5: Density of Estimation Error for Estimator and Its Log for Design 5

