I Z A

# GTL Regression: A Linear Model with Skewed and Thick-Tailed Disturbances

Wim Vijverberg
Takuya Hasebe

# GTL Regression: A Linear Model with Skewed and Thick-Tailed Disturbances

**Wim Vijverberg**
*City University of New York Graduate Center
and IZA*

**Takuya Hasebe**
*Sophia University
and University of California, Davis*

# ABSTRACT

# GTL Regression:
# A Linear Model with Skewed and Thick-Tailed Disturbances

If the disturbances of a linear regression model are skewed and/or thick-tailed, a maximum likelihood estimator is efficient relative to the customary Ordinary Least Squares (OLS) estimator. In this paper, we specify a highly flexible Generalized Tukey Lambda (GTL) distribution to model skewed and thick-tailed disturbances. The GTL-regression estimator is consistent and asymptotically normal. We demonstrate the potential gains of the GTL estimator over the OLS estimator in a Monte Carlo study and in five applications that are typical of applied economics research problems: log-wage equations, hedonic housing price equations, an analysis of speeding tickets, the issue of trade creation and trade diversion that result from preferential trade agreements, and the familiar CAPM model in financial economics.

Corresponding author:

Wim Vijverberg,
PhD Program in Economics
CUNY Graduate Center
365 5th Avenue
New York, NY 10016-4309
USA
E-mail: wvijverberg@gc.cuny.edu

# 1 Introduction

In estimating models of economic behavior, researchers pay more attention to the specification of the systematic component of the regression model than to the disturbances. After all, the researcher has full control over the manner in which the systematic component is built up from observable factors. The disturbance combines all unobservables in a single composite term that is both a nuisance because it prevents a perfect explanation of the outcome variable and a blessing because it permits a parsimonious representation of ignorance.

A researcher may resort to classical assumptions about the disturbances (independently and identically distributed with a zero mean and a constant finite variance), or he may choose to describe this aggregate disturbance factor with familiar tools such as heteroskedasticity, serial correlation, and ARCH and GARCH modeling. Such tools address patterns in the behavior among the disturbances. But consider a situation where such patterns do not exist: what options exist then? What is actually the nature of the distribution of the disturbances? The classical assumptions imply that the Ordinary Least Squares (OLS) estimator is efficient within the class of linear unbiased estimators. Moreover, if the disturbances are normally distributed, the OLS estimator coincides with the Maximum Likelihood (ML) estimator, which has optimality properties itself by virtue of the Cramer-Rao theorem. However, normality of the disturbances is not guaranteed by economic theory; in fact, economic theory rarely has anything to say about the distribution of the disturbances. Rather, the researcher is implicitly relying on the Central Limit Theorem, assuming that many unobservables play a role and none is dominant, thus yielding an approximately normally distributed aggregate disturbance. But that is an untested assumption.

This paper addresses the situation where disturbances are independently and identically distributed but are not "standard": they may be skewed; they may have long or short tails; and their moments may not even be defined. The Ordinary Least Squares estimator

may not work well under these circumstances. We offer an ML estimator based on a highly flexible Generalized Tukey Lambda (GTL) distribution, that approximately nests the normal distribution but can also handle thick, skewed tails and nonexisting moments. We show that a flexible description of the distribution yields a more efficient estimator than OLS.

The literature offers several other approaches. First, tail data may be dealt with by means of data trimming or winsorizing (Chen and Dixon, 1972; Yale and Forsythe, 1976; Chen et al., 2001). These techniques view tail observations as data contaminations arising from a different data generating process; the interior of the scatterplot contains the information that is relevant for the process under focus. However, for processes with thick-tailed data, while the interior of the scatterplot is indeed relevant for the location parameters, the tails are not considered contaminations but rather provide information about scale and shape (and also location) parameters. Moreover, trimming and winsorizing leads to biased parameter estimates if the disturbances are generated with a skewed distribution.

Second, M-estimation maximizes an objective function that may or may not coincide with the log-likelihood function but, more importantly, may be selected so as to downplay extreme values in order to reduce their impact on the slope (or other) estimates (Huber, 1964; Huber and Ronchetti, 2009). The least squares estimator that follows the trimming or winsorizing of the data may be seen as a special case of the M-estimator. Ultimately, the objective of M-estimation is to estimate particular characteristics of the data generating process, such as location (intercept and slope), dispersion, or skewness; the objective is not to describe the distribution of the disturbances *per se*. As a result, the information content of tail observations is not (fully) utilized. The same may be said about L-estimation, which formulates estimators from order statistics (or sample quantiles) and of which the Least Absolute Deviation estimator is a special case (Koenker and Bassett, 1978; Koenker and Portnoy, 1987; Koenker, 2005).

Third, by design, the semi-parametric estimation approach allows the distribution of the

disturbances to be more arbitrary. For example, the various semiparametric estimators of the single index model only require the existence of several moments of the dependent variable; at the very least, the variance of the disturbance must be finite.[1] Heavy-tailed disturbances, especially those for which higher-order moments do not exist, are likely to cause estimation problems for location parameters, and the nature of the distribution of the disturbances is not directly addressed.

The fourth approach is more direct, explicitly specifying the distribution of the disturbances. This paper fits within this approach. The current literature offers several alternatives. The first to come to mind is the Student's $t(\nu)$ distribution where variation in the degrees of freedom $\nu$ yields varying degrees of tail thickness. Other examples are the skewed-$t$, skewed generalized error, and asymmetric power distributions as well as a mixture of the beta and $t$ distributions (e.g., Fernandez and Steel, 1998; Ferreira and Steel, 2006; Komunjer, 2007; Harvey and Sucarrat, 2014).[2] Recently a series of studies rely on the family of so-called stable distributions, members of which are typically heavy-tailed and potentially skewed. Blattberg and Sargent (1971) and Samorodnitsky et al. (2007) develop a linear estimator for a model with a single explanatory variable in which both the disturbance and the explanatory variable are drawn from a stable distribution. Nolan and Ojeda-Revah (2013) develop an ML estimator of a multivariate linear (and nonlinear) model, and Hallin et al. (2011, 2013) use rank estimation as the strategy to deal with heavy-tailed stable disturbances. The family of stable distributions nests normality, but all other members do not have a second or higher moment and some even lack a first moment. All of these studies demonstrate the relative inefficiency of OLS estimators. Stable distributions are sometimes proposed as a way to model heavy-tailed disturbances in regression models especially in fi-

---

[1]For example, see Powell et al. (1989), Ichimura (1993), Carroll et al. (1997), Horowitz (1998), and Li and Racine (2007).

[2]However, Fernancez and Steel (1999) shows that the likelihood function of a $t(\nu)$-based regression model is unbounded for $\nu$ approaching 0, raising questions about models based on the $t$ distribution.

nance applications, but the fact that the variance of the disturbance does not exist for all but the normal-distribution member of the family could be seen as a disadvantage.

A separate literature examines the tails of the OLS estimator under heavy-tailed disturbances; e.g., see He et al. (1990), Jureckova et al. (2001), Mikosch and de Vries (2013). The GTL distribution can have tails of a kind that conforms to the type of distributions analyzed in this literature. One of the conclusions is that the OLS estimator behaves poorly if the disturbance is non-Gaussian. Moreover, the OLS estimator is adversely affected if the explanatory variables are heavy-tailed, possibly even when the disturbances are normally distributed.

This study specifies the GTL distribution as the source of variation in the disturbances. This distribution is discussed in Section 2 and is entered into the linear regression model in Section 3, where we prove the consistency and asymptotic normality of the ML estimator. We also provide an LM test for non-normality that can be applied to OLS residuals and points towards the GTL regression model if non-normality is discovered. Section 4 presents results of a Monte Carlo study that compares the performance of the OLS and ML estimators when data are generated with GTL disturbances and possibly also with heavy-tailed GTL-distributed explanatory variables. Section 5 presents five examples where OLS and GTL (ML) estimates are contrasted in common research questions that span the field of economics, from the standard topics of log-wage equations, hedonic housing prices, and capital asset pricing models to more specialized questions about the degree of police officer discretion in setting fines for speeding and the effect of preferential trade agreements on bilateral trade. In all of these cases, normality of the disturbances is resoundingly rejected. In three examples, slope estimates prove quite robust and the standard error of the GTL estimator are only about 5% smaller—but in a fourth example we find an efficiency gain of nearly 10% and essential differences in many of the main policy effects, and in a fifth example the OLS estimates are completely misleading and the GTL standard errors are less than one tenth of

4

the OLS standard errors. Section 6 concludes.

## 2 The GTL Distribution

Freimer et al. (1988) developed a generalization of the Tukey lambda distribution that permits not only varying levels of tail thickness (as the Tukey lambda distribution does) but also skewness and therefore is highly flexible. We refer to this distribution as the Generalized Tukey Lambda (GTL) distribution; in the literature, it is also known as the GLD-FMKL distribution.[3] In its canonical form, the $\text{GTL}(\alpha, \delta)$ distribution is described by its link function $G(u)$ with $u \in [0, 1]$:[4]

$$\epsilon = G(u) = \frac{u^{\alpha - \delta} - 1}{\alpha - \delta} - \frac{(1 - u)^{\alpha + \delta} - 1}{\alpha + \delta}. \tag{1}$$

$\alpha$ and $\delta$ are real-valued parameters potentially anywhere between $-\infty$ and $+\infty$. For $\alpha - \delta \to 0$, the first term converges to $\ln u$; for $\alpha + \delta \to 0$, the second term converges to $\ln(1 - u)$. Because $\alpha$ and $\delta$ often appear in pairs, we define $\lambda_1 = \alpha - \delta$ and $\lambda_2 = \alpha + \delta$

A link function is the inverse of a cumulative distribution function (cdf), but for general values of $\lambda_1$ and $\lambda_2$ the cdf does not have an analytical closed-form solution. The probability density function of $\epsilon$ is given by

$$f(\epsilon) = \frac{1}{u^{\lambda_1 - 1} + (1 - u)^{\lambda_2 - 1}} \equiv \frac{1}{G'(u)} \qquad \text{with} \qquad u = G^{-1}(\epsilon). \tag{2}$$

---

[3]Earlier, a generalization of the Tukey lambda distribution first appeared in the statistics literature in a paper by Ramberg and Schmeiser (1974) and has become known as the Generalized Lambda distribution (GLD), sometimes also referred to as GLD-RS. The parameter space of the GLD has gaps: as demonstrated in Karian et al. (1996), the GLD is not defined in the following regions of the shape parameters $\lambda_3$ and $\lambda_4$: $(i)$ $\lambda_3 \leq 0$ and $0 \leq \lambda_4 \leq 1$; $(ii)$ $-1 \leq \lambda_3 \leq 0$ and $\lambda_4 > 1$ and $(1 - \lambda_3)^{1 - \lambda_3}(\lambda_4 - 1)^{\lambda_4 - 1}(\lambda_4 - \lambda_3)^{\lambda_3 - \lambda_4}\lambda_4 \geq -\lambda_3$; and $(iii)$ symmetric regions relative to $(i)$ and $(ii)$ obtained by interchanging $\lambda_3$ and $\lambda_4$. As a result, the feasible parameter space for $(\lambda_3, \lambda_4)$ consists of four non-contiguous areas. The parameter space of the GTL distribution is free of such gaps and therefore more amenable to maximum likelihood estimation.

[4]Referring to (1) as "the canonical form" does not imply that the distribution of $\epsilon$ is standardized with mean 0 and variance 1. A few other well-known distributions that are commonly stated in their canonical form are the Student's $t$, logistic, and $\chi^2$ distributions. Many descriptions of the GTL distribution divide the expression in equation (1) by a scaling parameter and add a location parameter. This is not practical for our purposes as we fit GTL into a linear regression model that offers its own description of location.

The range of $\epsilon$ is not necessarily infinite: the lower bound equals $-\infty$ if $\lambda_1 \leq 0$ or $-1/\lambda_1$ if $\lambda_1 > 0$, and the upper bound equals $\infty$ if $\lambda_2 \leq 0$ or $1/\lambda_2$ if $\lambda_2 > 0$. The $k^{\text{th}}$ moment of $\epsilon$ exists only if $\min(\lambda_1, \lambda_2) > -1/k$.[5] In other words, the mean of $\epsilon$ does not exist if $\lambda_1 \leq -1$ or $\lambda_2 \leq -1$; if it does exist, $E[\epsilon] = -2\delta/((\lambda_1 - 1)(\lambda_2 - 1))$. Similarly, $\text{Var}(\epsilon)$ exists only if $\lambda_1 > -1/2$ and $\lambda_2 > -1/2$.[6] We shall denote $E[\epsilon]$ and $\text{Var}(\epsilon)$ of a canonical $\text{GTL}(\alpha, \delta)$-distributed $\epsilon$ as $\mu_\epsilon$ and $\sigma_\epsilon^2$.

The GTL distribution closely approximates many well-known distributions (Freimer et al., 1988; Vijverberg and Vijverberg, 2012). $\text{GLT}(0.1436, 0)$ is nearly indistinguishable from the normal distribution: the main difference is that under GTL $\epsilon$ ranges from $-6.96$ to $6.96$ and not from $-\infty$ to $\infty$. $\text{GTL}(\alpha, 0)$ for $-0.8416 \leq \alpha \leq 0.1436$ closely approximates the $t(\nu)$-distribution with $1 \leq \nu \leq \infty$ degrees of freedom. $\text{GTL}(0.1422, -0.2290)$ approximates the Gumbel distribution. With $\text{GTL}(0, 0)$, the GTL family nests the logistic distribution, and the GTL distribution simplifies to the uniform distribution for four different combinations of $(\alpha, \delta)$: $(1, 0)$, $(2, 0)$, $(\alpha, \alpha - 1)$ with $\alpha \to \infty$, and $(\alpha, 1 - \alpha)$ with $\alpha \to \infty$. Many other distributions can be approximated by matching moments or percentiles. The GTL distribution is quite flexible, indeed.

Figure 1 shows six examples of the GTL density. The distribution is symmetric when $\delta = 0$ or $\lambda_1 = \lambda_2$, right-skewed if $\delta < 0$ or $\lambda_1 > \lambda_2$, and left-skewed if $\delta > 0$ or $\lambda_1 < \lambda_2$.[7] Tails are longer and thicker if $\alpha$ is more negative.

---

[5]Equivalently, the $k^{\text{th}}$ moment exists only if $-\alpha - \frac{1}{k} < \delta < \alpha + \frac{1}{k}$.

[6] $\text{Var}(\epsilon) = E[\epsilon^2] - E[\epsilon]^2$, where $E[\epsilon^2] = \frac{2}{(\lambda_1+1)(2\lambda_1+1)} - \frac{1}{\lambda_1\lambda_2}(B(\lambda_1 + 1, \lambda_2 + 1) - \frac{1}{\lambda_1+1} - \frac{1}{\lambda_2+1} + 1) + \frac{2}{(\lambda_2+1)(2\lambda_2+1)}$ and $B(\cdot, \cdot)$ is the Beta function.

[7]As Freimer et al. (1988) show, the direction of the skew actually reverses for large values of $\alpha$ but the size of the skew is then small.

Figure 1: Standardized GTL densities for selected values of $\alpha$ and $\delta$
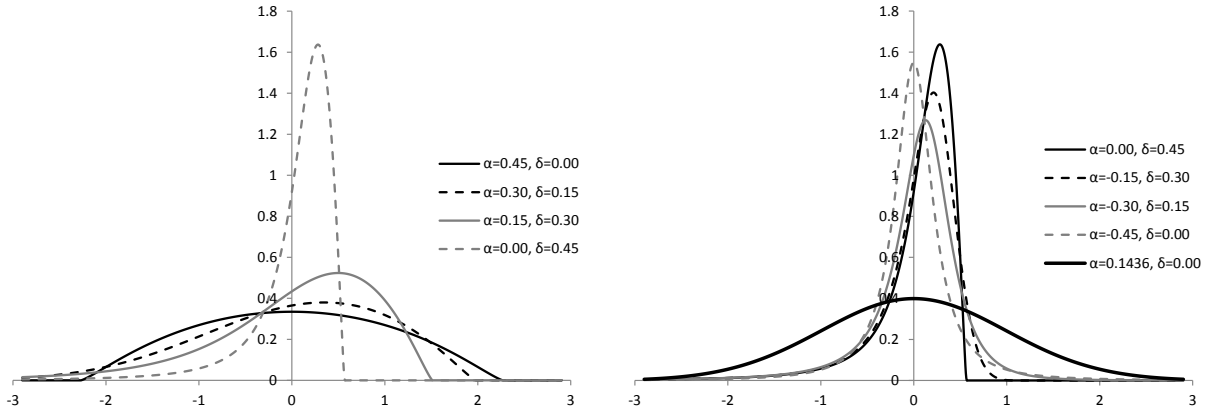


(a) Thin-tailed left-skewed densities



(b) Thick-tailed left-skewed densities

## 3 The GTL Regression Model

### 3.1 Specification

The GTL regression model is specified as follows:

$$y_i = x_i'\beta + \sigma\epsilon_i \qquad \text{where} \qquad \epsilon_i \sim \text{GTL}(\alpha, \delta). \tag{3}$$

where $i = 1, \ldots, n$ denotes observations (individuals, states, time periods, etc.). $x_i$ and $\beta$ are $k \times 1$ vectors. $\epsilon$ has a canonical $\text{GTL}(\alpha, \delta)$ distribution and is assumed independent of $x$, such that $E[\epsilon|x] = E[\epsilon]$ if indeed $\epsilon$ has a first moment. The existence of this first moment cannot be taken for granted if $\epsilon$ is generated by a GTL distribution with unknown parameters. In particular, if $\min(\lambda_1, \lambda_2) \leq -1$, $E[\epsilon]$ does not exist. For this reason also, we do not impose the restriction that $E[\epsilon] = 0$. $\sigma$ is merely a scaling parameter. The variance of the disturbance $(\sigma\epsilon)$ equals $\sigma^2\sigma_\epsilon^2$ whenever $\text{Var}(\epsilon)$ exists. However, $\text{Var}(\epsilon)$ does not exist

7

if $\min(\lambda_1, \lambda_2) \leq -1/2$.

If $E[\epsilon] \equiv \mu_\epsilon$ exists, it is a function of $(\alpha, \delta)$ and thus not generally equal to 0. Thus, in that case, $E[y|x] = x'\beta + \sigma\mu_\epsilon$. If we denote the intercept of the model with $\beta_1$, it follows that the magnitude of $\beta_1$ is sensitive to the mean of $\epsilon$. For the sake of comparability with other estimators such as OLS that assume $E[\epsilon|x] = 0$, we may compute an adjusted estimate of the intercept as $\hat{\beta}_1^* = \hat{\beta}_1 + \hat{\sigma}\hat{\mu}_\epsilon$. Its standard error is straightforwardly derived with the delta method.

Given the linear regression equation (3), the ML estimator maximizes the following likelihood function:

$$L(y, x, \theta) = -n\ln\sigma - \sum_{i=1}^{n}\ln G'(u_i) = -n\ln\sigma - \sum_{i=1}^{n}\ln\left(u_i^{\lambda_1-1} + (1-u_i)^{\lambda_2-1}\right) \qquad (4)$$

where $u_i = G^{-1}(\frac{1}{\sigma}(y_i - x_i'\beta))$.

## 3.2 OLS estimation of a GTL regression model

Equation (3) can still be estimated with Ordinary Least Squares (OLS), provided that $\epsilon$ has first and second moments. In that case (and subject to other assumptions),[8] the OLS estimator is most efficient in the class of linear unbiased estimators, regardless of the distribution of $\epsilon$. If $\epsilon$ is normally distributed, the OLS estimator is identical to the maximum likelihood estimator (MLE), but if $\epsilon$ has a GTL distribution, the MLE generally differs.[9] As proven below, the MLE of $\beta$ under the GTL distribution is consistent and asymptotically normal, even when $\epsilon$ does not possess first and second moments. The validity of OLS depends on the existence of these moments—and when these do exist, OLS is no longer efficient relative to MLE if the disturbance's distribution departs from normality.

---

[8] See White (1984) for a technical discussion.

[9] Violation of the normality assumption does not necessarily lead to inconsistency of OLS estimator under the theory of quasi (pseudo)-maximum likelihood (White, 1982; Gourieroux et al., 1984).

Another way to illustrate the difference between OLS and MLE under GTL disturbances follows from the first order conditions that are implied by maximization of the respective criterion functions. The first-order condition of OLS or MLE under a normal distribution yields:

$$\sum_{i=1}^{n} \tilde{\epsilon}_i x_i = 0 \tag{5}$$

where $\tilde{\epsilon}_i = (y_i - x_i'\beta)/(\sigma\sigma_\epsilon)$ is the standardized disturbance. The first-order condition of MLE under a GTL distribution yields:

$$\sum_{i=1}^{n} \frac{1}{\sigma} \frac{G''(u_i)}{(G'(u_i))^2} \, x_i = 0 \equiv \sum_{i=1}^{n} w(\tilde{\epsilon}_i)\tilde{\epsilon}_i x_i \equiv \sum_{i=1}^{n} \psi(\tilde{\epsilon}_i) x_i \tag{6}$$

where $u_i = G^{-1}(\sigma_\epsilon \tilde{\epsilon}_i + \mu_\epsilon)$. The term $w(\tilde{\epsilon}_i) = \frac{1}{\sigma} \frac{G''(u_i)}{(G'(u_i))^2 \tilde{\epsilon}_i}$ is the weight given to $x_i \epsilon_i$ for observation $i$. This weight distinguishes OLS from MLE: under OLS, it equals 1 for all $i$, whereas under MLE it varies with $\tilde{\epsilon}$. By varying the weight, the MLE is able to avoid the disruptive effect of data outliers.

Alternatively, we may define $\psi(\tilde{\epsilon}_i) = w(\tilde{\epsilon}_i)\tilde{\epsilon}_i$, such that $\psi(\tilde{\epsilon}_i) = \tilde{\epsilon}_i$ under OLS. Figure 2 depicts $\psi(\tilde{\epsilon})$ for various $(\alpha, \delta)$ combination since it is easier to depict in a graph than $w(\tilde{\epsilon})$. Figure 2a compares $\psi$ under OLS, or equivalently MLE under standard normality, with GTL(0.1436,0). Over the range $[-3, 3]$, which contains 99.5% of the distribution, the two lines are nearly identical; outliers beyond this range have more influence on the MLE estimator than OLS because for $(\alpha, \delta) = (0.1436, 0)$ the feasible range of $\tilde{\epsilon}$ is finite. In contrast, when tails become thicker (Figure 2b), observations with large $\tilde{\epsilon}$ are increasingly discounted.[10] Thus, a large $\tilde{\epsilon}$ is informative in a distribution with thin tails and a finite

---

[10] See Nolan and Ojeda-Revah (2013, p.193) for a similar pattern in the context of a stable distribution that permits thick-tailed and skewed disturbances. The distribution is characterized in part by an index of stability denoted as $\alpha$ (which has a different meaning than the parameter $\alpha$ in the GTL distribution), with $0 < \alpha \leq 2$. The comparision of the GTL and stable distributions is left as a future research topic, but suffice it to note that stable distributions have no moments of order $\alpha$ or higher—and thus no variance unless $\alpha = 2$, which is the special case that corresponds with the normal distribution (Samorodnitsky et al., 2007).

Figure 2: $\psi(\tilde{\epsilon})$ as a function of $\tilde{\epsilon}$ for selected values of $\alpha$ and $\delta$
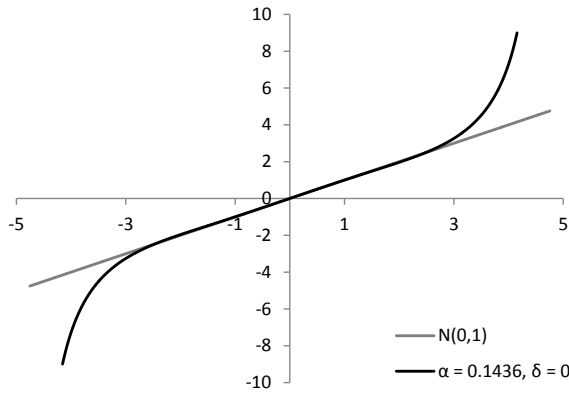
(a) Comparing $N(0,1)$ with GTL$(0.1436, 0)$
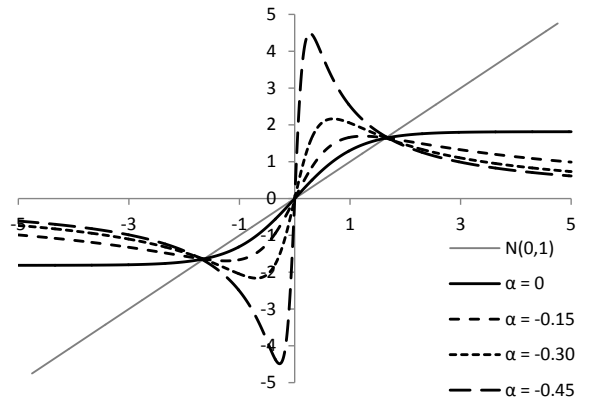
(b) Increasingly thick tails: $\delta = 0$
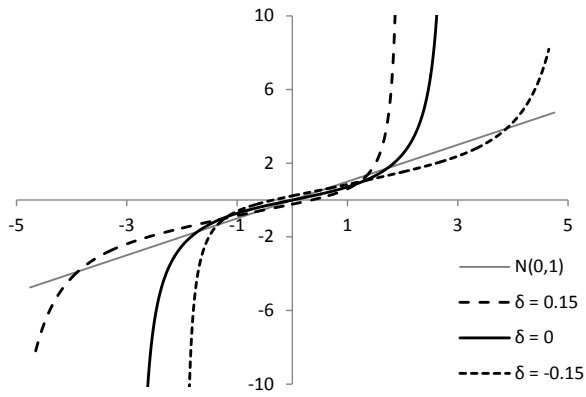
(c) Thin tails with skewness: $\alpha = 0.30$
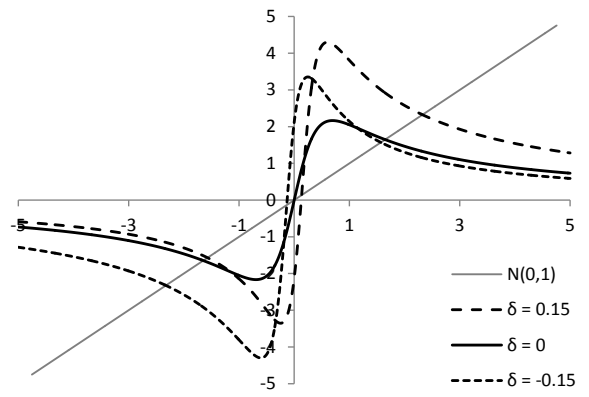
(d) Thick tails with skewness: $\alpha = -0.30$

10

range because it more clearly defines the set of plausible values of $\hat{\beta}$, whereas a large $\tilde{\epsilon}$ in the context of a distribution with thick tails is rather uninformative. Figures 2c and 2d shows that skewness in the distribution generates asymmetry in the $\psi$-function around 0. When $\delta$ is negative and the GTL distribution is right-skewed, the right tail is longer than the left and thus observations with large positive disturbances are weighted less heavily than observations with large negative disturbances.[11]

Non-normality may well be common in real-life data. For example, financial data are often reported to have thick tails and sometimes exhibit skewness as well. A recent study by Mikosch and de Vries (2013) shows that statistical inference based on OLS can be misleading if the disturbance has thick tails. GTL regression can gain efficiency by properly accounting for tail thickness and by properly favouring the information in one tail relative to the other as a result of asymmetry in the distribution. We examine this further in a Monte Carlo study in Section 4.

### 3.3 Asymptotic Properties of MLE of the GTL regression model

Let $\theta = (\beta', \ \sigma, \ \alpha \ , \delta)'$ be the complete parameter vector. Let $\theta_0$ denote the true parameter vector. Let $\Theta$ be the parameter space.

**Assumption A.1** $\Theta$ *is compact with* $\sigma > 0$.

**Assumption A.2** $(\alpha, \delta)$ *are such that* $\lambda_1 = \alpha - \delta < \frac{1}{2}$ *and* $\lambda_2 = \alpha + \delta < \frac{1}{2}$.

**Assumption A.3** $\theta_0 \in int(\Theta)$.

**Assumption A.4** *(i)* $x \in \mathcal{X} \subset \mathbb{R}^k$. *(ii)* $x$ *is weakly exogenous.* *(iii)* $x$ *has finite moments up to the fourth order.* *(iv)* $E[xx']$ *has full rank.*

---

[11]Graphs of $w(\tilde{\epsilon}_i)$ may be deduced from Figure 2: $w$ equals the ratio of GTL's $\psi$ to $N(0,1)$'s $\psi$. Thus, for configurations of Figure 2b, the typical $w$-curve exceeds 1 in the middle around the origin and trails off to 0 near the ends of the range. However, for skewed configurations, $\psi$ is negative for some values where $\tilde{\epsilon}$ is positive or vice versa, giving rise to locally negative values of $w$ near, and vertical asymptotes at, the origin. Because of this, graphs of $\psi$ provide better intuition.

**Assumption A.5** *(i) $\epsilon$ is independently and identically $GTL(\alpha, \delta)$-distributed, with the GTL distribution defined by the link function in equation (1) and the probability density function in equation (2). (ii) $\epsilon$ is independent of $x$.*

Define the range of $\epsilon$ as $\mathcal{E}(\theta)$. As indicated in Section 2, the range of $\epsilon$ may depend on $(\alpha, \delta)$. By equation (3), we may define the range of $y$ as $\mathcal{Y}(x, \theta)$, where $\mathcal{Y}(x, \theta)$ is $\mathcal{E}(\theta)$ scaled by $\sigma$ and shifted by $x'\beta$.

The first question is whether the MLE estimator $\hat{\theta}$ is consistent.

**Theorem 1** *Given Assumptions A.1, A.3, A.4, and A.5, $\hat{\theta}$ is a consistent estimator of $\theta_0$ for all $\theta_0 \in \Theta$ except for those $\theta_0$ for which $(\alpha_0, \delta_0) = (1, 0)$ or $(\alpha_0, \delta_0) = (2, 0)$.*

**Proof** See Appendix A. ∎

The exceptions stem from the fact that the GTL density simplifies to a uniform density for both of the stated $(\alpha, \delta)$ values. However, the exceptions are less problematic than they seem, as will become clear from Theorem 2 below.

As for the asymptotic distribution of $\hat{\theta}$, we verify the conditions of Theorem 13.2 of Wooldridge (2002, p.395) in relation to, in particular, the values of the parameters $\alpha$ and $\delta$.

**Lemma 1** *For each $(x, y) \in \mathcal{X} \times \mathcal{Y}(x, \theta)$, $L(y, x, \cdot)$ is twice continuously differentiable on int($\Theta$) for all $\theta \in \Theta$ except where $\lambda_1, \lambda_2 = 1, 2, 3$.*

**Proof** Derivatives are provided and examined in Appendix B. ∎

The dependence of $\mathcal{Y}(x, \theta)$ on $\theta$ is usually problematic, because proofs of asymptotic normality rely on the interchange of differentiation and integration, and the dependence of the bounds of $y$ on $\theta$ may prevent that. The following lemma addresses this issue:

**Lemma 2** *Define $g_y(y|x_i; \theta)$ as the conditional density of $y$. Define $\ell_i(\theta) = \ln g_y(y|x_i; \theta)$. Define $s_i(\theta) = \nabla_\theta \ell_i(\theta)$. Then, for all $x_i \in \mathcal{X}$ and $\theta \in int(\Theta)$ with $\lambda_1 < \frac{1}{2}$ and $\lambda_2 < \frac{1}{2}$, we have:*

*(i)* $\nabla_\theta \left( \int\limits_{\mathcal{Y}(x,\theta)} g_y(y|x_i; \theta) dy \right) = \int\limits_{\mathcal{Y}(x,\theta)} \nabla_\theta g_y(y|x_i; \theta) dy$ *, and*

*(ii)* $\nabla_\theta \left( \int\limits_{\mathcal{Y}(x,\theta)} s_i(\theta) g_y(y|x_i; \theta) dy \right) = \int\limits_{\mathcal{Y}(x,\theta)} \nabla_\theta s_i(\theta) g_y(y|x_i; \theta) dy$ *.*

**Proof** See Appendix C. ∎

**Lemma 3** *The elements of $\nabla_\theta^2 L(y, x, \theta)$ are bounded in absolute value by a function $b(y, x)$ with finite expectation for all $\theta \in \Theta$ where (i) $\lambda_1 < \frac{1}{2}$ and $\lambda_2 < \frac{1}{2}$, or (ii) $\lambda_1 > 2$ and $\lambda_2 > 2$ with $\lambda_1, \lambda_2 \neq 3$.*

**Proof** See Appendix B. ∎

**Lemma 4** *Define $A_0 = -E[H(\theta_0)]$ where $H(\theta_0) = \nabla_\theta^2 \ell(\theta_0)$. Then $A_0$ is positive definite.*

**Proof** $\nabla_\theta^2 \ln f(y|x_i; \theta_0)$ is a complicated expression, of which the expected value involves integrals that do not have an analytical solution. A direct proof that $A_0$ is positive definite is therefore not feasible, but Appendix D evaluates the needed integrals by Monte Carlo simulation over a grid of representative values of $(\alpha, \delta)$ and does not encounter any case where $A_0$ fails to be positive definite. ∎

In the light of these results, the following theorem defines the asymptotic distribution for the estimator of the GTL regression model; Assumption A.2 is added because of Lemma 2:

**Theorem 2** *Given the linear regression model of equation (3) and Assumptions A.1-A.5, the ML estimator $\hat{\theta}$ is asymptotically normally distributed $N(\theta_0, V_0)$, where $V_0 = n^{-1} A_0$ is estimated as*

$$V(\hat{\theta}) = -\left( \sum_{i=1}^n \nabla_{\theta\theta} \ell_i(\hat{\theta}) \right)^{-1} \tag{7}$$

13

### 3.4 An LM test for normality

As mentioned, the normal distribution may be approximated with the GTL$(0.1436, 0)$ distribution. Thus, a Lagrange Multiplier test for normality proceeds as follows. Let the restricted estimator of $\theta$ be denoted as $\breve{\theta} = (\breve{\beta}', \breve{\sigma}, 0.1436, 0)'$; let $s(\breve{\theta}) = \sum_{i=1}^{n} s_i(\breve{\theta})$ (see Lemma 2); and let $V(\breve{\theta})$ be the negative of the inverted hessian evaluated at $\breve{\theta}$ (see equation (7)).[12] Then approximate normality may be tested with:

$$LM = s(\breve{\theta})' \; V(\breve{\theta})^{-1} \; s(\breve{\theta}) \;\; \overset{a}{\sim} \;\; \chi^2(2) \tag{8}$$

Since this LM test is a test for approximate normality in the form of GTL$(0.1436, 0)$, the question arises how much this model actually deviates from true normality. Simulation results show that the power of the test to be only 7.4% in a large sample ($n = 5000$) in the context of the baseline design to be discussed in the next section. The LM does have power against minor deviations from GTL$(0.1436, 0)$: if disturbances are generated with slightly thicker tails $(\alpha, \delta) = (0.10, 0)$ (which closely approximates a Student's $t(20)$ distribution), power equals 84.9%; and similarly for a slightly thinner tail $(\alpha, \delta) = (0.20, 0)$ or slight skewness $(0.1436, 0.02)$, power is 99.5% and 77.0% respectively. Thus, evaluated by the LM test, pure normality and GTL$(0.1436, 0)$ are virtually indistinguishable.

Of course, LM is not the only feasible test statistic. Alternative tests are a Wald test that uses $(\hat{\alpha}, \hat{\delta})$ and $V(\hat{\theta})$; a likelihood ratio test; a Vuong test that examines the observations' contributions to the likelihood function under GTL$(\hat{\alpha}, \hat{\delta})$ and under normality; and a Jarque-

---

[12]Rather than actually estimating a GTL$(0.1436, 0)$ regression model, a simpler test procedure would be to estimate the linear regression model with OLS, deduce the proper estimate of $\sigma$ from the root mean squared error of the residuals, evaluate $s(\breve{\theta})$ and $V(\breve{\theta})$, and compute $LM$. In our experience, it does happen occasionally that the largest OLS residuals fall outside the feasible range $\mathcal{E}(\theta)$ of $\epsilon$, which for $(\alpha, \delta) = (0.1436, 0)$ equals $(-6.694, 6.694)$ if $\epsilon$ is GTL in canonical form or $(-4.819, 4.819)$ if $\epsilon$ is a standardized GTL disturbance (i.e., $\tilde{\epsilon}$). Since $s$ and $V$ are derived in terms of $u_i = G^{-1}(\epsilon_i) \in [0, 1]$ (see Appendix B), we might set $u_i = 0$ if $\tilde{\epsilon}_i < -4.819$ and $u_i = 1$ if $\tilde{\epsilon}_i > 4.819$ and proceed with the computation, which then typically yields a large $LM$ value. However, if $\epsilon$ is truly normally distributed, the chance of this happening is less than $1.5 \times 10^{-6}$, so such range violations indicate that $\epsilon$ may not be normally distributed anyway.

Bera test of normality of OLS residuals.

## 4  Monte Carlo evidence: the effect of GTL disturbances

By the Gauss-Markov Theorem, skewed and unusually-tailed data do not bias the OLS estimator or render it inefficient relative to other linear unbiased estimators, provided that the disturbance has first and second moments. Yet, non-linear estimators may exist that are efficient relative to OLS. The GTL-regression estimator is such an estimator. In this section, we report the results of a Monte Carlo analysis where data are generated with GTL disturbances. We employ two designs. In the first, observables are modelled as the outcome of a process that is conventional in Monte Carlo analyses; in the second, observables are thick-tailed as well.

### 4.1  The first design: a baseline

The data generating process is as follows: $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i$ for $i = 1, \ldots, n$ with $n = 250$ or $5000$, where $x_{2i}$ is a standard normal random variate, $x_{3i}$ is a $\chi^2(5)$ random variate that is standardized to have mean 0 and variance 1 and is independent of $x_{2i}$, and $\epsilon_i$ is a scaled $\mathrm{GTL}(\alpha, \delta)$ variate for various combinations of $\alpha$ and $\delta$. The values of $\beta_1$, $\beta_2$ and $\beta_3$ are all equal to 1. The skewness of $x_3$ inserts a bit of skewness into $y$. Because of the range of $(\alpha, \delta)$ combinations we cannot standardize the GTL distribution. Instead, we scale $\epsilon$ such that the range of the GTL distribution from the 0.1% quantile to the 99.9% quantile has the same length as that of a normal $N(0, 2)$ distribution; this scaling parameter is denoted as $\sigma$. This keeps the $R^2$ of the OLS regression in the neighborhood of 0.50 as long as the standard deviation of $\epsilon$ is defined, but the $R^2$ diminishes as the tails of the GTL distribution become longer. We replicate each design 1000 times, drawing new explanatory variables each time and using the same sets of draws across the different experiments.

Table 1 presents the root mean squared error (RMSE) of the OLS and GTL estimators

Table 1: RMSE of OLS and GTL estimators of slopes and intercept for GTL-generated data

| DGP | | | RMSE of OLS | | | RMSE of GTL | | |
|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\delta$ | $\sigma$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_1$ |
| A: GTL as an approximation of the standard normal distribution, $N = 250$ | | | | | | | | |
| 0.1436 | 0.00 | 1.188 | 0.107 | 0.110 | 0.110 | 0.108 | 0.111 | 0.126 |
| B: Various GTL distributions, small sample, $N = 250$ | | | | | | | | |
| 0.33 | -0.10 | 1.477 | 0.106 | 0.108 | 0.199 | 0.089 | 0.094 | 0.132 |
| 0.33 | 0.00 | 1.508 | 0.107 | 0.109 | 0.109 | 0.097 | 0.101 | 0.133 |
| 0.33 | 0.10 | 1.477 | 0.106 | 0.108 | 0.200 | 0.091 | 0.093 | 0.132 |
| -0.33 | -0.10 | 0.454 | 0.133 | 0.144 | 0.255 | 0.066 | 0.068 | 0.066 |
| -0.33 | 0.00 | 0.482 | 0.118 | 0.122 | 0.124 | 0.071 | 0.073 | 0.071 |
| -0.33 | 0.10 | 0.454 | 0.127 | 0.126 | 0.239 | 0.067 | 0.069 | 0.067 |
| -0.67 | -0.25 | 0.136 | 1.737 | 1.667 | 1.932 | 0.025 | 0.026 | 0.025 |
| -0.67 | 0.00 | 0.202 | 0.336 | 0.337 | 0.341 | 0.038 | 0.040 | 0.037 |
| -0.67 | 0.25 | 0.136 | 1.046 | 0.823 | 1.126 | 0.025 | 0.026 | 0.025 |
| -1.00 | -0.50 | 0.024 | 112.585 | 89.227 | 103.243 | 0.005 | 0.005 | 0.005 |
| -1.00 | 0.00 | 0.079 | 2.503 | 2.217 | 2.408 | 0.018 | 0.019 | 0.018 |
| -1.00 | 0.50 | 0.024 | 39.742 | 27.930 | 35.099 | 0.005 | 0.005 | 0.005 |
| C: Various GTL distributions, large sample, $N = 5000$ | | | | | | | | |
| 0.33 | -0.10 | 1.477 | 0.023 | 0.024 | 0.169 | 0.018 | 0.019 | 0.032 |
| 0.33 | 0.00 | 1.508 | 0.023 | 0.024 | 0.025 | 0.021 | 0.021 | 0.030 |
| 0.33 | 0.10 | 1.477 | 0.023 | 0.024 | 0.171 | 0.019 | 0.019 | 0.033 |
| -0.33 | -0.10 | 0.454 | 0.034 | 0.032 | 0.209 | 0.015 | 0.014 | 0.015 |
| -0.33 | 0.00 | 0.482 | 0.027 | 0.028 | 0.029 | 0.016 | 0.015 | 0.016 |
| -0.33 | 0.10 | 0.454 | 0.031 | 0.032 | 0.210 | 0.014 | 0.014 | 0.015 |
| -0.67 | -0.25 | 0.136 | 6.075 | 3.868 | 4.933 | 0.005 | 0.005 | 0.006 |
| -0.67 | 0.00 | 0.202 | 0.233 | 0.180 | 0.220 | 0.008 | 0.008 | 0.008 |
| -0.67 | 0.25 | 0.136 | 1.903 | 1.456 | 2.512 | 0.005 | 0.005 | 0.006 |
| -1.00 | -0.50 | 0.024 | 9916.534 | 5631.352 | 7274.680 | 0.001 | 0.001 | 0.001 |
| -1.00 | 0.00 | 0.079 | 12.606 | 7.908 | 10.359 | 0.004 | 0.004 | 0.004 |
| -1.00 | 0.50 | 0.024 | 1997.902 | 832.310 | 2381.450 | 0.001 | 0.001 | 0.001 |

of the slopes and intercept for different values of $(\alpha, \delta)$. In Panel A, a GTL(0.1436,0) distribution closely approximates a standard normal distribution. Not surprisingly, the OLS and GTL estimators perform nearly identically even when the sample at $n = 250$ is rather small. In Panel B with $n = 250$ and Panel C with $n = 5000$, the GTL distribution varies from having truncated tails (for $\alpha = 0.33$) to having very long tails (for $\alpha = -1$), with right skew if $\delta$ is negative and left skew if $\delta$ is positive. The direction of the skewness matters somewhat for the RMSEs because of the skewness inherent in one of the explanatory variables ($x_3$), but the overriding concern is the length of the tails. The more negative $\alpha$ is, the worse OLS

performs. The GTL estimator is hardly affected by the thickness or skewness of the tails; in fact, the RMSEs diminish as the GTL distribution becomes more heavy-tailed. For a larger negative $\alpha$, the GTL estimator outperforms the OLS estimator by a vast margin.

Detailed inspection of simulation results shows that the RMSEs of the slope estimators represent mostly the standard deviation of the simulated slope estimates. Given the exogeneity of the explanatory variables, the OLS estimator is unbiased as long as the mean of $\epsilon$ exists, but even when $(\alpha, \delta)$ is such that the first moment of $\epsilon$ no longer exists and the OLS estimator loses its theoretical moments also, the average of the simulated values is still close to 1. For the GTL estimator, the intercept estimator shows a bias as expected (Section 3.1), but it is minor compared to the standard deviation of the simulated values.

Table 2 examines the biases and standard deviations of the GTL estimator of the parameters of the GTL distribution that generates the disturbance, namely the shape parameters $\alpha$ and $\delta$ as well as the scaling parameter $\sigma$ that is determined by linking the length of the quantile range of the normal and GTL distributions as described above. $\sigma$ varies with $\alpha$ and $\delta$ as indicated in the table. Table 2 indicates a small bias for $\alpha = 0.33$, regardless of $\delta$; the reason is that the maximum likelihood function is augmented with a small regularity penalty function[13] to keep $(\alpha, \delta)$ within the feasible parameter area where the GTL estimator is consistent and asymptotically normal (Theorem 2). The penalty function is no longer needed when the values of $\alpha$ and $\delta$ of the data generating process are solidly inside this feasible area. The Monte Carlo results indicate that even for relatively small samples, bias in the GTL estimator of $\sigma$, $\alpha$ and $\delta$ is no issue.

Separate from concerns about bias and precision is the question whether the small-sample distribution of the estimators is approximately normal. One way to examine the small-sample properties is to compute coverage ratios, i.e., the rate at which the confidence intervals contain the population parameter values. Table 3 illustrates our findings with three experiments;

---

[13]See Appendix E for details.

17

Table 2: Bias and precision of GTL parameters for GTL-generated data

| DGP | | | Bias | | | Standard deviation | | |
|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\delta$ | $\sigma$ | $\sigma$ | $\alpha$ | $\delta$ | $\sigma$ | $\alpha$ | $\delta$ |
| A: GTL as an approximation of the standard normal distribution ($N = 250$) | | | | | | | | |
| 0.1436 | 0.00 | 1.188 | 0.028 | 0.022 | -0.001 | 0.128 | 0.064 | 0.033 |
| B: Various GTL distributions, small sample ($N = 250$) | | | | | | | | |
| 0.33 | -0.10 | 1.477 | -0.090 | -0.032 | 0.005 | 0.101 | 0.041 | 0.028 |
| 0.33 | 0.00 | 1.508 | -0.060 | -0.017 | -0.001 | 0.112 | 0.046 | 0.028 |
| 0.33 | 0.10 | 1.477 | -0.090 | -0.032 | -0.006 | 0.097 | 0.040 | 0.029 |
| -0.33 | -0.10 | 0.454 | 0.001 | 0.006 | -0.001 | 0.056 | 0.081 | 0.052 |
| -0.33 | 0.00 | 0.482 | 0.001 | 0.006 | -0.001 | 0.059 | 0.081 | 0.052 |
| -0.33 | 0.10 | 0.454 | 0.001 | 0.006 | 0.000 | 0.056 | 0.081 | 0.052 |
| -0.67 | -0.25 | 0.136 | 0.000 | 0.003 | -0.002 | 0.019 | 0.100 | 0.069 |
| -0.67 | 0.00 | 0.202 | -0.001 | 0.002 | -0.001 | 0.028 | 0.099 | 0.068 |
| -0.67 | 0.25 | 0.136 | 0.000 | 0.003 | 0.000 | 0.019 | 0.100 | 0.069 |
| -1.00 | -0.50 | 0.024 | 0.000 | 0.002 | -0.002 | 0.004 | 0.120 | 0.088 |
| -1.00 | 0.00 | 0.079 | 0.000 | -0.001 | -0.001 | 0.012 | 0.118 | 0.085 |
| -1.00 | 0.50 | 0.024 | 0.000 | 0.001 | 0.000 | 0.004 | 0.119 | 0.086 |
| C: Various GTL distributions, large sample ($N = 5000$) | | | | | | | | |
| 0.33 | -0.10 | 1.477 | -0.085 | -0.037 | 0.004 | 0.022 | 0.009 | 0.006 |
| 0.33 | 0.00 | 1.508 | -0.066 | -0.028 | 0.000 | 0.025 | 0.009 | 0.005 |
| 0.33 | 0.10 | 1.477 | -0.085 | -0.037 | -0.004 | 0.022 | 0.009 | 0.005 |
| -0.33 | -0.10 | 0.454 | 0.000 | 0.001 | 0.000 | 0.012 | 0.018 | 0.011 |
| -0.33 | 0.00 | 0.482 | 0.000 | 0.001 | 0.000 | 0.013 | 0.018 | 0.011 |
| -0.33 | 0.10 | 0.454 | 0.000 | 0.001 | 0.000 | 0.012 | 0.018 | 0.011 |
| -0.67 | -0.25 | 0.136 | 0.000 | 0.001 | 0.001 | 0.004 | 0.023 | 0.015 |
| -0.67 | 0.00 | 0.202 | 0.000 | 0.000 | 0.001 | 0.006 | 0.023 | 0.015 |
| -0.67 | 0.25 | 0.136 | 0.000 | 0.000 | 0.000 | 0.004 | 0.023 | 0.015 |
| -1.00 | -0.50 | 0.024 | 0.000 | 0.001 | 0.001 | 0.001 | 0.027 | 0.019 |
| -1.00 | 0.00 | 0.079 | 0.000 | 0.000 | 0.001 | 0.003 | 0.027 | 0.019 |
| -1.00 | 0.50 | 0.024 | 0.000 | 0.000 | 0.001 | 0.001 | 0.028 | 0.020 |

the detailed results for all experiments are provided in Appendix E. Surprisingly, in Panel A of Table 3, the coverage rates of the OLS slope estimators are essentially 95% in all experiments; the intercept is biased when the GTL distribution of $\epsilon$ is skewed as the GTL distribution is utilized in its canonical form. This would suggest that perhaps the sampling distribution of the OLS estimators is nearly normal after all. However, the Jarque-Bera test in Panel B squashes that conclusion: in the presence of GTL disturbances, the sampling distribution of the OLS estimator is not normal even for $n = 5000$.

How can the findings of Panels A and B be reconciled? Figure 3 presents QQ plots of the

Table 3: Diagnostics of the distribution of OLS and GTL estimators: baseline design

| DGP | | | OLS | | | GTL | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\delta$ | $\sigma$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\sigma$ | $\alpha$ | $\delta$ |
| A: Coverage rate at a 95% confidence level | | | | | | | | | | | |
| Small sample, $N = 250$ | | | | | | | | | | | |
| 0.1436 | 0.00 | 1.188 | 0.947 | 0.948 | 0.954 | 0.944 | 0.937 | 0.943 | 0.941 | 0.918 | 0.924 |
| -0.33 | -0.10 | 0.454 | 0.958 | 0.948 | 0.655 | 0.952 | 0.936 | 0.951 | 0.943 | 0.946 | 0.950 |
| -0.67 | -0.25 | 0.136 | 0.955 | 0.959 | 0.337 | 0.955 | 0.933 | 0.955 | 0.951 | 0.948 | 0.943 |
| Large sample, $N = 5000$ | | | | | | | | | | | |
| 0.1436 | 0.00 | 1.188 | 0.960 | 0.948 | 0.939 | 0.960 | 0.946 | 0.950 | 0.949 | 0.932 | 0.946 |
| -0.33 | -0.10 | 0.454 | 0.956 | 0.944 | 0.000 | 0.960 | 0.956 | 0.951 | 0.941 | 0.948 | 0.944 |
| -0.67 | -0.25 | 0.136 | 0.952 | 0.961 | 0.093 | 0.953 | 0.956 | 0.949 | 0.938 | 0.949 | 0.944 |
| B: $p$-values of Jarque Bera tests for normality of the sampling distribution of the estimator | | | | | | | | | | | |
| Small sample, $N = 250$ | | | | | | | | | | | |
| 0.1436 | 0.00 | 1.188 | 0.91 | 0.10 | 0.62 | 0.46 | 0.22 | 0.55 | 0.00 | 0.00 | 0.01 |
| -0.33 | -0.10 | 0.454 | 0.00 | 0.00 | 0.00 | 0.23 | 0.21 | 0.86 | 0.00 | 0.10 | 0.73 |
| -0.67 | -0.25 | 0.136 | 0.00 | 0.00 | 0.00 | 0.07 | 0.70 | 0.80 | 0.00 | 0.06 | 0.55 |
| Large sample, $N = 5000$ | | | | | | | | | | | |
| 0.1436 | 0.00 | 1.188 | 0.14 | 0.88 | 0.14 | 0.21 | 0.94 | 0.96 | 0.15 | 0.78 | 0.01 |
| -0.33 | -0.10 | 0.454 | 0.00 | 0.00 | 0.00 | 0.51 | 0.50 | 0.99 | 0.30 | 0.44 | 0.06 |
| -0.67 | -0.25 | 0.136 | 0.00 | 0.00 | 0.00 | 1.00 | 0.39 | 0.98 | 0.71 | 0.42 | 0.11 |
| C: Ratio of average variance to Monte Carlo variance | | | | | | | | | | | |
| Small sample, $N = 250$ | | | | | | | | | | | |
| 0.1436 | 0.00 | 1.188 | 1.045 | 1.000 | 0.979 | 1.010 | 0.961 | 0.988 | 0.905 | 0.878 | 0.856 |
| -0.33 | -0.10 | 0.454 | 1.073 | 0.926 | 0.915 | 1.018 | 0.956 | 1.031 | 0.970 | 0.992 | 0.929 |
| -0.67 | -0.25 | 0.136 | 0.875 | 1.010 | 0.866 | 1.005 | 0.945 | 1.051 | 1.007 | 1.019 | 0.933 |
| Large sample, $N = 5000$ | | | | | | | | | | | |
| 0.1436 | 0.00 | 1.188 | 1.051 | 1.005 | 0.935 | 1.048 | 1.004 | 0.978 | 0.951 | 0.935 | 0.990 |
| -0.33 | -0.10 | 0.454 | 1.050 | 1.046 | 0.901 | 1.027 | 1.036 | 0.992 | 0.912 | 0.908 | 0.951 |
| -0.67 | -0.25 | 0.136 | 1.839 | 1.897 | 0.998 | 1.006 | 1.058 | 0.996 | 0.918 | 0.932 | 0.940 |

OLS slope estimators, together with the 95% confidence intervals for each replication and a dashed reference line at 1. Thus, the coverage rate reflects how many of these confidence interval bracket the dashed line. Panels 3a and 3b illustrate the case of $(\alpha, \delta) = (0.1436, 0)$, where OLS is the best linear unbiased estimator: the QQ plot is completely supportive of the conclusion that the sampling distribution is normal, and the width of the confidence interval is stable across replications. In Panels 3c and 3d, the QQ plots gives evidence of

a sampling distribution with fat tails and a slight but systematic "sling" around the $45°$ line. The confidence interval is more irregularly sized and, unlike in Panels 3a and 3b, do not systematically fail to bracket the dashed line near the ends of the range. These patterns are all more prevalent for the experiment in Panels 3e and 3f, where the variance of the $\text{GTL}(-0.67, -0.25)$-distributed disturbance is not even defined, implying that $\text{Var}(\hat{\beta}_{OLS})$ is not defined. In all, the fact that the coverage rate happens to be 95% in all cases is purely incidental and not at all evidence that the usual statistical inference methods apply.

In contrast, the MLE results indicate a good coverage rate and a highly satisfactory QQ plot; Figure 4 shows the QQ plots for the experiment with $(\alpha, \delta) = (-0.67, -0.25)$, which gave OLS the most trouble in Figure 3. Other experiments yield similar results.

Panel C of Table 3 contributes one more facet about the small-sample behavior of the estimators. Figure 3 shows the confidence intervals becoming more variable in width as the GTL distribution differs more substantially from normality. The variance of the estimator, OLS or MLE, is itself estimated as well. Each replication yields another estimate of the variance, as well as an estimate of $\theta$: the average of the estimated variances ought to be close to the variance of the replicated $\hat{\theta}$. As the ratios of these two quantities in Panel C of Table 3 shows, the estimated OLS variance becomes problematic as GTL distribution differs more greatly from normality, and the ratio of the MLE variances stays close to 1.

## 4.2   The second design: thick-tailed observables

One might question the wisdom of a Monte Carlo design that specifies thick-tailed unobservables but regular-tailed observable determinants. Regression designs that contain thick-tailed explanatory variables may be problematic (e.g. Huber, 1981; He et al., 1990; Jureckova et al., 2001): the OLS estimator may lose its consistency property when the data generating process creates a few observations with outlying values of $x$ that have an outsized influence on the position of the regression line. However, such problems appear to be alleviated when the

Figure 3: QQ plots of OLS estimators of $\beta_2$ and $\beta_3$, $n = 250$
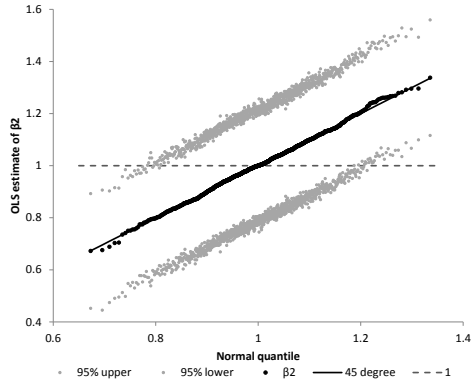
(a) $\hat{\beta}_2$ for $(\alpha, \delta) = (0.1436, 0)$
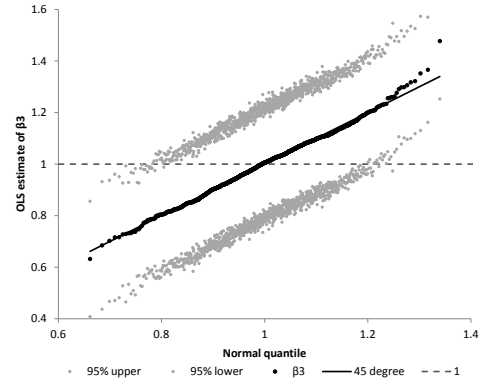
(b) $\hat{\beta}_3$ for $(\alpha, \delta) = (0.1436, 0)$
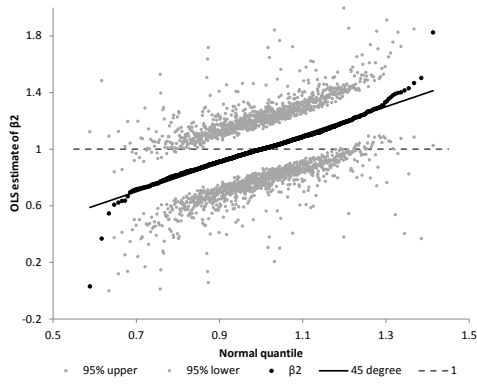
(c) $\hat{\beta}_2$ for $(\alpha, \delta) = (-0.33, -0.10)$
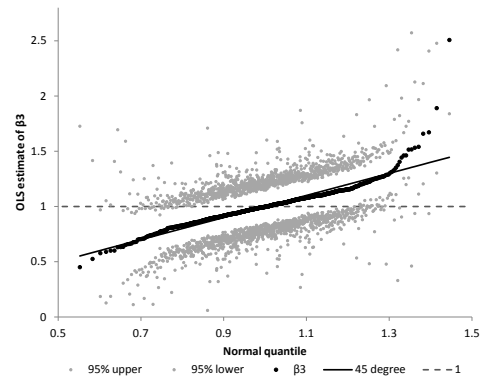
(d) $\hat{\beta}_3$ for $(\alpha, \delta) = (-0.33, -0.10)$
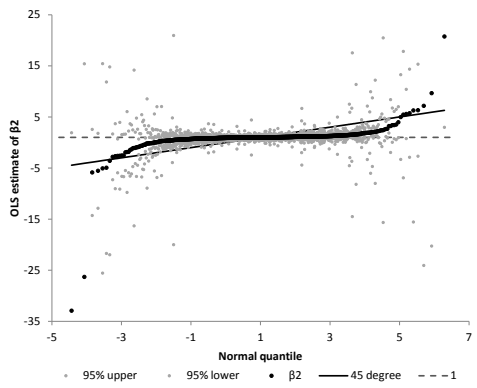
(e) $\hat{\beta}_2$ for $(\alpha, \delta) = (-0.67, -0.25)$
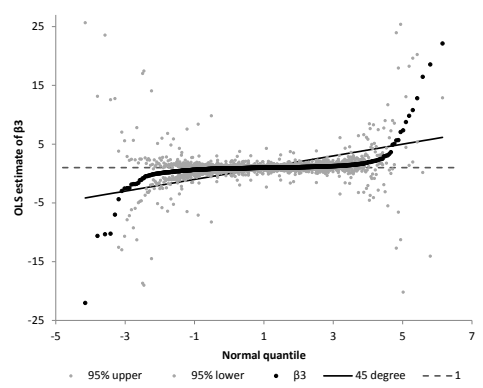
(f) $\hat{\beta}_3$ for $(\alpha, \delta) = (-0.67, -0.25)$

Figure 4: QQ plots of MLE estimators of $\beta_2$ and $\beta_3$, $n = 250$
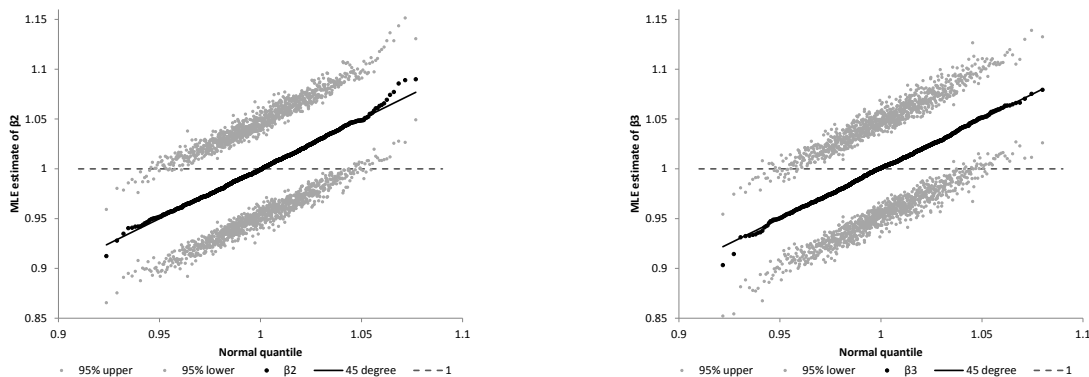


(a) $\hat{\beta}_2$ for $(\alpha, \delta) = (-0.67, -0.25)$    (b) $\hat{\beta}_3$ for $(\alpha, \delta) = (-0.67, -0.25)$

disturbance is normally distributed (Jureckova et al., 2001). Table 4 contains a few cases with thick-tailed disturbances and both regular- and thick-tailed $x$-variables, generated as GTL variates. As before, the GTL distribution of these GTL variates is scaled such that the distance between the 0.1% quantile and the 99.9% quantile is the same as that of a standard normal distribution.

The disturbances in Panel A are distributed GTL$(-0.67, 0.25)$ and thus are thick-tailed and skewed. The explanatory variables change from two standard normal ones to two thick-tailed skewed ones: the performance of the OLS estimator actually improves, as does the GTL estimator. Panel B shows that this also happens when the disturbances are standard normal (or, closely similar, GTL(0.1436,0)). These results seem at odds with the research of Huber (1981) that the OLS estimator is inconsistent with heavy-tailed explanatory variables. However, Panels A and B of Table 4 do not address the matter of consistency as sample size is unchanged between these experiments. The reason for the improved performance is that despite the scaling of the GTL distribution the variation in the explanatory variables increases dramatically as the tails thicken. The $R^2$ rises for many of these OLS regressions

22

Table 4: RMSE of OLS and GTL estimators of $\beta$ with thick-tailed determinants

| DGP for $x_2$ | | DGP for $x_3$ | | OLS | | | MLE | | |
|---|---|---|---|---|---|---|---|---|---|
| $\alpha_{x2}$ | $\delta_{x2}$ | $\alpha_{x3}$ | $\delta_{x3}$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_1$ |
| A: GTL-Disturbances are generated with $(\alpha, \delta) = (-0.67, 0.25)$ for $n = 250$ | | | | | | | | | |
| 0.1436 | 0.00 | 0.1436 | 0.00 | 1.0466 | 0.9638 | 1.1273 | 0.0250 | 0.0259 | 0.0248 |
| 0.1436 | 0.00 | -0.33 | 0.00 | 1.0464 | 0.5981 | 1.1254 | 0.0249 | 0.0152 | 0.0249 |
| 0.1436 | 0.00 | -0.33 | 0.10 | 1.0486 | 0.6754 | 1.1130 | 0.0249 | 0.0156 | 0.0250 |
| 0.1436 | 0.00 | -0.67 | 0.00 | 1.0456 | 0.5650 | 1.1226 | 0.0249 | 0.0134 | 0.0249 |
| 0.1436 | 0.00 | -0.67 | 0.25 | 1.0484 | 0.7007 | 1.1037 | 0.0249 | 0.0151 | 0.0252 |
| -0.67 | 0.00 | -0.67 | 0.00 | 0.4929 | 0.5712 | 1.1231 | 0.0135 | 0.0134 | 0.0248 |
| -0.67 | 0.25 | -0.67 | 0.25 | 0.4370 | 0.7064 | 1.1105 | 0.0155 | 0.0151 | 0.0255 |
| -0.67 | 0.25 | -0.67 | -0.25 | 0.4355 | 0.3487 | 1.1504 | 0.0155 | 0.0144 | 0.0253 |
| -0.67 | -0.25 | -0.67 | -0.25 | 0.3201 | 0.3622 | 1.1519 | 0.0145 | 0.0146 | 0.0253 |
| B: GTL-Disturbances are generated with $(\alpha, \delta) = (0.1436, 0)$ for $n = 250$ | | | | | | | | | |
| 0.1436 | 0.00 | 0.1436 | 0.00 | 0.107 | 0.111 | 0.110 | 0.108 | 0.113 | 0.126 |
| -0.67 | 0.00 | -0.67 | 0.00 | 0.049 | 0.049 | 0.110 | 0.049 | 0.050 | 0.125 |
| -0.67 | 0.25 | -0.67 | 0.25 | 0.051 | 0.051 | 0.113 | 0.052 | 0.052 | 0.129 |
| -0.67 | 0.25 | -0.67 | -0.25 | 0.051 | 0.051 | 0.113 | 0.052 | 0.051 | 0.128 |
| -0.67 | -0.25 | -0.67 | -0.25 | 0.052 | 0.051 | 0.113 | 0.052 | 0.051 | 0.128 |
| C: GTL-Disturbances are generated with $(\alpha, \delta) = (-0.67, 0.25)$ for $n = 5000$ | | | | | | | | | |
| 0.1436 | 0.00 | 0.1436 | 0.00 | 1.8968 | 1.5547 | 2.5085 | 0.0054 | 0.0054 | 0.0056 |
| 0.1436 | 0.00 | -0.33 | 0.00 | 1.9060 | 0.4835 | 2.5109 | 0.0054 | 0.0028 | 0.0056 |
| 0.1436 | 0.00 | -0.33 | 0.10 | 1.9043 | 0.4290 | 2.5738 | 0.0054 | 0.0026 | 0.0056 |
| 0.1436 | 0.00 | -0.67 | 0.00 | 1.9124 | 0.0693 | 2.5149 | 0.0054 | 0.0015 | 0.0056 |
| 0.1436 | 0.00 | -0.67 | 0.25 | 1.9111 | 0.0313 | 2.5252 | 0.0054 | 0.0009 | 0.0056 |
| -0.67 | 0.00 | -0.67 | 0.00 | 0.2560 | 0.0697 | 2.5241 | 0.0015 | 0.0015 | 0.0055 |
| -0.67 | 0.25 | -0.67 | 0.25 | 0.0940 | 0.0328 | 2.5778 | 0.0009 | 0.0009 | 0.0056 |
| -0.67 | 0.25 | -0.67 | -0.25 | 0.0940 | 0.0303 | 2.5676 | 0.0009 | 0.0010 | 0.0056 |
| -0.67 | -0.25 | -0.67 | -0.25 | 0.0784 | 0.0303 | 2.5334 | 0.0010 | 0.0010 | 0.0056 |

from around 0.33 to nearly 1. Across replications, the largest diagonal element of the hat-matrix (i.e., $X(X'X)^{-1}X'$) averages 0.0523 when both explanatory variables are standard normal and averages 0.7315 (ranging from 0.1223 to 0.9999) when both explanatory variables are drawn from a GTL(-0.67,0.25) distribution.

This leaves the question what happens when $n$ increases. In Panel C, we raise $n$ to 5000. If the estimator is $n^{1/2}$-consistent, the standard errors should diminish by a factor of $20^{1/2} = 4.47$, and the RMSE should follow suit since bias is not much of an issue (except for in the intercept). However, as $x$-variables are drawn from GTL($-0.33, \delta$) and GTL($-0.67, \delta$) for several values of $\delta$, they are so thick-tailed that the fourth moment of $x$ no longer exists

and Assumption A.4 is violated. Whenever $x_2$ is standard normal, the RMSE of $\hat{\beta}_{2OLS}$ rises with $n$, but when $x_2$ is GTL$(-0.67, \delta)$ for any $\delta$, it falls by a factor of approximately 4.5. As for the RMSE of $\hat{\beta}_{3OLS}$, it rises when $x_3$ is standard normal; it falls slightly when $x_3$ is GTL$(-0.33, \delta)$ for any $\delta$; and when $x_3$ is GTL$(-0.67, \delta)$ for any $\delta$, it falls by a factor of up to 22. The distinction between these cases is that the second moment of a GTL distribution is finite for GTL$(-0.33, \delta)$ but is not defined for GTL$(-0.67, \delta)$. It appears to be the relative thickness of the tails of the explanatory variables and the disturbances that determines the large-sample behavior of the OLS estimator.

As for the GTL estimator, the RMSE of $\hat{\beta}_{jOLS}$ decreases by a factor of approximately 4.5 when $x_j$ is standard normal or GTL$(-0.33, \delta)$, which indicates $\sqrt{n}$-consistency, and it decreases by a factor of 15 to 17.5 when $x_j$ is GTL$(-0.67, \delta)$, which might indicate a type of super consistency.

As with the baseline, we also examine the proximity to normality of the small-sample distribution of the estimators; see Tables E.4-E.6 in Appendix E. Thick-tailed explanatory variables cause problems for the OLS estimator even if $\epsilon$ is normally distributed. As for the GTL estimator, the slopes of thick-tailed explanatory variables also deviate from normality even for $n = 5000$, but the estimator of the GTL parameters $\alpha$, $\delta$, and $\sigma$ is still close to normally distributed, similar to the baseline case.

## 5 Applications

We apply the GTL regression model to five examples that are representative of many empirical research projects: a log-wage equation, a hedonic housing price equation, a study of the determinants of fines for speeding, the question whether preferential trade agreements are trade-creating or trade-diverting, and the familiar capital asset pricing model. In each case, we find some degree of skewness and a greater degree of kurtosis than the normal distribution allows for. In each example, we highlight the changes in the slopes estimates and the gain in

the precision of these estimates. Appendix F provides the data sources, variable definitions, and descriptive statistics.

## 5.1 Hourly wages

For the first example with real-world data, we estimate a log-wage equation that is specified in a typical fashion with data that are common. The data source is the Merged Outgoing Rotation Group (MORG) file of 1998, which is derived from information provided in the Current Population Survey. The sample consists of wage employees; self-employed workers are dropped from the MORG data. The sample is also restricted to individuals between the age of 16 and 65 and working full-time (between 30 and 70 hours in the week prior to the survey). The wage rate is computed as the earnings per week divided by the usual hours of work per week and refers to the current job of the respondent. Observations with a computed wage of less than \$1.00 per hour are discarded, as are observations for which information is allocated rather than directly measured. These restrictions leave us with a sample of 54,687 males and 46,045 females.

The log-wage equation contains the usual explanatory variables: years of schooling, a quadratic age profile (rather than experience which is likely endogenous for female workers), region, and race and ethnicity. Table 5 presents OLS and GTL estimates for male workers. Residuals of the OLS equation prove to be distinctly non-normally distributed: ($i$) a Jarque-Bera test examines the skewness and kurtosis of the residuals and rejects normality; ($ii$) a Wald test of the GTL estimates of ($\alpha, \delta$) indicates that ($\alpha, \delta$) is significantly different from $(0.1436, 0)$, with which the GTL distribution closely approximates a normal distribution; ($iii$) the $p$-value of the LM test is also less than 0.001, and there are as many as 46 range violations (nearly 1% of the sample); and ($iv$) a Vuong test that contrasts the OLS (modeled as a maximum likelihood estimator under normally distributed disturbances) and GTL estimators comes up strongly in favor of the GTL distribution. OLS estimates of the slopes deviate

Table 5: Log-wage equations, male wage and salaried workers, MORG 1998

| | OLS | | GTL | | $\frac{\hat{\beta}_{OLS}-\hat{\beta}_{GTL}}{\hat{\beta}_{GTL}}$ | $\frac{SE_{GTL}}{SE_{OLS}}$ |
| | Estimate | Stan.Err. | Estimate | Stan.Err. | | |
|---|---|---|---|---|---|---|
| Years of Schooling | 0.093 | 0.001 | 0.096 | 0.001 | -0.033 | 0.983 |
| Age$-16/10$ | 0.453 | 0.007 | 0.460 | 0.007 | -0.015 | 0.960 |
| $(\text{Age}-16)^2/100$ | -0.069 | 0.001 | -0.069 | 0.001 | 0.000 | 0.971 |
| MidAtla | 0.054 | 0.009 | 0.061 | 0.009 | -0.126 | 0.968 |
| EastNoC | 0.039 | 0.009 | 0.041 | 0.008 | -0.067 | 0.961 |
| WestNoC | -0.108 | 0.009 | -0.108 | 0.009 | 0.002 | 0.962 |
| SouthAtl | -0.031 | 0.009 | -0.032 | 0.008 | -0.030 | 0.965 |
| EastSoC | -0.105 | 0.011 | -0.104 | 0.011 | 0.008 | 0.958 |
| WestSoC | -0.082 | 0.010 | -0.080 | 0.009 | 0.020 | 0.969 |
| Mountain | -0.063 | 0.009 | -0.060 | 0.009 | 0.051 | 0.965 |
| Pacific | 0.051 | 0.009 | 0.055 | 0.009 | -0.078 | 0.970 |
| Afr. American | -0.175 | 0.007 | -0.176 | 0.007 | -0.006 | 0.967 |
| Hispanic | -0.185 | 0.007 | -0.186 | 0.007 | -0.003 | 0.970 |
| Asian | -0.124 | 0.010 | -0.116 | 0.010 | 0.064 | 1.000 |
| Indian | -0.141 | 0.019 | -0.134 | 0.018 | 0.043 | 0.970 |
| Intercept | 0.830 | 0.015 | 0.783 | 0.015 | | |
| $\sigma$ | | | 0.262 | 0.002 | | |
| $\alpha$ | | | 0.030 | 0.004 | | |
| $\delta$ | | | 0.029 | 0.002 | | |
| logL | -34199.7 | | -33548.0 | | | |
| (Absolute) Average | | | | | 0.038 | 0.969 |

Dependent variable: Log of hourly wage. Number of observations = 54674. Skewness and kurtosis of OLS residuals equal $-0.23$ and $4.26$; the Jarque-Bera test of normality of the OLS residuals has a $p$-value of less than 0.001. The Wald test of the GTL estimates of $(\alpha, \delta)$ equals 1127.5, rejecting normality with a $p$-value of less than 0.001. The LM test equals 583.7 with a $p$-value of less than 0.001 and with 42 range violations. The Vuong test that compares OLS and GTL equals $-11.16$ in favor of the GTL model with a $p$-value of less than 0.001.

by an average of 3.8 percent from the GTL estimates (which are preferred by virtue of the tests for normality); the largest deviation is nearly 12 percent. The GTL estimator is more precise, yielding an improvement of 3.1 percent in the standard errors. In Appendix G, Table G.1 offers the results of the same analysis of female workers: the gain achieved with GTL estimation is larger, as parameter estimates deviate by an average of 4.9 percent and GTL standard errors are 4.7 percent smaller. Figure 5 shows the estimated densities under the OLS (normal) and GTL estimation approaches for both male and female samples: GTL densities are thicker-tailed and left-skewed. The male-estimated GTL distribution implies a

Figure 5: Distributions estimated for male and female logwage equations



(a) Male                           (b) Female
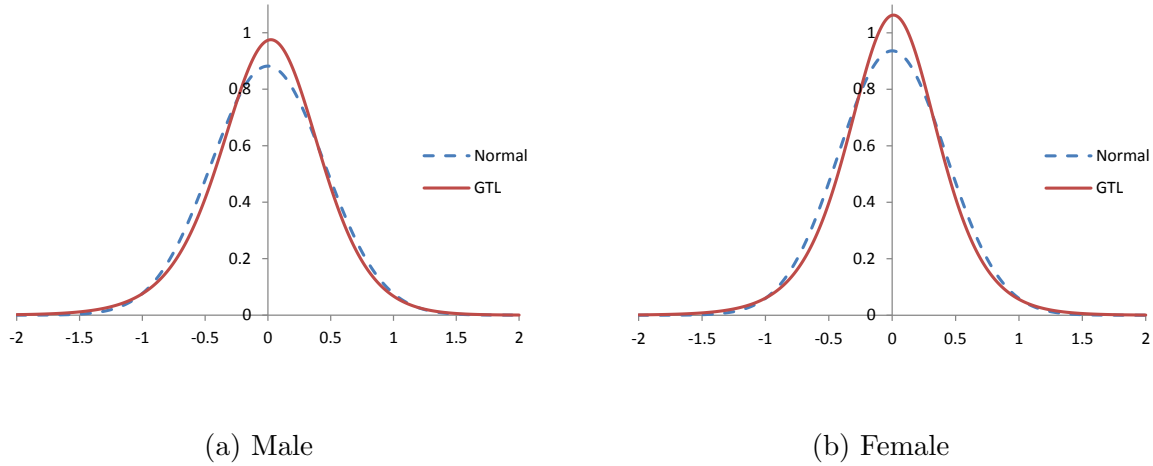
skewness $\kappa_3$ of $-0.21$ and a kurtosis $\kappa_4$ of 3.92; for females, the implied skewness and kurtosis are $-.11$ and 4.18.

## 5.2 Hedonic housing price equation

The second application concerns a hedonic housing price equation, estimated with data of housing sales in Windsor, Canada, in 1987. These data were first used in a study by Anglin and Gencay (1996) and subsequently in a nonparametric analysis by Parmeter et al. (2007) and Haupt et al. (2010), which led to the conclusion that the benchmark parametric model in Anglin and Gencay (1996) appears to be adequate. It is a slight variation of this parametric model that we reestimate now with GTL disturbances rather than with normally distributed disturbances.

For these data, the GTL estimator is preferred to the OLS estimator: the Jarque-Bera test rejects normality of the OLS residuals; the Wald and LM tests indicate that the GTL estimates $(\hat{\alpha}, \hat{\delta}) = (0.010, 0.034)$ differ significantly from $(0.1436, 0)$; and the Vuong test favors the GTL estimator. Because the sample is rather small (546 observations), the $p$-

Table 6: Hedonic housing price model, Windsor, Canada, 1987

| | OLS | | GTL | | $\frac{\hat{\beta}_{OLS}-\hat{\beta}_{GTL}}{\hat{\beta}_{GTL}}$ | $\frac{SE_{GTL}}{SE_{OLS}}$ |
|---|---|---|---|---|---|---|
| | Estimate | Stan.Err. | Estimate | Stan.Err. | | |
| lnLot | 0.302 | 0.026 | 0.292 | 0.025 | 0.033 | 0.957 |
| bdms3 | 0.067 | 0.024 | 0.065 | 0.023 | 0.020 | 0.941 |
| bdms4 | 0.075 | 0.032 | 0.062 | 0.031 | 0.210 | 0.970 |
| fb2 | 0.170 | 0.023 | 0.167 | 0.023 | 0.014 | 0.958 |
| fb3 | 0.347 | 0.065 | 0.385 | 0.064 | -0.100 | 0.970 |
| sty | 0.088 | 0.013 | 0.091 | 0.012 | -0.030 | 0.909 |
| drv | 0.112 | 0.028 | 0.115 | 0.027 | -0.021 | 0.956 |
| rec | 0.055 | 0.026 | 0.053 | 0.025 | 0.036 | 0.948 |
| ffin | 0.102 | 0.021 | 0.100 | 0.021 | 0.020 | 0.968 |
| ghw | 0.172 | 0.043 | 0.160 | 0.045 | 0.077 | 1.025 |
| ca | 0.163 | 0.021 | 0.158 | 0.020 | 0.037 | 0.944 |
| gar1 | 0.078 | 0.023 | 0.077 | 0.022 | 0.023 | 0.949 |
| gar2 | 0.104 | 0.024 | 0.095 | 0.024 | 0.095 | 0.966 |
| reg | 0.126 | 0.022 | 0.120 | 0.021 | 0.044 | 0.938 |
| Intercept | 7.972 | 0.214 | 8.065 | 0.208 | | |
| $\sigma$ | | | 0.116 | 0.009 | | |
| $\alpha$ | | | 0.010 | 0.050 | | |
| $\delta$ | | | 0.034 | 0.025 | | |
| log Likelihood | 85.46 | | 91.18 | | | |
| (Absolute) Average | | | | | 0.054 | 0.957 |

Dependent variable: Log of home price. Number of observations = 546. Skewness and kurtosis of OLS residuals equal $-0.19$ and 3.51; the Jarque-Bera test of normality of the OLS residuals equals 7.84 with a $p$-value of 0.02. The Wald test of the GTL estimates of $(\alpha, \delta)$ equals 8.50, rejecting normality with a $p$-value of less than 0.015. The LM test equals 8.17 with a $p$-value of 0.017 and with no range violations. The Vuong test that compares OLS and GTL equals $-1.69$ in favor of the GTL model with a $p$-value of less than 0.046.

values of these tests are not as low as 1 percent but this application shows that even in a sample of this small size ($n = 546$), it is possible to detect departures from normality. The effect on the slope estimates averages 5.4 percent, varying from 2 to 21 percent, and the GTL standard errors are 4.3 percent smaller than those of OLS. The GTL distribution is thicker-tailed ($\kappa_4 = 4.26$) and left-skewed ($\kappa_3 = -0.28$) relative to the normal distribution that best fits these data (Figure 6).

## 5.3 Speeding tickets, Massachusetts

Makowsky and Stratmann (2009) examine the determinants of traffic citations and fines for

Figure 6: Distributions estimated for the hedonic home price equation
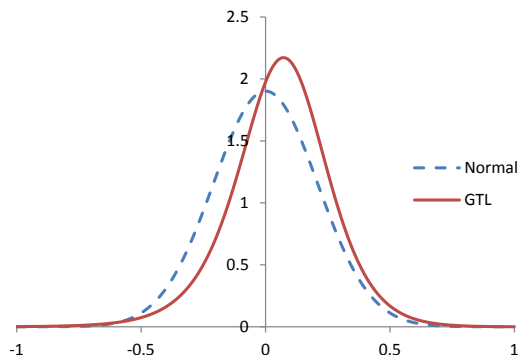


speeding, using a database that contains all speeding-related stops in Massachusetts from April 1, 2001 through May 31, 2001.

A traffic stop results in either a ticket or a warning. When a ticket is issued, a driver has to pay a fine. Whether a police officer issues a ticket or gives a warning is at the officer's discretion; in this database, about 46% of the 68,357 stops resulted in a speeding ticket. If a ticket is issued, state law provides a formula for the amount of the fine: $50 + $10 × (speed − (speed limit + 10)). Makowsky and Stratmann (2009, p.513) discuss the political economy hypothesis and the opportunity-cost hypothesis of officer behavior. The former relates the officers' decision to "the fiscal condition of the government that employs them and to whether the driver is a potential voter in local elections," and the latter predicts that "officers have a higher likelihood of issuing a ticket and issuing a larger fine amount when the opportunity cost for contesting the ticket is higher for drivers."

For the purpose of this paper, we ignore the selection issue whether a ticket is issued and concentrate on the amount of fine, which in this database averages $122. The regression model expresses this outcome variable in logarithmic form. The explanatory variables include the excess speed of the driver ("Mph over", in log form), driver characteristics (residence,

29

Table 7: Speeding tickets, Massachusetts, 2005

|  | OLS | | GTL | | $\frac{\hat{\beta}_{OLS}-\hat{\beta}_{GTL}}{\hat{\beta}_{GTL}}$ | $\frac{SE_{GTL}}{SE_{OLS}}$ |
|---|---|---|---|---|---|---|
|  | Estimate | Stan.Err. | Estimate | Stan.Err. |  |  |
| ln(Mph over) | 0.8649 | 0.0144 | 1.2332 | 0.0057 | -0.299 | 0.396 |
| OutTown | 0.0071 | 0.0107 | 0.0001 | 0.0006 | 119.524 | 0.054 |
| OutState | 0.0353 | 0.0070 | 0.0007 | 0.0004 | 52.026 | 0.062 |
| Afr.American | -0.0207 | 0.0095 | -0.0010 | 0.0004 | 20.402 | 0.045 |
| Hispanic | 0.0216 | 0.0101 | -0.0003 | 0.0007 | -85.838 | 0.064 |
| Female | -0.0564 | 0.0375 | 0.0003 | 0.0017 | -204.497 | 0.046 |
| ln(Age) | 0.0002 | 0.0073 | 0.0015 | 0.0007 | -0.878 | 0.092 |
| Female × ln(Age) | 0.0092 | 0.0105 | 0.0000 | 0.0005 | 946.064 | 0.049 |
| ln(CourtDist) | 0.0158 | 0.0035 | -0.0002 | 0.0003 | -81.445 | 0.074 |
| ln(Pvalue.pc) | -0.0043 | 0.0218 | -0.0007 | 0.0011 | 5.209 | 0.050 |
| OR | 0.0202 | 0.0700 | 0.0058 | 0.0020 | 2.502 | 0.028 |
| OR × OutTown | 0.0083 | 0.0630 | -0.0082 | 0.0019 | -2.008 | 0.031 |
| OR × ln(CourtDist) | 0.0019 | 0.0089 | 0.0003 | 0.0004 | 5.139 | 0.042 |
| SP | 0.0299 | 0.2949 | -0.0016 | 0.0155 | -19.662 | 0.053 |
| SP × OutTown | 0.0258 | 0.0187 | 0.0000 | 0.0008 | -4569.156 | 0.041 |
| SP × ln(CourtDist) | 0.0059 | 0.0040 | 0.0007 | 0.0004 | 7.853 | 0.101 |
| SP × ln(Pvalue.pc) | -0.0018 | 0.0265 | -0.0002 | 0.0014 | 9.790 | 0.051 |
| SP × OR | -0.0164 | 0.0319 | -0.0018 | 0.0017 | 8.270 | 0.053 |
| Intercept | 2.2716 | 0.2551 | 1.4817 | 0.0190 |  |  |
| $\sigma$ |  |  | 0.0030 | 0.0005 |  |  |
| $\alpha$ |  |  | -1.7612 | 0.0833 |  |  |
| $\delta$ |  |  | 1.3083 | 0.0483 |  |  |
| log Likelihood | -9440.55 |  | 25066.89 |  |  |  |
| (Absolute) Average |  |  |  |  | 360.723 | 0.075 |

Dependent variable: Log of amount of fine. Number of observations = 31674. Skewness and kurtosis of OLS residuals equal $-1.20$ and $5.07$; the Jarque-Bera test of normality of the OLS residuals has a $p$-value of less than 0.001. The Wald test of the GTL estimates of $(\alpha, \delta)$ equals 791.1, rejecting normality with a $p$-value of less than 0.001. The LM test equals 5072 with a $p$-value of less than 0.001 and with 8 range violations. The Vuong test that compares OLS and GTL equals $-52.67$ in favor of the GTL model with a $p$-value of less than 0.001.

race, ethnicity, gender, age, and the distance to court), and measures of the fiscal condition of a municipality (a dummy "OR" whether a municipality rejected a tax increase via an override referendum applicable to the operating budget of the 2001 fiscal year; property value per capita; and a dummy "SP" whether the traffic stop was made by a state police officer, who may have different incentives than a local police officer). The regression model includes several interactions as well, as indicated in the tables of results below.

The difference in the estimates is striking (Table 7). The first sign of the relative in-

Figure 7: Distributions estimated for the speeding ticket equation
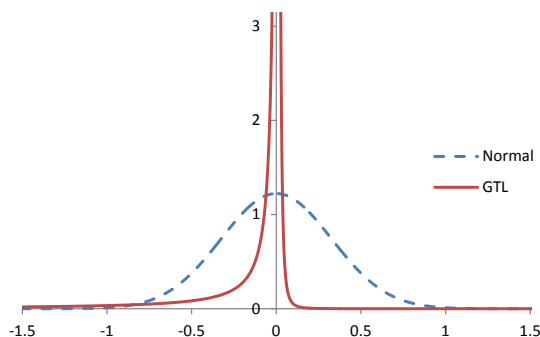


adequacy of the OLS estimator is seen in the skewness $(-1.20)$ and kurtosis $(5.07)$ of the residuals. Among the five applications discussed in this section, this application stands out in that none of the moments of the estimated GTL distribution exists. Obviously, then, the estimates of $(\alpha, \delta) = (-1.761, 1.308))$ are far from $(0.1436, 0)$ that represents normality. The disturbances are sharply peaked and have thick tails, with the left tail thicker than the right; see Figure 7 where the GTL density that peaks at 33.23 at $\epsilon = 0.0075$ is actually top-truncated to show the tails better. The two densities are drawn such that the median of both falls at 0. They intersect at $-0.8468$, $-0.0611$, $0.0362$, and $1.5974$, dividing the real axis in five segments. For normality, the probability of these segments equals 0.0047, 0.4210, 0.1185, 0.4558, and 4.9E-07; for GTL, the probability equals 0.1102, 0.1413, 0.7275, 0.0211, and 5.5E-06. This highlights the thick tails, especially on the left, and the great concentration of the GTL disturbances in the range from $-0.0611$ to $0.0362$. That means that the actual log-fine generally deviates little from the predicted log-fine—but if it deviates, it can deviate much.

About the determinants of the log-fine, both estimates show that excess speed (ln(Mph over)) is the major determinant, as it should be, but the two estimates have different im-

plications. The OLS slope estimate is less than 1, indicating that the fine is inelastic with respect to the severity of the speeding violation (miles over the speed limit). On the other hand, the GTL estimate indicates an elasticity greater than 1: the fine is elastic. This is more intuitive: a more severe speeding violation draws an increasingly severe penalty; this also corresponds with the prescription in state law.[14]

The other slopes are also much different. The OLS estimates suggest that African-Americans (and perhaps young females) pay lower fines, and that Hispanic and out-of-state drivers as well as those living farther away from the courthouse pay more. Regardless of whether these effects are intuitively plausible, the GTL estimates of none of these variables is economically significant any longer, even if a few of them are still statistically significant. In other words, the amount of fine is not varying at the discretion of the police officer in response to observable factors, in contrast with the findings by Makowsky and Stratmann (2009), but unobservable factors can occasionally cause major deviations from the fine that state law prescribes.

## 5.4 Trade creation and diversion

Many studies have examined the effect of preferential trade agreements (PTAs) on trade. Two trading partners signing on to the same PTA may trade more with each other than before and less with other partners who are not part of the PTA because of changes in the relative cost of trading. Thus, following work by Ghosh and Yamarik (2004) and Subramanianl and Wei (2007), Eicher et al. (2012) estimate a regression model that relates bilateral imports from country $j$ to country $i$ to ($a$) trade creation dummy variables that indicate whether countries $i$ and $j$ are both members of various PTAs and ($b$) trade diversion dummy variables that indicate whether only one of the $(i, j)$ country pair belongs to a given PTA, augmented

---

[14]Simple algebra with the formula for the amount of fine reveals that the elasticity exceeds 1 as long as the driver exceeded the speed limit by 5 miles, and the elasticity is not constant as it is in the estimated double-log model. In the entire sample, only 1% of the stopped drivers were going less than five miles over speed limit; one fifth of them received a ticket.

Figure 8: Distributions estimated for the trade creation and diversion equation
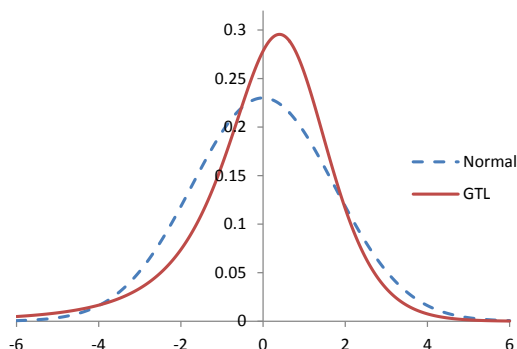


with $(c)$ various control variables such as the national income of each country, the distance between them, the colonial history between them, the similarity in language, differences in human capital and income level, indices of remoteness, trade policy, exchange rate volatility, and so forth. These variables are defined in Table F.4 in the Appendix, which also provides descriptive statistics.[15] The data describe trade between 97 countries arranged in 4069 importer-exporter country pairs over five-year intervals from 1960 to 2000 with a total of 37983 observations in an unbalanced panel.

The estimation results in Table 8 correspond with specification 2 in Table III of Eicher et al. (2012). Table 8 highlights the estimates of the trade creation and trade diversion effects, whereas further results concerning the slopes of the control variables may be found in Table G.2 in the Appendix. The normality assumption that is consistent with the OLS estimation method is rejected convincingly by every indicator; the distribution of the disturbances is left-skewed (as $\hat{\delta} > 0$, with $\kappa_3 = -2.34$) and quite thick-tailed (as $\hat{\alpha} < 0$, with $\kappa_4 = 40.17$). Figure 8 illustrates the density functions.

More importantly, the OLS and GTL slopes differ by an average of nearly 30 percent, and

---

[15]More precise definitions of the PTAs may be found in Ghosh and Yamarik (2004), Subramanianl and Wei (2007), and Eicher et al. (2012).

Table 8: Trade creation and diversion, 1960-2005

| | OLS | | GTL | | $\frac{\hat{\beta}_{OLS}-\hat{\beta}_{GTL}}{\hat{\beta}_{GTL}}$ | $\frac{SE_{GTL}}{SE_{OLS}}$ |
|---|---|---|---|---|---|---|
| | Estimate | Stan.Err. | Estimate | Stan.Err. | | |
| Trade creation dummy variables | | | | | | |
| tc.nafta | 0.371 | 0.252 | 0.500 | 0.220 | -0.257 | 0.874 |
| tc.eu | 0.427 | 0.115 | 0.196 | 0.098 | 1.182 | 0.853 |
| tc.efta | 0.685 | 0.129 | 0.518 | 0.116 | 0.323 | 0.896 |
| tc.eea | 0.179 | 0.095 | 0.261 | 0.079 | -0.312 | 0.837 |
| tc.caricom | 2.823 | 0.513 | 2.417 | 0.459 | 0.168 | 0.894 |
| tc.ap | 0.828 | 0.186 | 0.804 | 0.159 | 0.030 | 0.852 |
| tc.mercosur | 1.086 | 0.306 | 1.035 | 0.315 | 0.049 | 1.030 |
| tc.asean | 0.467 | 0.216 | 0.492 | 0.185 | -0.051 | 0.858 |
| tc.anzcerta | 0.969 | 0.141 | 0.748 | 0.130 | 0.295 | 0.920 |
| tc.apec | 1.599 | 0.095 | 1.291 | 0.085 | 0.238 | 0.889 |
| tc.laia | -0.133 | 0.141 | -0.432 | 0.134 | -0.691 | 0.950 |
| tc.cacm | 2.314 | 0.150 | 1.931 | 0.139 | 0.198 | 0.927 |
| tc.bilateralPTA | 0.110 | 0.128 | 0.098 | 0.117 | 0.122 | 0.916 |
| Trade diversion dummy variables | | | | | | |
| td.nafta | 0.151 | 0.073 | 0.081 | 0.061 | 0.875 | 0.841 |
| td.eu | 0.651 | 0.051 | 0.434 | 0.047 | 0.500 | 0.908 |
| td.efta | 0.376 | 0.059 | 0.202 | 0.054 | 0.866 | 0.921 |
| td.eea | -0.142 | 0.048 | -0.101 | 0.043 | 0.409 | 0.894 |
| td.caricom | -0.577 | 0.100 | -0.539 | 0.097 | 0.071 | 0.972 |
| td.ap | 0.105 | 0.074 | 0.115 | 0.068 | -0.088 | 0.925 |
| td.mercosur | 0.030 | 0.073 | -0.019 | 0.063 | -2.554 | 0.869 |
| td.asean | 0.474 | 0.070 | 0.395 | 0.061 | 0.198 | 0.869 |
| td.anzcerta | -0.759 | 0.098 | -0.657 | 0.086 | 0.156 | 0.879 |
| td.apec | 0.439 | 0.049 | 0.341 | 0.042 | 0.288 | 0.871 |
| td.laia | -0.561 | 0.060 | -0.533 | 0.054 | 0.051 | 0.913 |
| td.cacm | -0.174 | 0.078 | -0.120 | 0.074 | 0.451 | 0.945 |
| td.bilateralPTA | -0.292 | 0.054 | -0.275 | 0.045 | 0.064 | 0.832 |
| $\sigma$ | | | 0.809 | 0.013 | | |
| $\alpha$ | | | -0.078 | 0.008 | | |
| $\delta$ | | | 0.139 | 0.005 | | |
| log Likelihood | -74773.2 | | -72585.9 | | | |
| (Absolute) Average | | | | | 0.294 | 0.911 |

Dependent variable: Log of bilateral imports. The model also includes control variables (reported in Table G.2 in the Appendix) and time dummy variables (not reported). Number of observations = 37983. Skewness and kurtosis of OLS residuals equal $-0.71$ and $4.90$; the Jarque-Bera test of normality of the OLS residuals has a $p$-value of less than 0.001. The Wald test of the GTL estimates of $(\alpha, \delta)$ equals 1326.8, rejecting normality with a $p$-value of less than 0.001. The LM test equals 1299 with a $p$-value of less than 0.001 and with 24 range violations (all in the left tail). The Vuong test that compares OLS and GTL equals $-24.96$ in favor of the GTL model with a $p$-value of less than 0.001.

the standard errors of the GTL slopes are 8.9 percent smaller on average. The degree of trade

creation varies substantially by estimation method. For example, the European Union (EU)

effect drops from 42.7 percent to 19.6 percent; NAFTA rises from 37.1 percent to 50 percent; the effect of the ANZCERTA agreement between Australia and New Zealand is nearly one third smaller than the OLS suggests (though still large); and the *negative* effect of LAIA (the Latin America Free Trade Association/Latin America Integration Agreement) is tripled and statistically significant under GTL relative to OLS. As for the trade diversion effect of PTAs, the GTL estimates that correspond with the five positive and statistically significant OLS estimates are all substantially smaller, whereas all of the negative OLS slopes estimates are matched by similar GTL estimates. Thus, the evidence in favor of a PTA trade diversion effect is actually stronger than the OLS estimates suggest, although the GTL estimates still present a mixed picture.

Eicher et al. (2012) estimate more elaborate models than the one presented in Table 8. Their preferred model appears to be one with import-export-pair fixed effects in order to account for unobserved time-invariant country-pair heterogeneity (their specification 4 in Table IV). This introduces 7342 dummy variables, which in OLS estimation are dealt with by computing within-pair deviations. GTL estimates are obtained by maximum likelihood, which does not permit within-pair deviations as an estimation shortcut. This implies that, because of the short panel, the inconsistency of the fixed effects estimator spills over to the estimator of all other slopes.[16] Practically, however, estimating another 7342 slopes is not even feasible. As a substitute, we explore a specification with 36 continent import-export dummies, which are likely to pick up at least some of the unobserved time-invariant country-pair heterogeneity. These results are presented in Table G.3, which yields similar conclusions as Table 8, though the magnitude of the estimated trade effects differ.

---

[16]A country pair may be in the sample for a maximum of nine period. The average number of times a country pair appears is 5.17.

## 5.5 The extended CAPM model

The capital asset pricing model (CAPM) is a standard example of OLS regression analysis in econometrics; yet, it is estimated with financial data that are regularly found to be heavy-tailed. Thus, we compute both OLS estimates and GTL estimates of standard formulations of the CAPM model[17], of which we consider the standard one-factor model

$$R_{it} - R_{ft} = \beta_{1i} + \beta_{2i}(R_{mt} - R_{ft}) + \epsilon_{it} \tag{9}$$

and the extended three-factor model

$$R_{it} - R_{ft} = \beta_{1i} + \beta_{2i}(R_{mt} - R_{ft}) + \beta_{3i}F_{SMB,t} + \beta_{4i}F_{HML,t} + \epsilon_{it} \tag{10}$$

In these equations, $R_{it}$ is the return to portfolio $i$ in month $t$; $R_{ft}$ is the risk-free interest rate, which is the one-month treasury bill rate; $R_{mt}$ is the market return; $F_{SMB,t}$ is the factor that measures the difference in returns of portfolios of small and big stock; and $F_{HML,t}$ is the factor that measures the difference in returns of portfolios of stocks with high and low book-to-market ratios (e.g., Fama and French (2004)).

We use monthly data from the Fama-French website that pertain to the time period January 1960 - December 2012. Specifically, we examine the 25 portfolios that are constructed on the basis of the intersection of five size (market equity) and five book-to-market ratio portfolios.

For all 25 portfolios, residuals of the standard CAPM model are not normally distributed: the $p$-value of the Jarque-Bera test is less than 0.00001 for 24 of 25 cases. Panel A of Figure 9 shows why: estimated GTL regression models yield values of $(\alpha, \delta)$ that are far from $(0.1436, 0)$, which is marked with an $\times$, indicating that the distribution is considerably

---

[17]See, for example, Fama and French (1993, 2004) and Carhart (1997)

Figure 9: Estimates of $\alpha$ and $\delta$ of the basic and extended CAPM models
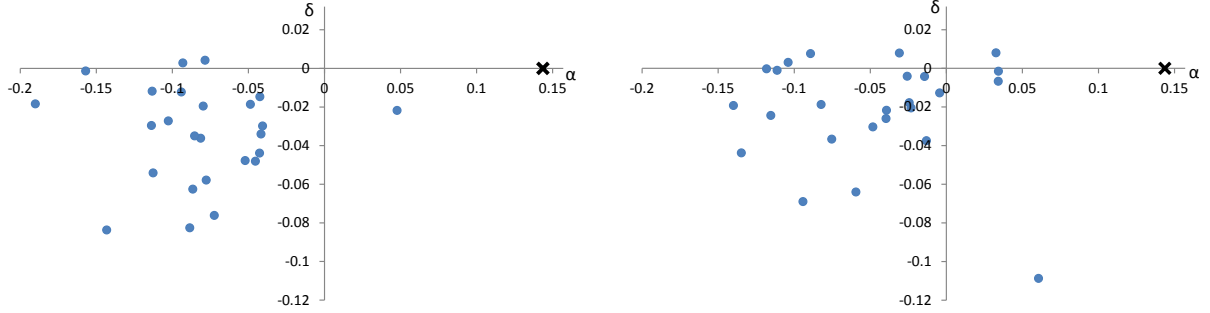


(a) $(\hat{\alpha}, \hat{\delta})$ for equation (9)



(b) $(\hat{\alpha}, \hat{\delta})$ for equation (10)

heavy-tailed ($\hat{\alpha} < 0$) and quite often right-skewed ($\hat{\delta} < 0$). The deviation from normality is only slightly less severe in the extended CAPM model: 18 of 25 residuals yield Jarque-Bera tests with a $p$-value of less than 0.00001, but still all $p$-values are less than 0.02; all Wald tests reject normality at a 1.2% significance level or better; and 17 of 25 Vuong tests favor GTL over OLS at a 5% two-tailed significance level.[18] For CAPM, the GTL-implied skewness $\kappa_3$ ranges from $-0.05$ to 1.97, averaging 0.49; kurtosis $\kappa_4$ runs from 3.68 to as high as 48.88 with an average of 9.43. All of these values are tend to be smaller for the extended CAPM model: $-0.10 \leq \kappa_3 \leq 0.83$ with an average of 0.25, and $3.77 \leq \kappa_4 \leq 12.54$ with an average of 6.16. Thus, the addition of the two factors $F_{SMB}$ and $F_{HML}$ remove a little of the extreme non-normality of the disturbances.

The question arises which types of portfolios the disturbances have the most skewed and thick-tailed distributions. Table 9 reports the estimates of $\alpha$ and $\delta$ by portfolio category

---

[18]The rejection of normality in these data is notable, since each separate portfolio blends the performance of, on average, 146 separate stocks (ranging from a minimum of 5 to a maximum of 1191 stocks). Such averaging might make a normality assumption plausible, but, as shown, normality still is soundly rejected.

Table 9: Basic CAPM and Extended CAPM: estimates of $\alpha$ and $\delta$ for 25-portfolio data

| | B/M:1 | B/M:2 | B/M:3 | B/M:4 | B/M:5 | B/M:1 | B/M:2 | B/M:3 | B/M:4 | B/M:5 |
|---|---|---|---|---|---|---|---|---|---|---|
| | Basic CAPM | | | | | | | | | |
| | | | $\hat{\alpha}$ | | | | | $\hat{\delta}$ | | |
| ME:1 | -0.143 | -0.113 | -0.087 | -0.089 | -0.072 | -0.084 | -0.054 | -0.063 | -0.083 | -0.076 |
| ME:2 | -0.049 | -0.043 | -0.042 | -0.041 | -0.114 | -0.019 | -0.044 | -0.034 | -0.030 | -0.030 |
| ME:3 | -0.093 | -0.042 | -0.103 | -0.080 | -0.078 | 0.003 | -0.015 | -0.027 | -0.020 | -0.058 |
| ME:4 | -0.081 | -0.085 | -0.190 | -0.045 | -0.052 | -0.036 | -0.035 | -0.018 | -0.048 | -0.048 |
| ME:5 | 0.048 | -0.079 | -0.113 | -0.157 | -0.094 | -0.022 | 0.004 | -0.012 | -0.001 | -0.012 |
| Average | | | -0.081 | | | | | -0.034 | | |
| | Extended CAPM | | | | | | | | | |
| | | | $\hat{\alpha}$ | | | | | $\hat{\delta}$ | | |
| ME:1 | -0.135 | -0.059 | -0.013 | 0.061 | -0.040 | -0.044 | -0.064 | -0.037 | -0.109 | -0.026 |
| ME:2 | -0.014 | -0.075 | -0.082 | -0.031 | -0.024 | -0.004 | -0.037 | -0.019 | 0.008 | -0.018 |
| ME:3 | 0.033 | -0.118 | -0.111 | -0.104 | -0.094 | 0.008 | 0.000 | -0.001 | 0.003 | -0.069 |
| ME:4 | -0.048 | -0.115 | -0.140 | -0.004 | 0.034 | -0.030 | -0.024 | -0.019 | -0.013 | -0.007 |
| ME:5 | 0.034 | -0.026 | -0.039 | -0.023 | -0.089 | -0.001 | -0.004 | -0.022 | -0.020 | 0.008 |
| Average | | | -0.049 | | | | | -0.022 | | |

Estimates of $\alpha$ and $\delta$ of the GTL densities belonging to the disturbance $\epsilon$ in equation (9) for the basic CAPM and equation (9) for the extended CAPM. B/M refers to the book-to-market equity ratio, divided into five groups; ME refers to market equity, divided into five groups.

(book-to-market and market size). Greater rightward skewness is found in the portfolios of the portfolios of smallest (ME:1) companies. Thicker-tailed distributions are found both for portfolios the smallest (ME:1) companies and for portfolios of median type companies in terms of both size (ME:3 and ME:4) and book-to-market ratio (B/M:2 and B/M:3).

A full description of the estimates of all 25 portfolios is impractical. Table 10 focuses on the comparison of OLS and GTL estimates of $\beta_2$ and $\beta_1$ (which is adjusted in the case of GTL; see section 3.1). As expected, the portfolios of smaller companies draw estimates of $\beta_2$ greater than 1. However, the GTL estimates tend to be greater than the OLS estimates for portfolios with the smallest B/M ratio and smaller for the other portfolio types; the intercepts ($\beta_1$) tend to be smaller for the smallest-B/M portfolio and larger for the other types.

Most significantly, the GTL estimates of the basic CAPM model tend to have smaller standard errors, especially for the portfolios built of smaller (ME:1-3 and B/M:1-2) com-

Table 10: Basic CAPM and Extended CAPM: estimates of $\beta_2$ and $\beta_1$ for 25-portfolio data

| | B/M:1 | B/M:2 | B/M:3 | B/M:4 | B/M:5 | B/M:1 | B/M:2 | B/M:3 | B/M:4 | B/M:5 |
|---|---|---|---|---|---|---|---|---|---|---|
| Basic CAPM | | | | | | | | | | |
| | | | $\hat{\beta}_{2,GTL}$ | | | | | $\hat{\beta}^*_{1,GTL}$ | | |
| ME:1 | 1.399 | 1.230 | 1.101 | 1.015 | 1.073 | -0.637 | 0.048 | 0.185 | 0.398 | 0.518 |
| ME:2 | 1.388 | 1.188 | 1.065 | 1.029 | 1.105 | -0.264 | 0.104 | 0.349 | 0.412 | 0.405 |
| ME:3 | 1.322 | 1.104 | 1.038 | 0.987 | 1.016 | -0.152 | 0.211 | 0.214 | 0.365 | 0.449 |
| ME:4 | 1.207 | 1.087 | 1.047 | 0.965 | 1.044 | -0.027 | 0.015 | 0.180 | 0.339 | 0.264 |
| ME:5 | 0.995 | 0.941 | 0.892 | 0.860 | 0.917 | -0.047 | 0.043 | 0.057 | 0.115 | 0.139 |
| | | $\hat{\beta}_{2,OLS} - \hat{\beta}_{2,GTL}$ | | | | | $\hat{\beta}_{1,OLS} - \hat{\beta}^*_{1,GTL}$ | | | |
| ME:1 | 0.025 | -0.004 | -0.010 | -0.001 | 0.003 | 0.192 | 0.105 | 0.086 | 0.087 | 0.072 |
| ME:2 | 0.007 | -0.018 | -0.010 | -0.015 | 0.007 | 0.008 | 0.023 | 0.017 | 0.019 | 0.048 |
| ME:3 | 0.007 | 0.006 | -0.035 | -0.026 | 0.010 | -0.019 | -0.003 | 0.040 | 0.022 | 0.054 |
| ME:4 | 0.014 | -0.011 | -0.022 | -0.013 | -0.006 | 0.005 | 0.023 | 0.026 | 0.022 | 0.038 |
| ME:5 | 0.000 | -0.009 | -0.023 | -0.033 | -0.028 | -0.006 | 0.005 | 0.023 | 0.014 | 0.037 |
| | | $SE(\hat{\beta}_{2,GTL})/SE(\hat{\beta}_{2,OLS})$ | | | | | $SE(\hat{\beta}^*_{1,GTL})/SE(\hat{\beta}_{1,OLS})$ | | | |
| ME:1 | 0.856 | 0.949 | 0.883 | 0.900 | 1.024 | 0.862 | 0.946 | 0.896 | 0.888 | 1.029 |
| ME:2 | 0.920 | 1.024 | 1.016 | 1.038 | 0.959 | 0.866 | 0.967 | 0.963 | 0.932 | 0.922 |
| ME:3 | 0.996 | 1.066 | 0.983 | 0.886 | 0.952 | 0.927 | 0.972 | 0.916 | 0.815 | 0.898 |
| ME:4 | 1.017 | 1.080 | 1.001 | 1.013 | 0.951 | 0.939 | 0.973 | 0.931 | 0.963 | 0.838 |
| ME:5 | 1.015 | 1.024 | 0.986 | 0.985 | 0.943 | 0.960 | 0.907 | 0.943 | 0.964 | 0.915 |
| Average | | | 0.979 | | | | | 0.925 | | |
| Extended CAPM | | | | | | | | | | |
| | | | $\hat{\beta}_{2,GTL}$ | | | | | $\hat{\beta}^*_{1,GTL}$ | | |
| ME:1 | 1.076 | 0.968 | 0.921 | 0.888 | 0.970 | -0.600 | -0.142 | -0.054 | -0.007 | 0.082 |
| ME:2 | 1.107 | 1.015 | 0.957 | 0.970 | 1.079 | -0.200 | -0.076 | 0.072 | 0.090 | -0.053 |
| ME:3 | 1.090 | 1.003 | 0.973 | 0.982 | 1.047 | -0.035 | 0.108 | 0.029 | 0.076 | -0.077 |
| ME:4 | 1.062 | 1.045 | 1.040 | 1.022 | 1.132 | 0.089 | -0.101 | -0.023 | 0.023 | -0.128 |
| ME:5 | 0.977 | 1.003 | 0.974 | 0.992 | 1.059 | 0.146 | 0.022 | -0.069 | -0.162 | -0.182 |
| | | $\hat{\beta}_{2,OLS} - \hat{\beta}_{2,GTL}$ | | | | | $\hat{\beta}_{1,OLS} - \hat{\beta}^*_{1,GTL}$ | | | |
| ME:1 | 0.003 | -0.007 | -0.007 | -0.005 | 0.010 | 0.046 | 0.015 | -0.001 | -0.049 | 0.006 |
| ME:2 | 0.008 | -0.004 | 0.009 | 0.000 | 0.004 | -0.005 | -0.010 | -0.007 | -0.008 | -0.006 |
| ME:3 | 0.000 | 0.030 | 0.022 | 0.012 | 0.014 | -0.004 | -0.053 | -0.039 | -0.026 | 0.043 |
| ME:4 | -0.005 | 0.025 | 0.036 | -0.007 | 0.008 | 0.002 | -0.018 | -0.032 | 0.011 | -0.004 |
| ME:5 | -0.002 | -0.004 | -0.001 | -0.006 | -0.002 | 0.004 | -0.001 | 0.001 | -0.005 | 0.013 |
| | | $SE(\hat{\beta}_{2,GTL})/SE(\hat{\beta}_{2,OLS})$ | | | | | $SE(\hat{\beta}^*_{1,GTL})/SE(\hat{\beta}_{1,OLS})$ | | | |
| ME:1 | 0.853 | 0.971 | 1.024 | 0.974 | 1.012 | 0.858 | 0.984 | 1.014 | 0.957 | 1.015 |
| ME:2 | 0.941 | 0.953 | 0.967 | 1.008 | 1.027 | 0.949 | 0.930 | 0.894 | 0.908 | 0.976 |
| ME:3 | 0.970 | 0.927 | 0.942 | 0.924 | 0.982 | 0.983 | 0.921 | 0.907 | 0.876 | 0.959 |
| ME:4 | 0.935 | 0.974 | 0.905 | 1.034 | 0.967 | 1.118 | 0.970 | 0.901 | 0.996 | 0.977 |
| ME:5 | 0.993 | 0.999 | 0.911 | 1.019 | 0.967 | 0.967 | 0.980 | 0.917 | 1.016 | 0.927 |
| Average | | | 0.967 | | | | | 0.956 | | |

Estimates of $\beta_2$ and $\beta_1$ of the regression models in equation (9) for the basic CAPM and equation (10) for the extended CAPM. B/M refers to the book-to-market equity ratio, divided into five groups; ME refers to market equity, divided into five groups. SE refers to the standard error of the indicated estimator.

Table 11: CAPM model of a (ME:1,B/M:1) portfolio, Jan.1960 - Dec.2012

| | OLS | | GTL | | $\hat{\beta}_{OLS} - \hat{\beta}_{GTL}$ | $\frac{SE_{GTL}}{SE_{OLS}}$ |
| | Estimate | Stan.Err. | Estimate | Stan.Err. | | |
|---|---|---|---|---|---|---|
| A: Simple CAPM | | | | | | |
| $R_m - R_f$ | 1.423 | 0.043 | 1.399 | 0.037 | 0.025 | 0.856 |
| Intercept | -0.445 | 0.196 | -0.637 | 0.169 | 0.192 | 0.862 |
| $\sigma$ | | | 2.000 | 0.138 | | |
| $\alpha$ | | | -0.143 | 0.044 | | |
| $\delta$ | | | -0.084 | 0.026 | | |
| log-likelihood | -1914.1 | | -1852.3 | | | |
| Average | | | | | | 0.859 |
| B: Extended CAPM | | | | | | |
| $R_m - R_f$ | 1.079 | 0.025 | 1.076 | 0.021 | 0.003 | 0.853 |
| $F_{SMB}$ | 1.377 | 0.035 | 1.359 | 0.036 | 0.019 | 1.015 |
| $F_{HML}$ | -0.278 | 0.038 | -0.285 | 0.034 | 0.006 | 0.907 |
| Intercept | -0.481 | 0.104 | -0.527 | 0.089 | 0.046 | 0.858 |
| $\sigma$ | | | 1.068 | 0.069 | | |
| $\alpha$ | 0.144 | | -0.135 | 0.041 | | |
| $\delta$ | 0.000 | | -0.044 | 0.026 | | |
| log-likelihood | -1498.3 | | -1443.0 | | | |
| Average | | | | | | 0.908 |

Dependent variable: Monthly return net of a risk-free interest rate in a portfolio constructed from companies in the lowest market size (ME:1) and lowest book-to-market ratio (B/M:1) category. Number of observations = 636. Skewness and kurtosis of OLS residuals equal 1.14 and 9.75 in the case of CAPM and 0.74 and 9.11 in the case of the extended CAPM; for both models, the Jarque-Bera test of normality of the OLS residuals has a $p$-value of less than 0.001. The Wald test of the GTL estimates of $(\alpha, \delta)$ equals 47.6 for both models, rejecting normality with a $p$-value of less than 0.001. The LM test equals 42.7 (1 range violation) and 48.7 (2 range violations), both with a $p$-value of less than 0.001. The Vuong tests that compares OLS and GTL equal $-2.63$ and $-2.72$, respectively, in favor of the GTL model with a $p$-value of 0.005 and 0.004.

panies in particular. Overall, the standard errors for the extended CAPM model are 3.3% smaller for the estimates of $\beta_2$ and 4.4% smaller for the estimates of $\beta_1$.

Table 11 provides a more detailed report of the estimates of the basic and extended CAPM models under OLS and GTL estimation for the portfolio data that appear to be more affected by non-normality than most: the portfolio constructed from companies in the lowest market size (ME:1) and lowest book-to-market ratio (B/M:1) category. The slope estimates are quite robust to the difference in estimation approach; the intercept is relatively more sensitive; and the standard errors for this portfolio are at least 10% smaller. The addition of the factors $F_{SMB}$ and $F_{HML}$ has virtually the same impact on the estimate

Figure 10: Normal and GTL densities for CAPM and Extended CAPM models, portfolio (ME:1, B/M:1)
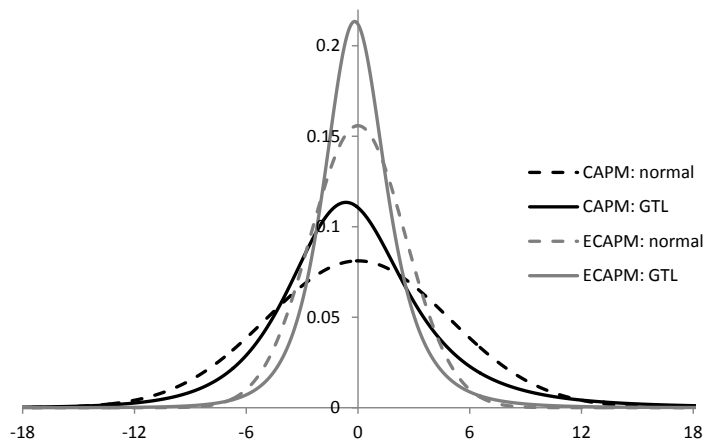


of $\beta_2$ whether the equation is estimated with OLS or with GTL.

Figure 10 shows the estimated densities of the disturbances. The pair of densities of the extended CAPM model are less spread out than those of the basic CAPM model, thanks to the addition of the $F_{SMB}$ and $F_{HML}$ factors to the model that are effective in explaining variation in the returns of this portfolio. For both the basic and extended CAPM models, the difference between the normal and GTL densities is dominated more by the difference in kurtosis than the difference in skewness—GTL has a higher mode and thicker tails and also is right-skewed, whereas the normal density is symmetric.[19] The scale of the GTL density may seem to be narrower than the normal density for both pairs, but the standard deviation of the GTL and normal distribution is, respectively, 5.02 and 4.91 for the CAPM model and 2.54 and 2.56 for the extended CAPM model. Thus, the concentration of GTL's probability mass around the mode (near 0) is compensated for by long tails, especially on the right side

---

[19]The GTL-estimated $\kappa_3$ and $\kappa_4$ equal 1.97 and 48.88 for the basic CAPM model and 0.83 and 12.54 for the extended CAPM model, as opposed to 0 and 3 under normality.

since the distributions are right-skewed.

## 6   Concluding remarks

The classical linear regression model is the first model novice applied econometricians are exposed to; the next concept is often the Gauss-Markov theorem that states that the OLS estimator is the best linear unbiased estimator. Out of the latter four words, applied econometricians tend to remember the word "best" most vividly. As a result, OLS is the workhorse approach to linear regression models, to be discarded only if there is a clear violation of its basic assumptions, such as endogenous regressors.

Skewed and thick-tailed disturbances do not constitute a violation of those assumptions, but OLS is no longer the best estimator. To be sure, if the moments of the distribution of the disturbances exist, OLS is still the best linear unbiased estimator, but ($i$) those moments do not always exist, and ($ii$) a nonlinear estimator may have a smaller variance than OLS.

This paper develops the maximum likelihood estimator of a linear regression model with GTL-distributed disturbances. Statistically, it has good properties (consistent and asymptotically normal). Monte Carlo comparisons demonstrate its dominance when disturbances are non-normally distributed. Five applications highlight the practical relevance of the GTL regression approach. First, they illustrate the fact that disturbances are often not normally distributed but rather exhibit skewness and a higher degree of kurtosis than normality. Second, they illustrate that in some research problems the location parameters (the slopes and intercept of the regression model) are highly sensitive to variations in the distributional assumption and in other research problems these parameters are very robust. This difference in the effect of non-normality is *a priori* not foreseeable; the only way to find out is to test for normality of the residuals of OLS and, if normality is rejected, to estimate the regression model with a GTL estimator. Third, the applications illustrate that, robust or not, the location parameters can be estimated with greater precision with a GTL estimator. Effi-

ciency gains vary between research applications and generally are larger when disturbances are more distinctly non-normal.

Disturbances need not be GTL-distributed. The assumption of a GTL distribution is one of several alternatives the applied econometrician can choose from when confronted with non-normal disturbances. But the GTL distribution has much to commend it: it is highly flexible and can accommodate both thin (truncated) tails and thick tails as well any shape between symmetry and extreme degrees of skewness. This flexibility also makes the GTL distribution appealing for, e.g., discrete choice and GARCH modelling, where, just as in linear regression models, applied econometricians may benefit from a wider range of modelling tools.

# Appendices

## A  Proof of Theorem 1

The proof of Theorem 1 relies on Theorem 2.5 of Newey and McFadden (1994), which consists on three components. First, $\epsilon_i$ must be i.i.d., as we indeed assume. Second, the log-likelihood function must be continuous at every $\theta$ with probability 1, which is trivial to verify. Third, $\theta$ must be identified. Lemma 2.2 of Newey and McFadden (1994) provides conditions for identification:

> If $\theta_0$ is identified and $E[|\ln g(z|\theta)|] < \infty$ for all $\theta$, then $Q_0(\theta) = E[\ln g(z|\theta)]$ has a unique maximum at $\theta_0$.

We shall first check whether $E[|\ln g(z|\theta)|] < \infty$ for all $\theta$.

The regression model is given in equation (3). Define the range of $\epsilon$ as $\mathcal{E}(\theta)$. Define $z = (y, x)$ with $x \in \mathcal{X}$ and $y \in \mathcal{Y}(x, \theta)$, where $\mathcal{Y}(x, \theta)$ is $\mathcal{E}(\theta)$ shifted by $x'\beta$. Given the independence of $x$ and $\epsilon$, $g(z) = g_y(y|x, \theta)g_x(x)$ with $g_y(y|x, \theta) = \frac{1}{\sigma}f(\frac{y-x'\beta}{\sigma}|\theta)$, where $f(\cdot|\theta)$ is the GTL density. Successive transformation of variables ($\epsilon = (y - x'\beta)/\sigma$ and $u = G^{-1}(\epsilon)$) yields

$$
\begin{aligned}
E[|\ln g(z|\theta)|] &= \int_{\mathcal{X}} \int_{\mathcal{Y}(x,\theta)} |\ln g(y, x|\theta)| g(y, x|\theta) dy dx & \text{(A.1)} \\
&= \int_{\mathcal{X}} \int_{\mathcal{E}(\theta)} |\ln\{\frac{1}{\sigma}f(\epsilon|\theta)g_x(x)\}| f(\epsilon|\theta)g_x(x) d\epsilon dx & \text{(A.2)} \\
&= \int_{\mathcal{X}} \left( \int_0^1 |-\ln\sigma - \ln G'(u) + \ln g_x(x)| du \right) g_x(x) dx. & \text{(A.3)}
\end{aligned}
$$

since the density of $u$ is uniform on $[0, 1]$. With $\sigma > 0$ and $x$ being properly distributed as a vector of explanatory variables, the focus of equation (A.3) is on $G'(u)$. Thus, write the

term in the large parentheses as

$$\int_0^1 |A - \ln G'(u)| du = \int_0^1 |A - \ln(u^{\lambda_1 - 1} + (1-u)^{\lambda_2 - 1})| du. \qquad (A.4)$$

$G'(u)$ is well-behaved for any $u \in [u_\eta, 1 - u_\eta]$ for a small $u_\eta$ but may have asymptotes at $u = 0$ and/or $u = 1$. Thus, we must check for the behavior of $G'(u)$ in the intervals $[0, u_\eta]$ and $[1 - u_\eta, 1]$. For $\lambda_1 > 1$ and $u \downarrow 0$, $G'(u) \to 1$ since $u^{\lambda_1 - 1} \downarrow 0$ and $(1-u)^{\lambda_2 - 1} \to 1$. For $\lambda_1 = 1$, $u^{\lambda_1 - 1} = 1$, so as $u \downarrow 0$, $G'(u) \to 2$ since $(1-u)^{\lambda_2 - 1} \to 1$. For $\lambda_1 < 1$, consider that $G'(u) \to \infty$ as $u \downarrow 0$. Thus, choose $u_\eta$ such that $\ln G'(u_\eta) \geq A$. Rewrite $-\ln G'(u) = -\ln(u^{\lambda_1 - 1}(1 + u^{1 - \lambda_1}(1-u)^{\lambda_2 - 1})) \equiv -((\lambda_1 - 1)\ln u + \ln B(u))$, where $B(0) = 1$ and $B(u) > 1$ for $u \in (0, u_\eta)$ for a small enough $u_\eta$. Thus, since $A - \ln G'(u) \leq 0$ for all $u \in [0, u_\eta]$, we have

$$\int_0^{u_\eta} |A - \ln G'(u)| du = \int_0^{u_\eta} ((\lambda_1 - 1)\ln u + \ln B(u) - A) du < \int_0^{u_\eta} ((\lambda_1 - 1)\ln u - A) du + u_\eta \ln B(u_\eta) < \infty,$$

since $\int_0^{u_\eta} \ln u = [u \ln u - u]_0^{u_\eta} = u_\eta \ln u_\eta - u_\eta$ is finite. Thus, $G'(u)$ is sufficiently well-behaved for any $\lambda_1$ at the left bound of the interval $[0, 1]$. By a similar argument, the right bound yields no problems for any $\lambda_2$ either. Thus, $E[|\ln g(z|\theta)|] < \infty$ for all $\theta$.

The other part of Lemma 2.2 of Newey and McFadden (1994) refers to identification of $\theta_0$. As indicated in Section 2, the GTL$(\alpha, \delta)$ distribution turns into a uniform distribution for four parameter pairs: $(\alpha, \delta) = (1, 0), (2, 0), (\alpha, 1 - \alpha)$ for $\alpha \to \infty$, and $(\alpha, \alpha - 1)$ for $\alpha \to \infty$. If the parameter space $\Theta$ is compact, the latter two parameter pairs are immediately ruled out, but that still leaves two parameter pairs: $(1, 0)$ and $(2, 0)$. Thus, $\beta_0$ and $\sigma_0$ are always identified; $\alpha_0$ and $\delta_0$ are identified anywhere in a compact $\Theta$ except at $(1, 0)$ and $(2, 0)$. Note that Theorem 2 implies that these two parameter pairs fall in a region where the MLE

45

estimator is not asymptotically normally distributed.

## B   Proof of Lemmas 1 and 3

In this appendix, we first examine derivatives of the function $G(u)$ and then list the derivatives of the log-likelihood function. This leads directly to the proof of Lemma 1 and sets up the examination of the second order derivatives for 3.

We have $G(u) = \frac{1}{\lambda_1}(u^{\lambda_1} - 1) - \frac{1}{\lambda_2}((1-u)^{\lambda_2} - 1)$, where $\lambda_1 = \alpha - \delta$ and $\lambda_2 = \alpha + \delta$. In the following, apostrophes refer to derivatives with respect to $u$, and subscripts indicate derivatives with respect to parameters.

$$G'(u) = u^{\lambda_1 - 1} + (1-u)^{\lambda_2 - 1} \tag{B.1}$$

$$G''(u) = (\lambda_1 - 1)u^{\lambda_1 - 2} - (\lambda_2 - 1)(1-u)^{\lambda_2 - 2} \tag{B.2}$$

$$G'''(u) = (\lambda_1 - 1)(\lambda_1 - 2)u^{\lambda_1 - 3} + (\lambda_2 - 1)(\lambda_2 - 2)(1-u)^{\lambda_2 - 3} \tag{B.3}$$

$$G_\alpha(u) = \frac{1}{\lambda_1}u^{\lambda_1} \ln u - \frac{1}{\lambda_1^2}(u^{\lambda_1} - 1) - \frac{1}{\lambda_2}(1-u)^{\lambda_1} \ln(1-u) + \frac{1}{\lambda_2^2}((1-u)^{\lambda_2} - 1) \tag{B.4}$$

$$G_\alpha'(u) = u^{\lambda_1 - 1} \ln u + (1-u)^{\lambda_2 - 1} \ln(1-u) \tag{B.5}$$

$$G_\alpha''(u) = u^{\lambda_1 - 2}(1 + (\lambda_1 - 1) \ln u) - (1-u)^{\lambda_2 - 2}(1 + (\lambda_2 - 1) \ln(1-u)) \tag{B.6}$$

$$\begin{aligned} G_{\alpha\alpha}(u) = {} & \frac{1}{\lambda_1}u^{\lambda_1} \ln u \left( \ln u - \frac{2}{\lambda_1} \right) + \frac{2}{\lambda_1^3}(u^{\lambda_1} - 1) \\ & - \frac{1}{\lambda_2}(1-u)^{\lambda_2} \ln(1-u)\left( \ln(1-u) - \frac{2}{\lambda_2} \right) - \frac{2}{\lambda_2^3}((1-u)^{\lambda_2} - 1) \end{aligned} \tag{B.7}$$

$$G_{\alpha\alpha}'(u) = u^{\lambda_1 - 1}(\ln u)^2 + (1-u)^{\lambda_2 - 1}(\ln(1-u))^2 \tag{B.8}$$

$$G_\delta(u) = -\frac{1}{\lambda_1}u^{\lambda_1} \ln u + \frac{1}{\lambda_1^2}(u^{\lambda_1} - 1) - \frac{1}{\lambda_2}(1-u)^{\lambda_1} \ln(1-u) + \frac{1}{\lambda_2^2}((1-u)^{\lambda_2} - 1) \tag{B.9}$$

$$G_\delta'(u) = -u^{\lambda_1 - 1} \ln u + (1-u)^{\lambda_2 - 1} \ln(1-u) \tag{B.10}$$

$$G_\delta''(u) = -u^{\lambda_1 - 2}(1 + (\lambda_1 - 1) \ln u) - (1-u)^{\lambda_2 - 2}(1 + (\lambda_2 - 1) \ln(1-u)) \tag{B.11}$$

$$\begin{aligned} G_{\delta\delta}(u) = {} & \frac{1}{\lambda_1}u^{\lambda_1} \ln u \left( \ln u - \frac{2}{\lambda_1} \right) + \frac{2}{\lambda_1^3}(u^{\lambda_1} - 1) \\ & - \frac{1}{\lambda_2}(1-u)^{\lambda_2} \ln(1-u)\left( \ln(1-u) - \frac{2}{\lambda_2} \right) - \frac{2}{\lambda_2^3}((1-u)^{\lambda_2} - 1) \end{aligned} \tag{B.12}$$

$$G'_{\delta\delta}(u) \;=\; u^{\lambda_1-1}(\ln u)^2 + (1-u)^{\lambda_2-1}(\ln(1-u))^2 \tag{B.13}$$

$$\begin{aligned}
G_{\alpha\delta}(u) \;=\;& -\frac{1}{\lambda_1}u^{\lambda_1}\ln u\Big(\ln u - \frac{2}{\lambda_1}\Big) - \frac{2}{\lambda_1^3}(u^{\lambda_1}-1) \\
& -\frac{1}{\lambda_2}(1-u)^{\lambda_2}\ln(1-u)\Big(\ln(1-u) - \frac{2}{\lambda_2}\Big) - \frac{2}{\lambda_2^3}((1-u)^{\lambda_2}-1)
\end{aligned} \tag{B.14}$$

$$G'_{\alpha\delta}(u) \;=\; -u^{\lambda_1-1}(\ln u)^2 + (1-u)^{\lambda_2-1}(\ln(1-u))^2 \tag{B.15}$$

$\lambda_1$ and $\lambda_2$ appear in the denominator in $G$, $G_\alpha$, $G_\delta$, $G_{\alpha\alpha}$, $G_{\alpha\delta}$ and $G_{\delta\delta}$. Nevertheless, these functions are well-defined when $\lambda_1 \to 0$ or $\lambda_2 \to 0$. Because of symmetry, it is necessary to show this only for $\lambda_1 \to 0$. In $G(u)$, we have

$$\lim_{\lambda_1\to 0}\frac{u^{\lambda_1}-1}{\lambda_1} = \frac{u^{\lambda_1}\ln u}{1} = \ln u. \tag{B.16}$$

by L'Hôpital's Rule. Similarly, in $G_\alpha(u)$ and $G_\delta(u)$, we find

$$\lim_{\lambda_1\to 0}\frac{\lambda_1 u^{\lambda_1}\ln u - (u^{\lambda_1}-1)}{\lambda_1^2} = \lim_{\lambda_1\to 0}\frac{u^{\lambda_1}\ln u + \lambda_1 u^{\lambda_1}\ln u - u^{\lambda_1}\ln u}{2\lambda_1} = \frac{1}{2}(\ln u)^2, \tag{B.17}$$

and in regard to $G_{\alpha\alpha}(u)$, $G_{\alpha\delta}(u)$ and $G_{\delta\delta}(u)$, we find

$$\lim_{\lambda_1\to 0}\frac{\lambda_1^2 u^{\lambda_1}(\ln u)^2 - 2\lambda_1 u^{\lambda_1}\ln u + 2(u^{\lambda_1}-1)}{\lambda_1^3} = \lim_{\lambda_1\to 0}\frac{\lambda_1^2 u^{\lambda_1}(\ln u)^3}{3\lambda_1^2} = \frac{1}{3}(\ln u)^3. \tag{B.18}$$

A more serious issue exists at $\lambda_1 = 1$. For example, as $u \downarrow 0$, $G'(u) \to 1$ when $\lambda > 1$; $G'(u) \to 2$ when $\lambda_1 = 1$; and $G'(u) \to \infty$ when $\lambda < 1$. Thus, $G'(0)$ is not a continuous function of $\lambda_1$. Other derivatives ($G'_\alpha$ and others) are similarly impacted at $\lambda_1 = 1$. Moreover, $G''(0)$ is discontinuous in $\lambda_1$ at $\lambda_1 = 2$ and $G'''(0)$ is discontinuous in $\lambda_1$ at $\lambda_1 = 3$. Similar discontinuities exist at $u = 1$ for $\lambda_2 = 1, 2, 3$.

Next, we present the first and second order derivatives of the log-likelihood function: $L = \sum_{i=1}^{n} \ell_i(\theta)$. For ease of notation, we drop the argument $u_i$ from the $G$-function and its

derivatives. The first order derivatives of $\ell_i(\theta)$ are:

$$\nabla_\beta \ell_i \;=\; \frac{1}{\sigma}\frac{G''}{(G')^2}x_i \tag{B.19}$$

$$\nabla_\sigma \ell_i \;=\; -\frac{1}{\sigma}+\frac{1}{\sigma^2}\frac{G''}{(G')^2}(y_i-x_i\beta) \tag{B.20}$$

$$\nabla_\alpha \ell_i \;=\; \frac{G''G_\alpha}{(G')^2}-\frac{G'_\alpha}{G'} \tag{B.21}$$

$$\nabla_\delta \ell_i \;=\; \frac{G''G_\delta}{(G')^2}-\frac{G'_\delta}{G'} \tag{B.22}$$

The second order derivatives are:

$$\nabla_{\beta\beta}\ell_i \;=\; -\frac{1}{\sigma^2}\frac{G'G'''-2(G'')^2}{(G')^4}x_ix_i' \tag{B.23}$$

$$\nabla_{\beta\sigma}\ell_i \;=\; -\frac{1}{\sigma}\nabla_\beta\ell_i-\frac{1}{\sigma^3}\frac{G'G'''-2(G'')^2}{(G')^4}(y_i-x_i\beta)x_i \tag{B.24}$$

$$\nabla_{\beta\alpha}\ell_i \;=\; -\frac{1}{\sigma}\frac{G'G'''-2(G'')^2}{(G')^4}G_\alpha x_i+\frac{G'G''_\alpha-2G''G'_\alpha}{(G')^3}x_i \tag{B.25}$$

$$\nabla_{\beta\delta}\ell_i \;=\; -\frac{1}{\sigma}\frac{G'G'''-2(G'')^2}{(G')^4}G_\delta x_i+\frac{G'G''_\delta-2G''G'_\delta}{(G')^3}x_i \tag{B.26}$$

$$\nabla_{\sigma\sigma}\ell_i \;=\; -\frac{2}{\sigma}\nabla_\sigma\ell_i-\frac{1}{\sigma^2}-\frac{1}{\sigma^2}\frac{G'G'''-2(G'')^2}{(G')^4}(y_i-x_i\beta)^2 \tag{B.27}$$

$$\nabla_{\sigma\alpha}\ell_i \;=\; -\frac{1}{\sigma^2}\frac{G'G'''-2(G'')^2}{(G')^4}G_\alpha(y_i-x_i\beta)+\frac{1}{\sigma^2}\frac{G'G''_\alpha-2G''G'_\alpha}{(G')^3}(y_i-x_i\beta) \tag{B.28}$$

$$\nabla_{\sigma\delta}\ell_i \;=\; -\frac{1}{\sigma^2}\frac{G'G'''-2(G'')^2}{(G')^4}G_\alpha(y_i-x_i\beta)+\frac{1}{\sigma^2}\frac{G'G''_\delta-2G''G'_\delta}{(G')^3}(y_i-x_i\beta) \tag{B.29}$$

$$\nabla_{\alpha\alpha}\ell_i \;=\; -\frac{G'G'''-2(G'')^2}{(G')^4}G_\alpha^2+2\frac{G'G''_\alpha-2G''G'_\alpha}{(G')^3}G_\alpha+\frac{G''G_{\alpha\alpha}-G'G'_{\alpha\alpha}+(G'_\alpha)^2}{(G')^2} \tag{B.30}$$

$$\nabla_{\alpha\delta}\ell_i \;=\; -\frac{G'G'''-2(G'')^2}{(G')^4}G_\alpha G_\delta+\frac{G'G''_\alpha-2G''G'_\alpha}{(G')^3}G_\delta$$

$$+\frac{G'G''_\delta-2G''G'_\delta}{(G')^3}G_\alpha+\frac{G''G_{\alpha\delta}-G'G'_{\alpha\delta}+G'_\alpha G'_\delta}{(G')^2} \tag{B.31}$$

$$\nabla_{\delta\delta}\ell_i \;=\; -\frac{G'G'''-2(G'')^2}{(G')^4}G_\delta^2+2\frac{G'G''_\delta-2G''G'_\delta}{(G')^3}G_\delta+\frac{G''G_{\delta\delta}-G'G'_{\delta\delta}+(G'_\delta)^2}{(G')^2} \tag{B.32}$$

As for Lemma 1, the properties of $G$ and all its derivatives imply continuity of all second

48

order derivatives in $\theta$ except at $\lambda_1, \lambda_2 = 1, 2, 3$.

As for Lemma 3, we replace $(y_i - x_i'\beta)/\sigma$ by $G(u_i)$ in the first and second order derivatives. This yields expressions in $u_i$, multiplied in some cases with $x_i$ or $x_i x_i'$. Thus, relative to $x$, Lemma 3 requires that $\mathrm{E}[x]$ and $\mathrm{E}[xx']$ are finite. Taking expectations with respect to $y_i$ turns into an integration over $u_i$, which has a uniform density, which simplifies the analysis considerably. Thus, we examine whether $\nabla_{\theta\theta}$ is integrable for $u \in [0, 1]$. Since $G$ and its derivatives are well-behaved for any $u \in [u_\eta, 1 - u_\eta]$ for a small $u_\eta$, we must examine the behavior of $\nabla_{\theta\theta}$ over the range $[0, u_\eta]$; the argument for the range $[1 - u_\eta, 1]$ is similar.

In $\nabla_{\theta\theta}$, $G'$ appears in the denominator frequently. Since $G' \to \infty$ for $u \downarrow 0$ when $\lambda_1 < 1$, we consider the cases of $\lambda_1 < 1$, $\lambda_1 = 1$ and $\lambda_1 > 1$ separately. Furthermore, note that as $u \downarrow 0$, terms with $1 - u$ in equations (B.19) to (B.32) contribute at most a constant to the limit of the expression for any finite $\lambda_2$.

**The case of $\lambda_1 < 1$.**

We start with $\nabla_{\beta\beta}\ell_i$, dividing the ratio in equation (B.23) into two parts:

$$
\begin{aligned}
\frac{G'''}{(G')^3} &\to \frac{(\lambda_1 - 1)(\lambda - 2)u^{\lambda_1 - 3} + (\lambda_2 - 1)(\lambda_2 - 2)}{(u^{\lambda_1 - 1} + 1)^3} \\
&= \frac{u^{\lambda_1 - 3}\big((\lambda_1 - 1)(\lambda - 2) + (\lambda_2 - 1)(\lambda_2 - 2)u^{3 - \lambda_1}\big)}{u^{3\lambda_1 - 3}(1 + u^{1 - \lambda_1})^3} \\
&\to \frac{u^{\lambda_1 - 3}(\lambda_1 - 1)(\lambda - 2)}{u^{3\lambda_1 - 3}} = u^{-2\lambda_1}(\lambda_1 - 1)(\lambda - 2) \qquad\text{(B.33)} \\
\frac{-2(G'')^2}{(G')^4} &\to -2\frac{((\lambda_1 - 1)u^{\lambda_1 - 2} - (\lambda_2 - 1))^2}{(u^{\lambda_1 - 1} + 1)^4} \\
&= -2\frac{u^{2\lambda_1 - 4}((\lambda_1 - 1) - (\lambda_2 - 1)u^{2 - \lambda_1})^2}{u^{4\lambda_1 - 4}(1 + u^{1 - \lambda_1})^4} \\
&\to -2\frac{(\lambda_1 - 1)^2 u^{2\lambda_1 - 4}}{u^{4\lambda_1 - 4}} = -2(\lambda_1 - 1)^2 u^{-2\lambda_1} \qquad\text{(B.34)}
\end{aligned}
$$

since $u^{m - \lambda_1} \to 0$ for $m = 1, 2, 3$. Combining these two terms yields

$$\frac{G'G''' - 2(G'')^2}{(G')^4} \to -\lambda_1(\lambda_1 - 1)u^{-2\lambda_1} \tag{B.35}$$

which has a finite integral on $[0, u_\eta]$ if $-2\lambda_1 \geq -1$ or $\lambda_1 \leq 1/2$.[20]

For the other terms, we follow the same strategy. For $\nabla_{\beta\sigma}\ell_i$, we have:

$$\frac{G''}{G'} \quad \to \quad \frac{(\lambda_1 - 1)u^{\lambda_1 - 2} - (\lambda_2 - 1)}{(u^{\lambda_1 - 1} + 1)^2} \to (\lambda_1)u^{-\lambda_1} \tag{B.36}$$

$$\frac{G'G''' - 2(G'')^2}{(G')^4}G \quad \to \quad -\lambda_1(\lambda_1 - 1)u^{-2\lambda_1}\frac{u^{\lambda_1} - 1}{\lambda_1} = -(\lambda_1 - 1)(u^{-\lambda_1} - u^{-2\lambda_1}) \tag{B.37}$$

which, altogether, has a finite integral if $\lambda_1 \leq 1/2$.

For $\nabla_{\beta\alpha}\ell_i$, we have:

$$\frac{G'G''' - 2(G'')^2}{(G')^4}G_\alpha \quad \to \quad -\lambda_1(\lambda_1 - 1)u^{-2\lambda_1}\left(\frac{1}{\lambda_1}u^{\lambda_1}\ln u - \frac{1}{\lambda_1^2}(u^{\lambda_1} - 1)\right)$$

$$= \quad (\lambda_1 - 1)u^{-\lambda_1}\ln u - \frac{\lambda_1 - 1}{\lambda_1}(u^{-\lambda_1} - u^{-2\lambda_1}) \tag{B.38}$$

$$\frac{G'G_\alpha'' - 2G''G_\alpha'}{(G')^3} \quad \to \quad u^{-\lambda_1} - u^{-2\lambda_1 + 2} - (\lambda_1 - 1)u^{-\lambda_1}\ln u \tag{B.39}$$

The first term in (B.38) cancels against the last term of (B.38). $u^{-\lambda_1}$, $u^{-2\lambda_1}$, and $u^{2-2\lambda_1}$ all have a finite integral if $\lambda_1 < 1/2$. The result for $\nabla_{\beta\delta}\ell_i$ is parallel to this case.

For $\nabla_{\sigma\sigma}\ell_i$, we have:

$$\frac{G''}{(G')^2}G \quad \to \quad \left(\frac{(\lambda_1 - 1)u^{-\lambda_1 - 2} - (\lambda_2 - 1)}{(u^{\lambda_1 - 1} + 1)^2}\right)\left(\frac{u^{\lambda_1} - 1}{\lambda_1}\right)$$

$$\to \quad \frac{1}{\lambda_1}\left((\lambda_1 - 1)(1 - u^{-\lambda_1}) - (\lambda_2 - 1)(u^{2-\lambda_1} - u^{2-2\lambda_1})\right) \tag{B.40}$$

---

[20]Lemma 3 requires $\nabla_{\beta\beta}\ell_i(\theta)$ to be bounded in absolute value by a function $b(y, x)$. As equations (B.33) and (B.34), the expression in equation (B.35) omits multiplicative functions that converge to 1 as $u \downarrow 0$. Thus, $b(y, x)$ is found by multiplying (B.35) with the largest value that such functions take over the interval $[0, u_\eta]$, which is finite.

$$\frac{G'G''' - 2(G'')^2}{(G')^4}G^2 \; \rightarrow \; -\lambda_1(\lambda_1 - 1)u^{-2\lambda_1}\Big(\frac{u^{\lambda_1} - 1}{\lambda_1}\Big)^2$$

$$= \; -\frac{(\lambda_1 - 1)}{\lambda_1}(1 - 2u^{-\lambda_1} + u^{-2\lambda_1}) \tag{B.41}$$

In (B.40) and (B.41), $u^{-\lambda_1}$, $u^{2-\lambda_1}$, $u^{2-2\lambda_1}$ and $u^{-2\lambda_1}$ all have a finite integral if $\lambda_1 < 1/2$.

For $\nabla_{\sigma\alpha}\ell_i$, we have with the aid of (B.35) and (B.39):

$$\nabla_{\sigma\alpha}\ell_i \rightarrow \frac{1}{\lambda_1^2}\big(1 - \lambda_1(u^{2-\lambda_1} - u^{2-2\lambda_1}) + (\lambda_1 - 2)u^{-\lambda_1} - (\lambda_1 - 1)u^{-2\lambda_1}\big) \tag{B.42}$$

Once again, finite integrability depends on $u^{-2\lambda_1}$, such that $\lambda_1$ must be less than $1/2$. The result for $\nabla_{\sigma\delta}\ell_i$ is parallel to this case.

With respect to $\nabla_{\alpha\alpha}\ell_i$, we have three components:

$$-\frac{G'G''' - 2(G'')^2}{(G')^4}G_\alpha^2 \; \rightarrow \; \frac{\lambda_1 - 1}{\lambda_1}(\ln u)^2 - \frac{2(\lambda_1 - 1)}{\lambda_1^2}\ln u + \frac{2(\lambda_1 - 1)}{\lambda_1^2}u^{-\lambda_1}\ln u$$

$$+ \frac{\lambda_1 - 1}{\lambda_1^3}(1 - 2u^{-\lambda_1} + u^{-2\lambda_1}) \tag{B.43}$$

$$2\frac{G'G''_\alpha - 2G''G'_\alpha}{(G')^3}G_\alpha \; \rightarrow \; -2\frac{\lambda_1 - 1}{\lambda_1}(\ln u)^2 + \frac{2(\lambda_1 - 1)}{\lambda_1^2}\ln u - \frac{2(\lambda_1 - 1)}{\lambda_1^2}u^{-\lambda_1}\ln u$$

$$+ \frac{2}{\lambda_1^2}\big(\lambda_1\ln u - \lambda_1 u^{2-\lambda_1}\ln u\big)$$

$$+ \frac{2}{\lambda_1^2}\big(u^{-\lambda_1} - 1 + u^{2-\lambda_1} - u^{2-2\lambda_1}\big) \tag{B.44}$$

$$\frac{G''G_{\alpha\alpha} - G'G'_{\alpha\alpha} + (G'_\alpha)^2}{(G')^2} \; \rightarrow \; \frac{\lambda_1 - 1}{\lambda_1}(\ln u)^2 - 2\frac{\lambda_1 - 1}{\lambda_1^2}\ln u + 2\frac{\lambda_1 - 1}{\lambda_1^3}(1 - u^{-\lambda_1}) \tag{B.45}$$

The first line of (B.43) and (B.44) and the first term of (B.45) cancel against each other. On the second line of (B.44), $u^{2-\lambda_1}\ln u$ vanishes more rapidly as $u \downarrow 0$ than $\ln u$, which itself has a finite integral. The remainder has a finite integral if $\lambda_1 \leq 1/2$. The results for $\nabla_{\alpha\delta}\ell_i$ and $\nabla_{\delta\delta}\ell_i$ are parallel to this case.

**The case of $\lambda_1 \geq 1$.**

From Lemma 1, it is clear that $\nabla_{\theta\theta}\ell_i$ has discontinuities at $\lambda_1 = 1, 2, 3$. With methods similar to above, we have in regard to $\nabla_{\beta\beta}\ell_i$, as $u \downarrow 0$:

$$1 < \lambda_1 < 2 \quad : \quad \frac{G'G''' - 2(G'')^2}{(G')^4} \rightarrow (\lambda_1 - 1)(\lambda_1 - 2)u^{\lambda_1 - 3} - 2(\lambda_1 - 1)^2 u^{2\lambda_1 - 4} \quad \text{(B.46)}$$

$$2 < \lambda_1 < 3 \quad : \quad \frac{G'G''' - 2(G'')^2}{(G')^4} \rightarrow (\lambda_1 - 1)(\lambda_1 - 2)u^{\lambda_1 - 3} - 2(\lambda_2 - 1)^2 \quad \text{(B.47)}$$

$$3 < \lambda_1 < \infty \quad : \quad \frac{G'G''' - 2(G'')^2}{(G')^4} \rightarrow -\lambda_2(\lambda_2 - 1) \quad \text{(B.48)}$$

In (B.47), the expression has a finite integral if $\lambda_1 - 3 > -1$ or $\lambda_1 > 2$, which is compatible with the range of $\lambda_1$ considered in the derivation of (B.47). In (B.46) however, for the expression to have a finite integral over $[0, u_\eta]$, $\lambda_1$ must exceed 2, which is beyond the range considered. With similar derivations, it can be shown that every component of $\nabla_{\theta\theta}\ell_i$ has a finite integral only if $\lambda_1 > 2$ and $\lambda_2 > 2$ with a discontinuity at $\lambda_1 = \lambda_2 = 3$.

## C    Proof of Lemma 2

For $\lambda_1 > 0$, the lower bound of $\epsilon$ is $\underline{\epsilon} = \frac{-1}{\lambda_1}$, and for $\lambda_2 > 0$, the upper bound of $\epsilon$ is $\bar{\epsilon} = \frac{1}{\lambda_2}$. Thus the bounds on $y$ are $\underline{y} = x'\beta + \sigma\underline{\epsilon}$ and $\bar{y} = x'\beta + \sigma\bar{\epsilon}$, respectively. To prove Lemma 2, we abbreviate notation slightly and examine

$$\nabla_\theta \left( \int_{\underline{y}}^{\bar{y}} g_y(y|\theta) dy \right) = \int_{\underline{y}}^{\bar{y}} \nabla_\theta g_y(y|\theta) - g_y(\underline{y}|\theta)\underline{y}_\theta + g_y(\bar{y}|\theta)\bar{y}_\theta. \quad \text{(C.1)}$$

Thus, interchanging differentiation and integration is permissible as long as $g_y(\underline{y}|\theta) = g_y(\bar{y}|\theta) = 0$. We have $g_y(y|\theta) = [G'(u)]^{-1}/\sigma$ with $u = G^{-1}(\epsilon)$ with $\epsilon = \frac{1}{\sigma}(y - x'\beta)$. Therefore, as $y \rightarrow \underline{y}$, we have $\epsilon \rightarrow \underline{\epsilon}$ and $u \rightarrow 0$, and thus for $\lambda_1 < 1$, $G'(u) \rightarrow \infty$ and $g_y(y|\theta) \rightarrow 0$. Similarly, for $\lambda_2 < 1$, $g_y(\bar{y}|\theta) \rightarrow 0$.

Since $\int s(\theta)g_y(y|\theta)dy = \int \nabla_\theta g_y(y|\theta)dy$, part $(ii)$ of Lemma 2 concerns the derivative of

the first term in equation (C.1):

$$\nabla_\theta \Big( \int_{\underline{y}}^{\bar{y}} \nabla_\theta g_y(y|\theta) dy \Big) = \int_{\underline{y}}^{\bar{y}} \nabla_{\theta\theta} g_y(y|\theta) - \nabla_\theta g_y(\underline{y}|\theta)\underline{y}_\theta + \nabla_\theta g_y(\bar{y}|\theta)\bar{y}_\theta. \tag{C.2}$$

$\nabla_\theta g_y(y|\theta)$ approaches the following functions as $u \downarrow 1$:

$$
\begin{aligned}
\nabla_\beta g_y &= \frac{G''(u)x}{\sigma^2(G'(u))^3} \to \big( (\lambda_1 - 1)u^{1-2\lambda_1} - (\lambda_2 - 1)u^{3-3\lambda_1} \big) && \text{(C.3)} \\
\nabla_\sigma g_y &= -\frac{1}{\sigma^2 G'(u)} - \frac{G''(u)(y - x'\beta)}{\sigma^3(G'(u))^3} \\
&\to -\sigma^{-2}\big( u^{1-\lambda_1} - \lambda_1^{-1}((\lambda_1 - 1)u^{1-2\lambda_1} - (\lambda_2 - 1)u^{3-3\lambda_1}) \big) && \text{(C.4)} \\
\nabla_\alpha g_y &= \frac{G''(u)G_\alpha}{\sigma(G'(u))^3} \\
&\to \frac{\lambda_1 - 1}{\lambda_1}u^{1-\lambda_1}\ln u - \frac{\lambda_2 - 1}{\lambda_1}u^{3-2\lambda_1}\ln u \\
&\quad -\frac{\lambda_1 - 1}{\lambda_1^2}(u^{1-\lambda_1} - u^{1-2\lambda_1}) + \frac{\lambda_2 - 1}{\lambda_1^2}(u^{3-2\lambda_1} - u^{3-3\lambda_1}) && \text{(C.5)} \\
\nabla_\delta g_y &= \frac{G''(u)G_\delta}{\sigma(G'(u))^3} \\
&\to -\frac{\lambda_1 - 1}{\lambda_1}u^{1-\lambda_1}\ln u + \frac{\lambda_2 - 1}{\lambda_1}u^{3-2\lambda_1}\ln u \\
&\quad +\frac{\lambda_1 - 1}{\lambda_1^2}(u^{1-\lambda_1} - u^{1-2\lambda_1}) - \frac{\lambda_2 - 1}{\lambda_1^2}(u^{3-2\lambda_1} - u^{3-3\lambda_1}) && \text{(C.6)}
\end{aligned}
$$

As $u \downarrow 1$, each of these functions goes to 0 if and only if $\lambda < 1/2$. By a similar argument, $\lambda_2 < 1/2$ is necessary for $\nabla_\theta g_y(y|\theta)$ to go to 0 as $u \to 1$.[21]

## D  Proof of Lemma 4

$E[H(\theta)]$ may be written as integrals of expressions found in equations (B.23)-(B.32) (multiplied by $g_x(x)$) over $u$ and $x$. As functions of $u$, integration of equations (B.23)-(B.32) has

---

[21]The derivative of the second and third terms of (C.1) equals 0 as well when $\lambda_1 < 1/2$ and $\lambda_2 < 1/2$. That is, for both the first and second order derivatives of $\int g_y(y|\theta)dy$, differentiation and integration is interchangeable.

no analytical solution. Thus, to provide evidence that $A_0 = -E[H(\theta)]$ is positive definite, we resort to numerical integration by simulation. This evidence is tied to a model that must therefore be specified in a generic fashion.

We specify $x$ as a vector of three elements: $x_1 = 1$ to allow for an intercept, $x_2$ is standard normal, and $x_3$ is a standardized $\chi^2(5)$ variable that introduces some skewness into the explanatory variables. $\beta$ is set to $(1, 1, 1)'$. Note that $x'\beta$ generates a location shift only. $\lambda_1$ and $\lambda_2$ vary from $-3$ to $0.49$, the former out of concern for numerical over- and underflow in the computations and the latter in deference to the upper bound of the feasible parameter space. $\sigma$ cannot really be held constant because variations in $\alpha$ and $\delta$ generate $\epsilon$s of a different scale such that models with a high $\lambda_1$ and $\lambda_2$ have a much higher signal-to-noise ratio. Since $x'\beta$ has a variance of 2 by design, we choose $\sigma = (2IQR_{N(0,1)}/IQR_{GTL(\alpha,\delta)})^{0.5}$. Thus, if the GTL density is close to the standard normal density, the variance of $\epsilon$ is close to 2 and the signal-to-noise ratio is about 1. $\sigma$ declines as $\lambda_1$ and $\lambda_2$ fall and the interquartile range of $\epsilon$ rises.[22] For other values of $\lambda_1$ and $\lambda_2$, the signal-to-noise ratio probably differs from 1 but should be in the neighborhood of it.

For the given range of $\lambda_1$ and $\lambda_2$, Table D.1 reports the smallest eigenvalues of the simulated matrix $\hat{A}_0$, which itself is computed on the basis of 10,000 replications. The smallest eigenvalue is always solidly positive: at least for the model that is examined here, $A_0$ appears to be positive definite.

Table D.1: Simulated values of the smallest eigenvalue of $\hat{A}_0$

| | $\lambda_2$ for $N = 100$ | | | | | $\lambda_2$ for $N = 1000$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\lambda_1$ | -3 | -2 | -1 | 0 | 0.49 | -3 | -2 | -1 | 0 | 0.49 |
| -3 | 6.56 | 8.01 | 9.46 | 11.14 | 11.93 | 65.29 | 79.81 | 94.12 | 110.77 | 117.87 |
| -2 | 8.04 | 10.69 | 13.10 | 16.27 | 17.91 | 79.82 | 106.30 | 130.31 | 161.48 | 175.66 |
| -1 | 9.48 | 13.13 | 15.17 | 17.16 | 18.07 | 94.15 | 130.36 | 152.74 | 170.78 | 178.06 |
| 0 | 11.16 | 16.30 | 17.17 | 24.22 | 23.79 | 110.84 | 161.62 | 170.99 | 242.28 | 234.88 |
| 0.49 | 11.94 | 17.92 | 18.09 | 23.69 | 37.17 | 118.08 | 176.12 | 178.49 | 235.53 | 352.26 |

---

[22]Recall that $\text{Var}\,\epsilon$ does not exist if $\min(\lambda_1, \lambda_2)$ is less than $-1/2$.

# E   Monte Carlo Results: Diagnostics of small-sample distributions

Tables E.1 - E.6 present diagnostic results of the small-sample distributions of the OLS and MLE estimators of the various experiments that are discussed in Section 4 of the paper. Some remarks are in order.

Theoretically, since the GTL regression estimator is a maximum likelihood estimator, the variance (denoted as $V_h$) of the estimator may be calculated as the inverse of the negative Hessian. However, when a penalty function[23] is added to the log-likelihood function in order to keep the estimator of $(\alpha, \delta)$ in bounds, $V_h$ is impacted by the curvature in this penalty function and thus may well be biased. More properly, the log-likelihood-with-penalty function may be seen as a criterion function within the context of quasi-likelihood estimation (White, 1982; Gourieroux et al., 1984), in which case the familiar sandwich estimator (denoted as $V_{sw}$) ought to be used to compute the variance. In our set of experiments, we added a penalty function only for the experiments with $\alpha = 0.33$ and $\delta = -0.10, 0.00, 0.10$. For other experiments, the sandwich estimator should not be needed. However, as a check, we compute the sandwich estimator also when $\alpha = -0.33$.

Table E.3 shows that for $n = 5000$ and $\alpha = 0.33$, $V_{sw}$ yields an estimated variance that is much closer to the variance of the Monte Carlo draws than $V_h$. But its use is questionable when $\alpha = -0.33$: the variance ratio moves away from 1 for four of the six parameters, increasing it for $\hat{\beta}_1$ and $\hat{\beta}_2$ and decreasing it for $\hat{\alpha}$ and $\hat{\delta}$. For small samples ($n = 250$), when $\alpha = 0.33$, $V_h$ tends to be too large; $V_{sw}$ tends to overcorrect and ends up to be often too small

---

[23] The penalty function takes the following form. Let $\lambda_1 = \alpha - \delta$ and $\lambda_2 = \alpha + \delta$. Let the lower and upper limit on $\lambda_j$ be denoted as $\lambda_{jL}$ and $\lambda_{jU}$, respectively; in our Monte Carlo analysis, we set $\lambda_{jL} = -3$ and $\lambda_{jU} = 0.5$. Let $\lambda_{jR} = (\lambda_{jU} - \lambda_{jL})/2$ measure half of the feasible range of $\lambda_j$. As always, $n$ denotes the number of observations. Then the penalty function is written as $P(\lambda_1, \lambda_2, n) = 0.005\ n\ \big(p(\lambda_1) + p(\lambda_2)\big)$

$$p(\lambda_j) = \ln\left\{(\lambda_j - \lambda_{jL})/\lambda_{jR}\right\} + \ln\left\{(\lambda_{jL} - \lambda_j)/\lambda_{jR}\right\}$$

This function $p$ takes on a value of 0 at the midpoint $-1.25$ of the feasible range of $\lambda_j$, is symmetric around this midpoint, and for $\lambda_j = 0.20, 0.40, 0.45, 0.49, 0.499$ equals $-0.00580, -0.01099, -0.01438, -0.02237, -0.03387$ respectively. It has an asymptote of $-\infty$ at the endpoints of the range.

but is usually closer to the Monte Carlo variance. For $\alpha = -0.33$, the sandwich estimator yields a better value of the variance for $\hat{\beta}_2$ but worsens the variance for $\hat{\beta}_3$, $\hat{\beta}_1$, $\hat{\alpha}$, and $\hat{\delta}$.

As for coverage ratios (the proportion of confidence intervals that include the true population parameter), Table E.2 considers a significance level of 0.05. Ideally, therefore, the coverage ratios should equal 0.95. Use of the sandwich estimator when $\alpha = -0.33$ makes little difference when $n = 5000$ and worsens coverage when $n = 250$. When $\alpha = 0.33$, the coverage ratio improves for $\hat{\beta}_2$ and $\hat{\beta}_3$; the other parameters suffer from bias especially when $\delta \neq 0$ and thus exhibit poor coverage performance.[24]

Based on these comparisons, the use of a sandwich estimator is recommended when a penalty function is added to the log-likelihood function but may do more harm than good when the regression model is estimated without a penalty function.

Whereas the determinants in the baseline design are well-behaved variables ($N(0,1)$ and $\chi^2(5)$), the second design specifies the determinants to be thick-tailed as well. Specifically, $X_j$ is distributed $GTL(\alpha_{xj}, \delta_{xj})$ for $j = 2, 3$ with $\alpha_{xj} = -0.67$ and $\delta_{xj} = -0.25, 0.00, 0.25$. In this design, if $X$ denotes the matrix of explanatory variables including a column of ones for the intercept, $plim\ X'X/n$ is not defined. It is well-known that the OLS estimator is no longer consistent; the GTL-regression estimator is likely similarly impacted. Tables E.4 to E.6 illustrate the relative performance of the OLS and GTL-regression estimators.

Table E.4 illustrates that normality of the OLS slope estimators is destroyed by thick tails in the disturbances or explanatory variables. The GTL-regression estimator of the slopes displays normality for large $n$ in the estimators of $\alpha$, $\delta$, $\sigma$, and $\beta_1$. If $X_2$ is well-behaved and $X_3$ is thick-tailed, $\hat{\beta}_2$ is normally distributed but $\hat{\beta}_3$ is not. Table E.6 shows that the inverted negative hessian underestimates the Monte Carlo variance of the GTL-regression estimator

---

[24]The penalty function keeps $(\hat{\alpha}, \hat{\delta})$ within the feasible area but also has the effect of driving it away from the boundary of the feasible area. This produces a small but significant bias in $(\hat{\alpha}, \hat{\delta})$ in experiments with $\alpha = 0.33$. Experimentation with different penalty function specifications indicated that functions that become active only near the boundary (and thus are less likely to cause bias) typically cause non-convergence during some replications.

Table E.1: *p*-values of Jarque-Bera tests for normality of OLS and GTL estimators: baseline design

| DGP | | | OLS | | | GTL | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\delta$ | $\sigma$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\sigma$ | $\alpha$ | $\delta$ |
| A: GTL as an approximation of the standard normal distribution, $N = 250$ | | | | | | | | | | | |
| 0.1436 | 0.00 | 1.188 | 0.91 | 0.10 | 0.62 | 0.46 | 0.22 | 0.55 | 0.00 | 0.00 | 0.01 |
| B: Various GTL distributions, small sample, $N = 250$ | | | | | | | | | | | |
| 0.33 | -0.10 | 1.477 | 0.80 | 0.07 | 0.48 | 0.36 | 0.07 | 0.30 | 0.29 | 0.00 | 0.03 |
| 0.33 | 0.00 | 1.508 | 0.90 | 0.11 | 0.51 | 0.47 | 0.51 | 0.16 | 0.08 | 0.00 | 0.43 |
| 0.33 | 0.10 | 1.477 | 0.93 | 0.17 | 0.51 | 0.17 | 0.46 | 0.14 | 0.31 | 0.00 | 0.10 |
| -0.33 | -0.10 | 0.454 | 0.00 | 0.00 | 0.00 | 0.23 | 0.21 | 0.86 | 0.00 | 0.10 | 0.73 |
| -0.33 | 0.00 | 0.482 | 0.00 | 0.00 | 0.02 | 0.11 | 0.10 | 0.72 | 0.00 | 0.12 | 0.59 |
| -0.33 | 0.10 | 0.454 | 0.00 | 0.00 | 0.00 | 0.07 | 0.07 | 0.54 | 0.00 | 0.18 | 0.29 |
| -0.67 | -0.25 | 0.136 | 0.00 | 0.00 | 0.00 | 0.07 | 0.70 | 0.80 | 0.00 | 0.06 | 0.55 |
| -0.67 | 0.00 | 0.202 | 0.00 | 0.00 | 0.00 | 0.00 | 0.43 | 0.78 | 0.00 | 0.04 | 0.63 |
| -0.67 | 0.25 | 0.136 | 0.00 | 0.00 | 0.00 | 0.04 | 0.51 | 0.41 | 0.00 | 0.08 | 0.21 |
| -1.00 | -0.50 | 0.024 | 0.00 | 0.00 | 0.00 | 0.53 | 0.93 | 0.64 | 0.00 | 0.09 | 0.33 |
| -1.00 | 0.00 | 0.079 | 0.00 | 0.00 | 0.00 | 0.00 | 0.57 | 0.81 | 0.00 | 0.02 | 0.57 |
| -1.00 | 0.50 | 0.024 | 0.00 | 0.00 | 0.00 | 0.40 | 0.53 | 0.23 | 0.00 | 0.04 | 0.10 |
| C: Various GTL distributions, large sample, $N = 5000$ | | | | | | | | | | | |
| 0.33 | -0.10 | 1.477 | 0.09 | 0.99 | 0.07 | 0.68 | 0.23 | 0.78 | 0.71 | 0.78 | 0.09 |
| 0.33 | 0.00 | 1.508 | 0.10 | 0.96 | 0.16 | 0.65 | 0.55 | 0.70 | 0.48 | 0.69 | 0.10 |
| 0.33 | 0.10 | 1.477 | 0.14 | 0.89 | 0.33 | 0.87 | 0.67 | 0.61 | 0.59 | 0.56 | 0.26 |
| -0.33 | -0.10 | 0.454 | 0.00 | 0.00 | 0.00 | 0.51 | 0.50 | 0.99 | 0.30 | 0.44 | 0.06 |
| -0.33 | 0.00 | 0.482 | 0.83 | 0.07 | 0.21 | 0.35 | 0.39 | 0.98 | 0.23 | 0.44 | 0.07 |
| -0.33 | 0.10 | 0.454 | 0.00 | 0.00 | 0.00 | 0.19 | 0.41 | 0.97 | 0.15 | 0.42 | 0.08 |
| -0.67 | -0.25 | 0.136 | 0.00 | 0.00 | 0.00 | 1.00 | 0.39 | 0.98 | 0.71 | 0.42 | 0.11 |
| -0.67 | 0.00 | 0.202 | 0.00 | 0.00 | 0.00 | 0.91 | 0.24 | 0.87 | 0.29 | 0.61 | 0.20 |
| -0.67 | 0.25 | 0.136 | 0.00 | 0.00 | 0.00 | 0.40 | 0.40 | 0.73 | 0.13 | 0.50 | 0.30 |
| -1.00 | -0.50 | 0.024 | 0.00 | 0.00 | 0.00 | 0.75 | 0.49 | 0.93 | 0.98 | 0.42 | 0.07 |
| -1.00 | 0.00 | 0.079 | 0.00 | 0.00 | 0.00 | 0.83 | 0.21 | 0.76 | 0.30 | 0.86 | 0.30 |
| -1.00 | 0.50 | 0.024 | 0.00 | 0.00 | 0.00 | 0.18 | 0.45 | 0.47 | 0.04 | 0.74 | 0.53 |

Note: These test results pertain to simulations reported in Tables 1 and 2.

if determinants are thick-tailed. However, the measured deviation is not as large as it is for the OLS estimator. (Note that this "deviation" cannot be called a "bias" since the mean of the variance estimator is not defined.) Table E.5 shows that these poor estimates impact the coverage ratio as well: for OLS, unlike in the baseline design, thick-tailed determinants lead to a excessive coverage ratio (too large confidence intervals), and for the GTL-regression estimator, the coverage ratios are too low (too small confidence intervals).

Table E.2: Coverage rates of OLS and GTL estimators under a 95% confidence level: baseline design

| DGP | | | OLS | | | GTL | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\delta$ | $\sigma$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\sigma$ | $\alpha$ | $\delta$ |
| A: GTL as an approximation of the standard normal distribution, $N = 250$ | | | | | | | | | | | |
| 0.1436 | 0.00 | 1.188 | 0.947 | 0.948 | 0.954 | 0.944 | 0.937 | 0.943 | 0.941 | 0.918 | 0.924 |
| B: Various GTL distributions, small sample, $N = 250$ | | | | | | | | | | | |
| 0.33 | -0.10 | 1.477 | 0.948 | 0.951 | 0.658 | 0.959 | 0.947 | 0.942 | 0.905 | 0.970 | 0.965 |
| | | | | | | *0.935* | *0.915* | *0.949* | *0.853* | *0.899* | *0.913* |
| 0.33 | 0.00 | 1.508 | 0.943 | 0.950 | 0.951 | 0.951 | 0.944 | 0.944 | 0.935 | 0.976 | 0.964 |
| | | | | | | *0.938* | *0.912* | *0.948* | *0.909* | *0.909* | *0.909* |
| 0.33 | 0.10 | 1.477 | 0.944 | 0.950 | 0.651 | 0.959 | 0.948 | 0.946 | 0.908 | 0.976 | 0.963 |
| | | | | | | *0.941* | *0.914* | *0.948* | *0.862* | *0.893* | *0.915* |
| -0.33 | -0.10 | 0.454 | 0.958 | 0.948 | 0.655 | 0.952 | 0.936 | 0.951 | 0.943 | 0.946 | 0.950 |
| -0.33 | 0.00 | 0.482 | 0.952 | 0.946 | 0.951 | 0.956 | 0.944 | 0.946 | 0.944 | 0.945 | 0.947 |
| -0.33 | 0.10 | 0.454 | 0.956 | 0.945 | 0.632 | 0.950 | 0.939 | 0.948 | 0.939 | 0.936 | 0.936 |
| -0.67 | -0.25 | 0.136 | 0.955 | 0.959 | 0.337 | 0.955 | 0.933 | 0.955 | 0.951 | 0.948 | 0.943 |
| -0.67 | 0.00 | 0.202 | 0.954 | 0.947 | 0.962 | 0.955 | 0.936 | 0.955 | 0.944 | 0.955 | 0.937 |
| -0.67 | 0.25 | 0.136 | 0.952 | 0.946 | 0.326 | 0.950 | 0.935 | 0.952 | 0.938 | 0.943 | 0.935 |
| -1.00 | -0.50 | 0.024 | 0.951 | 0.957 | 0.571 | 0.949 | 0.926 | 0.953 | 0.951 | 0.955 | 0.940 |
| -1.00 | 0.00 | 0.079 | 0.951 | 0.946 | 0.972 | 0.954 | 0.929 | 0.956 | 0.940 | 0.949 | 0.934 |
| -1.00 | 0.50 | 0.024 | 0.947 | 0.956 | 0.558 | 0.949 | 0.939 | 0.951 | 0.939 | 0.944 | 0.932 |
| C: Various GTL distributions, large sample, $N = 5000$ | | | | | | | | | | | |
| 0.33 | -0.10 | 1.477 | 0.964 | 0.950 | 0.000 | 0.966 | 0.960 | 0.913 | 0.052 | 0.024 | 0.932 |
| | | | | | | *0.955* | *0.948* | *0.913* | *0.026* | *0.007* | *0.868* |
| 0.33 | 0.00 | 1.508 | 0.963 | 0.949 | 0.938 | 0.961 | 0.950 | 0.942 | 0.290 | 0.285 | 0.977 |
| | | | | | | *0.955* | *0.941* | *0.942* | *0.214* | *0.142* | *0.950* |
| 0.33 | 0.10 | 1.477 | 0.963 | 0.951 | 0.000 | 0.965 | 0.954 | 0.910 | 0.065 | 0.032 | 0.946 |
| | | | | | | *0.953* | *0.936* | *0.911* | *0.027* | *0.007* | *0.889* |
| -0.33 | -0.10 | 0.454 | 0.956 | 0.944 | 0.000 | 0.960 | 0.956 | 0.951 | 0.941 | 0.948 | 0.944 |
| -0.33 | 0.00 | 0.482 | 0.959 | 0.944 | 0.944 | 0.954 | 0.951 | 0.948 | 0.942 | 0.942 | 0.943 |
| -0.33 | 0.10 | 0.454 | 0.957 | 0.947 | 0.001 | 0.956 | 0.952 | 0.947 | 0.941 | 0.940 | 0.945 |
| -0.67 | -0.25 | 0.136 | 0.952 | 0.961 | 0.093 | 0.953 | 0.956 | 0.949 | 0.938 | 0.949 | 0.944 |
| -0.67 | 0.00 | 0.202 | 0.959 | 0.952 | 0.967 | 0.956 | 0.952 | 0.946 | 0.942 | 0.948 | 0.941 |
| -0.67 | 0.25 | 0.136 | 0.954 | 0.954 | 0.100 | 0.947 | 0.956 | 0.944 | 0.949 | 0.937 | 0.944 |
| -1.00 | -0.50 | 0.024 | 0.951 | 0.954 | 0.486 | 0.952 | 0.954 | 0.949 | 0.939 | 0.952 | 0.948 |
| -1.00 | 0.00 | 0.079 | 0.950 | 0.964 | 0.978 | 0.953 | 0.958 | 0.948 | 0.947 | 0.949 | 0.943 |
| -1.00 | 0.50 | 0.024 | 0.958 | 0.956 | 0.481 | 0.953 | 0.943 | 0.945 | 0.947 | 0.934 | 0.939 |

Notes: These test results pertain to simulations reported in Tables 1 and 2. For single-row sets of results, the coverage ratio is based on a variance that is computed from the information matrix (inverse negative Hessian). For double-row sets of results, the first row uses a variance that is computed from the information matrix; the second row (in italics) uses a sandwich estimator to compute the variance.

Table E.3: Ratio of average estimated variance to Monte Carlo variance: baseline design

| DGP | | | OLS | | | GTL | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\delta$ | $\sigma$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\sigma$ | $\alpha$ | $\delta$ |
| A: GTL as an approximation of the standard normal distribution, $N = 250$ | | | | | | | | | | | |
| 0.1436 | 0.00 | 1.188 | 1.045 | 1.000 | 0.979 | 1.010 | 0.961 | 0.988 | 0.905 | 0.878 | 0.856 |
| B: Various GTL distributions, small sample, $N = 250$ | | | | | | | | | | | |
| 0.33 | -0.10 | 1.477 | 1.033 | 0.995 | 0.972 | 1.131 | 1.028 | 0.959 | 1.280 | 1.606 | 1.226 |
|  |  |  |  |  |  | *0.932* | *0.827* | *0.994* | *0.974* | *0.865* | *0.812* |
| 0.33 | 0.00 | 1.508 | 1.037 | 0.994 | 0.984 | 1.064 | 1.006 | 0.979 | 1.255 | 1.460 | 1.157 |
|  |  |  |  |  |  | *0.932* | *0.858* | *1.001* | *1.017* | *0.869* | *0.765* |
| 0.33 | 0.10 | 1.477 | 1.042 | 0.995 | 0.996 | 1.080 | 1.055 | 0.960 | 1.392 | 1.691 | 1.167 |
|  |  |  |  |  |  | *0.888* | *0.851* | *0.994* | *1.064* | *0.916* | *0.773* |
| -0.33 | -0.10 | 0.454 | 1.073 | 0.926 | 0.915 | 1.018 | 0.956 | 1.031 | 0.970 | 0.992 | 0.929 |
| -0.33 | 0.00 | 0.482 | 1.073 | 1.004 | 0.960 | 1.005 | 0.945 | 1.024 | 0.966 | 0.996 | 0.929 |
| -0.33 | 0.10 | 0.454 | 1.037 | 1.071 | 1.018 | 0.989 | 0.933 | 1.016 | 0.960 | 0.994 | 0.928 |
| -0.67 | -0.25 | 0.136 | 0.875 | 1.010 | 0.866 | 1.005 | 0.945 | 1.051 | 1.007 | 1.019 | 0.933 |
| -0.67 | 0.00 | 0.202 | 0.960 | 0.994 | 0.931 | 0.997 | 0.922 | 1.045 | 0.994 | 1.026 | 0.937 |
| -0.67 | 0.25 | 0.136 | 0.794 | 1.368 | 1.042 | 0.987 | 0.912 | 1.032 | 0.977 | 1.020 | 0.946 |
| -1.00 | -0.50 | 0.024 | 0.700 | 1.235 | 0.844 | 0.991 | 0.940 | 1.052 | 1.039 | 1.019 | 0.933 |
| -1.00 | 0.00 | 0.079 | 0.830 | 1.132 | 0.893 | 0.978 | 0.896 | 1.054 | 1.008 | 1.032 | 0.939 |
| -1.00 | 0.50 | 0.024 | 0.777 | 1.707 | 0.996 | 0.991 | 0.903 | 1.033 | 0.993 | 1.030 | 0.961 |
| C: Various GTL distributions, large sample, $N = 5000$ | | | | | | | | | | | |
| 0.33 | -0.10 | 1.477 | 1.053 | 1.006 | 0.952 | 1.177 | 1.137 | 0.958 | 1.342 | 1.609 | 1.338 |
|  |  |  |  |  |  | *1.031* | *0.992* | *0.967* | *0.992* | *0.931* | *0.974* |
| 0.33 | 0.00 | 1.508 | 1.051 | 1.008 | 0.942 | 1.111 | 1.058 | 0.965 | 1.201 | 1.420 | 1.346 |
|  |  |  |  |  |  | *1.038* | *0.986* | *0.963* | *0.954* | *0.913* | *0.989* |
| 0.33 | 0.10 | 1.477 | 1.049 | 1.010 | 0.931 | 1.177 | 1.110 | 0.957 | 1.262 | 1.583 | 1.415 |
|  |  |  |  |  |  | *1.031* | *0.970* | *0.965* | *0.932* | *0.914* | *1.029* |
| -0.33 | -0.10 | 0.454 | 1.050 | 1.046 | 0.901 | 1.027 | 1.036 | 0.992 | 0.912 | 0.908 | 0.951 |
| -0.33 | 0.00 | 0.482 | 1.035 | 1.006 | 0.897 | 1.034 | 1.038 | 0.997 | 0.907 | 0.905 | 0.949 |
| -0.33 | 0.10 | 0.454 | 1.076 | 0.999 | 0.921 | 1.042 | 1.035 | 1.003 | 0.907 | 0.906 | 0.948 |
| -0.67 | -0.25 | 0.136 | 1.839 | 1.897 | 0.998 | 1.006 | 1.058 | 0.996 | 0.918 | 0.932 | 0.940 |
| -0.67 | 0.00 | 0.202 | 1.307 | 1.446 | 0.980 | 1.019 | 1.066 | 1.005 | 0.908 | 0.921 | 0.931 |
| -0.67 | 0.25 | 0.136 | 1.529 | 2.627 | 1.020 | 1.035 | 1.034 | 1.011 | 0.920 | 0.918 | 0.925 |
| -1.00 | -0.50 | 0.024 | 1.676 | 0.981 | 1.011 | 0.995 | 1.046 | 1.002 | 0.948 | 0.967 | 0.945 |
| -1.00 | 0.00 | 0.079 | 2.727 | 0.966 | 0.998 | 1.008 | 1.080 | 1.017 | 0.923 | 0.943 | 0.924 |
| -1.00 | 0.50 | 0.024 | 20.456 | 0.382 | 1.005 | 1.024 | 0.998 | 1.014 | 0.944 | 0.922 | 0.909 |

Notes: These test results pertain to simulations reported in Tables 1 and 2. For single-row sets of results, the numerator of the ratio is based on a variance that is computed from the information matrix (inverse negative Hessian). For double-row sets of results, the first row uses a variance that is computed from the information matrix; the second row (in italics) uses a sandwich estimator to compute the variance in the numerator.

Table E.4: $p$-values of Jarque-Bera tests for normality of OLS and GTL estimators: design with thick-tailed determinants

| DGP for $x_1$ | | DGP for $x_2$ | | OLS | | | MLE | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha_{x2}$ | $\delta_{x2}$ | $\alpha_{x3}$ | $\delta_{x3}$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\sigma$ | $\alpha$ | $\delta$ |
| A: GTL-Disturbances are generated with $(\alpha,\delta)=(-0.67,0.25)$ for $n=250$ | | | | | | | | | | | | |
| 0.1436 | 0.00 | 0.1436 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.62 | 0.43 | 0.00 | 0.11 | 0.21 |
| 0.1436 | 0.00 | -0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.02 | 0.32 | 0.00 | 0.15 | 0.15 |
| 0.1436 | 0.00 | -0.33 | 0.10 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.39 | 0.00 | 0.16 | 0.19 |
| 0.1436 | 0.00 | -0.67 | 0.00 | 0.00 | 0.00 | 0.00 | 0.09 | 0.00 | 0.39 | 0.00 | 0.12 | 0.12 |
| 0.1436 | 0.00 | -0.67 | 0.25 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.48 | 0.00 | 0.13 | 0.20 |
| -0.67 | 0.00 | -0.67 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.54 | 0.00 | 0.07 | 0.12 |
| -0.67 | 0.25 | -0.67 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.55 | 0.00 | 0.09 | 0.24 |
| -0.67 | 0.25 | -0.67 | -0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.36 | 0.00 | 0.10 | 0.06 |
| -0.67 | -0.25 | -0.67 | -0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.35 | 0.00 | 0.07 | 0.04 |
| B: GTL-Disturbances are generated with $(\alpha,\delta)=(0.1436,0)$ for $n=250$ | | | | | | | | | | | | |
| 0.1436 | 0.00 | 0.1436 | 0.00 | 0.94 | 0.27 | 0.64 | 0.48 | 0.48 | 0.54 | 0.00 | 0.00 | 0.00 |
| -0.67 | 0.00 | -0.67 | 0.00 | 0.00 | 0.00 | 0.61 | 0.00 | 0.00 | 0.41 | 0.00 | 0.08 | 0.14 |
| -0.67 | 0.25 | -0.67 | 0.25 | 0.00 | 0.00 | 0.37 | 0.00 | 0.00 | 0.66 | 0.00 | 0.07 | 0.05 |
| -0.67 | 0.25 | -0.67 | -0.25 | 0.00 | 0.00 | 0.55 | 0.00 | 0.00 | 0.75 | 0.00 | 0.06 | 0.04 |
| -0.67 | -0.25 | -0.67 | -0.25 | 0.00 | 0.00 | 0.47 | 0.00 | 0.00 | 0.31 | 0.00 | 0.07 | 0.17 |
| C: GTL-Disturbances are generated with $(\alpha,\delta)=(-0.67,0.25)$ for $n=5000$ | | | | | | | | | | | | |
| 0.1436 | 0.00 | 0.1436 | 0.00 | 0.00 | 0.00 | 0.00 | 0.38 | 0.24 | 0.72 | 0.14 | 0.50 | 0.29 |
| 0.1436 | 0.00 | -0.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.39 | 0.37 | 0.72 | 0.14 | 0.50 | 0.30 |
| 0.1436 | 0.00 | -0.33 | 0.10 | 0.00 | 0.00 | 0.00 | 0.39 | 0.06 | 0.92 | 0.14 | 0.51 | 0.29 |
| 0.1436 | 0.00 | -0.67 | 0.00 | 0.00 | 0.00 | 0.00 | 0.41 | 0.00 | 0.76 | 0.13 | 0.51 | 0.31 |
| 0.1436 | 0.00 | -0.67 | 0.25 | 0.00 | 0.00 | 0.00 | 0.39 | 0.00 | 0.88 | 0.14 | 0.52 | 0.30 |
| -0.67 | 0.00 | -0.67 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.76 | 0.13 | 0.51 | 0.30 |
| -0.67 | 0.25 | -0.67 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.87 | 0.13 | 0.54 | 0.31 |
| -0.67 | 0.25 | -0.67 | -0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.61 | 0.12 | 0.51 | 0.32 |
| -0.67 | -0.25 | -0.67 | -0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.12 | 0.51 | 0.31 |

Note: These test results pertain to simulations reported in Table 4.

Table E.5: Coverage rates of OLS and GTL estimators under a 95% confidence level: design with thick-tailed determinants

| DGP for $x_1$ | | DGP for $x_2$ | | OLS | | | MLE | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha_{x2}$ | $\delta_{x2}$ | $\alpha_{x3}$ | $\delta_{x3}$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\sigma$ | $\alpha$ | $\delta$ |
| A: GTL-Disturbances are generated with $(\alpha, \delta) = (-0.67, 0.25)$ for $n = 250$ | | | | | | | | | | | | |
| 0.1436 | 0.00 | 0.1436 | 0.00 | 0.952 | 0.947 | 0.323 | 0.951 | 0.944 | 0.951 | 0.939 | 0.945 | 0.935 |
| 0.1436 | 0.00 | -0.33 | 0.00 | 0.953 | 0.950 | 0.325 | 0.953 | 0.915 | 0.952 | 0.937 | 0.944 | 0.935 |
| 0.1436 | 0.00 | -0.33 | 0.10 | 0.951 | 0.951 | 0.328 | 0.952 | 0.909 | 0.950 | 0.937 | 0.944 | 0.935 |
| 0.1436 | 0.00 | -0.67 | 0.00 | 0.953 | 0.957 | 0.329 | 0.951 | 0.838 | 0.951 | 0.938 | 0.946 | 0.936 |
| 0.1436 | 0.00 | -0.67 | 0.25 | 0.951 | 0.966 | 0.334 | 0.954 | 0.793 | 0.944 | 0.941 | 0.944 | 0.934 |
| -0.67 | 0.00 | -0.67 | 0.00 | 0.954 | 0.957 | 0.328 | 0.839 | 0.831 | 0.953 | 0.938 | 0.942 | 0.935 |
| -0.67 | 0.25 | -0.67 | 0.25 | 0.952 | 0.965 | 0.358 | 0.780 | 0.791 | 0.950 | 0.941 | 0.945 | 0.934 |
| -0.67 | 0.25 | -0.67 | -0.25 | 0.951 | 0.964 | 0.358 | 0.790 | 0.788 | 0.950 | 0.940 | 0.946 | 0.932 |
| -0.67 | -0.25 | -0.67 | -0.25 | 0.964 | 0.963 | 0.342 | 0.779 | 0.790 | 0.949 | 0.936 | 0.943 | 0.930 |
| B: GTL-Disturbances are generated with $(\alpha, \delta) = (0.1436, 0)$ for $n = 250$ | | | | | | | | | | | | |
| 0.1436 | 0.00 | 0.1436 | 0.00 | 0.948 | 0.947 | 0.954 | 0.946 | 0.934 | 0.943 | 0.939 | 0.917 | 0.925 |
| -0.67 | 0.00 | -0.67 | 0.00 | 0.945 | 0.943 | 0.953 | 0.945 | 0.938 | 0.945 | 0.946 | 0.920 | 0.931 |
| -0.67 | 0.25 | -0.67 | 0.25 | 0.952 | 0.947 | 0.947 | 0.946 | 0.943 | 0.939 | 0.948 | 0.930 | 0.927 |
| -0.67 | 0.25 | -0.67 | -0.25 | 0.951 | 0.961 | 0.951 | 0.945 | 0.955 | 0.948 | 0.946 | 0.924 | 0.929 |
| -0.67 | -0.25 | -0.67 | -0.25 | 0.948 | 0.961 | 0.951 | 0.946 | 0.954 | 0.948 | 0.947 | 0.922 | 0.929 |
| C: GTL-Disturbances are generated with $(\alpha, \delta) = (-0.67, 0.25)$ for $n = 5000$ | | | | | | | | | | | | |
| 0.1436 | 0.00 | 0.1436 | 0.00 | 0.954 | 0.947 | 0.098 | 0.945 | 0.954 | 0.945 | 0.950 | 0.935 | 0.945 |
| 0.1436 | 0.00 | -0.33 | 0.00 | 0.955 | 0.948 | 0.099 | 0.946 | 0.950 | 0.945 | 0.949 | 0.936 | 0.943 |
| 0.1436 | 0.00 | -0.33 | 0.10 | 0.954 | 0.952 | 0.102 | 0.946 | 0.931 | 0.947 | 0.949 | 0.935 | 0.943 |
| 0.1436 | 0.00 | -0.67 | 0.00 | 0.954 | 0.972 | 0.098 | 0.948 | 0.854 | 0.948 | 0.947 | 0.936 | 0.943 |
| 0.1436 | 0.00 | -0.67 | 0.25 | 0.955 | 0.982 | 0.099 | 0.947 | 0.816 | 0.948 | 0.949 | 0.937 | 0.942 |
| -0.67 | 0.00 | -0.67 | 0.00 | 0.980 | 0.972 | 0.098 | 0.863 | 0.857 | 0.947 | 0.947 | 0.936 | 0.945 |
| -0.67 | 0.25 | -0.67 | 0.25 | 0.988 | 0.982 | 0.098 | 0.782 | 0.816 | 0.950 | 0.948 | 0.937 | 0.945 |
| -0.67 | 0.25 | -0.67 | -0.25 | 0.988 | 0.987 | 0.101 | 0.783 | 0.792 | 0.946 | 0.950 | 0.936 | 0.944 |
| -0.67 | -0.25 | -0.67 | -0.25 | 0.985 | 0.987 | 0.101 | 0.789 | 0.793 | 0.943 | 0.948 | 0.936 | 0.944 |

Note: These test results pertain to simulations reported in Table 4.

Table E.6: Ratio of average estimated variance to Monte Carlo variance: design with thick-tailed determinants

| $\alpha_{x2}$ | $\delta_{x2}$ | $\alpha_{x3}$ | $\delta_{x3}$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\sigma$ | $\alpha$ | $\delta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| DGP for $x_1$ | | DGP for $x_2$ | | OLS | | | MLE | | | | | |

A: GTL-Disturbances are generated with $(\alpha, \delta) = (-0.67, 0.25)$ for $n = 250$

| $\alpha_{x2}$ | $\delta_{x2}$ | $\alpha_{x3}$ | $\delta_{x3}$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\sigma$ | $\alpha$ | $\delta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.1436 | 0.00 | 0.1436 | 0.00 | 0.792 | 0.958 | 1.038 | 0.987 | 0.920 | 1.031 | 0.977 | 1.020 | 0.946 |
| 0.1436 | 0.00 | -0.33 | 0.00 | 0.792 | 0.883 | 1.043 | 0.992 | 0.860 | 1.026 | 0.974 | 1.018 | 0.945 |
| 0.1436 | 0.00 | -0.33 | 0.10 | 0.788 | 0.667 | 1.079 | 0.993 | 0.830 | 1.028 | 0.976 | 1.018 | 0.945 |
| 0.1436 | 0.00 | -0.67 | 0.00 | 0.794 | 0.792 | 1.050 | 0.995 | 0.743 | 1.025 | 0.982 | 1.022 | 0.943 |
| 0.1436 | 0.00 | -0.67 | 0.25 | 0.788 | 0.568 | 1.109 | 0.996 | 0.630 | 1.013 | 0.985 | 1.027 | 0.941 |
| -0.67 | 0.00 | -0.67 | 0.00 | 0.833 | 0.773 | 1.050 | 0.785 | 0.707 | 1.029 | 0.976 | 1.026 | 0.938 |
| -0.67 | 0.25 | -0.67 | 0.25 | 0.718 | 0.558 | 1.105 | 0.559 | 0.664 | 1.007 | 0.986 | 1.033 | 0.935 |
| -0.67 | 0.25 | -0.67 | -0.25 | 0.724 | 2.408 | 1.018 | 0.564 | 0.588 | 1.019 | 0.975 | 1.025 | 0.931 |
| -0.67 | -0.25 | -0.67 | -0.25 | 2.115 | 2.285 | 1.028 | 0.548 | 0.608 | 1.014 | 0.970 | 1.023 | 0.939 |

B: GTL-Disturbances are generated with $(\alpha, \delta) = (0.1436, 0)$ for $n = 250$

| $\alpha_{x2}$ | $\delta_{x2}$ | $\alpha_{x3}$ | $\delta_{x3}$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\sigma$ | $\alpha$ | $\delta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.1436 | 0.00 | 0.1436 | 0.00 | 1.045 | 0.973 | 0.979 | 1.010 | 0.933 | 0.986 | 0.903 | 0.876 | 0.852 |
| -0.67 | 0.00 | -0.67 | 0.00 | 1.051 | 1.000 | 0.972 | 1.037 | 0.983 | 0.989 | 0.921 | 0.903 | 0.880 |
| -0.67 | 0.25 | -0.67 | 0.25 | 1.038 | 1.007 | 0.960 | 1.024 | 0.992 | 0.965 | 0.922 | 0.906 | 0.870 |
| -0.67 | 0.25 | -0.67 | -0.25 | 1.037 | 1.002 | 0.961 | 1.018 | 0.980 | 0.975 | 0.909 | 0.895 | 0.878 |
| -0.67 | -0.25 | -0.67 | -0.25 | 1.006 | 1.011 | 0.963 | 0.993 | 0.991 | 0.980 | 0.919 | 0.904 | 0.884 |

C: GTL-Disturbances are generated with $(\alpha, \delta) = (-0.67, 0.25)$ for $n = 5000$

| $\alpha_{x2}$ | $\delta_{x2}$ | $\alpha_{x3}$ | $\delta_{x3}$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_1$ | $\sigma$ | $\alpha$ | $\delta$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.1436 | 0.00 | 0.1436 | 0.00 | 2.707 | 0.589 | 0.989 | 1.034 | 1.039 | 1.010 | 0.918 | 0.917 | 0.924 |
| 0.1436 | 0.00 | -0.33 | 0.00 | 2.714 | 1.299 | 0.986 | 1.033 | 1.029 | 1.011 | 0.918 | 0.918 | 0.924 |
| 0.1436 | 0.00 | -0.33 | 0.10 | 2.710 | 1.914 | 0.935 | 1.037 | 1.019 | 1.005 | 0.918 | 0.917 | 0.924 |
| 0.1436 | 0.00 | -0.67 | 0.00 | 2.710 | 7.029 | 0.979 | 1.036 | 0.842 | 1.013 | 0.919 | 0.918 | 0.925 |
| 0.1436 | 0.00 | -0.67 | 0.25 | 1.516 | 85.169 | 1.013 | 1.039 | 0.753 | 1.008 | 0.918 | 0.918 | 0.926 |
| -0.67 | 0.00 | -0.67 | 0.00 | 1.836 | 6.924 | 0.980 | 0.836 | 0.840 | 1.015 | 0.917 | 0.918 | 0.925 |
| -0.67 | 0.25 | -0.67 | 0.25 | 28.386 | 77.614 | 0.978 | 0.685 | 0.758 | 1.007 | 0.917 | 0.919 | 0.926 |
| -0.67 | 0.25 | -0.67 | -0.25 | 14.215 | 12.190 | 0.983 | 0.683 | 0.623 | 1.010 | 0.919 | 0.919 | 0.924 |
| -0.67 | -0.25 | -0.67 | -0.25 | 3.485 | 12.194 | 0.982 | 0.555 | 0.621 | 1.020 | 0.918 | 0.919 | 0.926 |

Note: These test results pertain to simulations reported in Table 4.

# F  Applications: Variable Definitions and Descriptive Statistics

Table F.1: Hourly wages, males and females, MORG 1998: Descriptive statistics

| Variable | Definition | Males | | Females | |
|---|---|---|---|---|---|
| | | Mean | Std. | Mean | Std. |
| lnWage | Log of hourly wage (dependent variable) | 2.635 | 0.558 | 2.412 | 0.518 |
| YrEduc | Years of schooling | 13.693 | 2.297 | 13.798 | 2.137 |
| Age16 | $(Age - 16)/10$ | 2.259 | 1.122 | 2.278 | 1.126 |
| Age16sq | $(Age - 16)^2/100$ | 6.360 | 5.447 | 6.457 | 5.434 |
| MidAtla | Dummy, =1 if residing in a Mid-Atlantic state | 0.126 | 0.332 | 0.126 | 0.332 |
| EastNoC | Dummy, =1 if residing in a East North-Central | 0.156 | 0.362 | 0.148 | 0.355 |
| WestNoC | Dummy, =1 if residing in a West North-Central | 0.098 | 0.298 | 0.104 | 0.305 |
| SouthAtl | Dummy, =1 if residing in a South-Atlantic stat | 0.157 | 0.364 | 0.168 | 0.374 |
| eastsoC | Dummy, =1 if residing in a East South-Central | 0.048 | 0.213 | 0.050 | 0.218 |
| WestSoC | Dummy, =1 if residing in a West South-Central | 0.090 | 0.286 | 0.091 | 0.287 |
| Mountain | Dummy, =1 if residing in a Mountain state | 0.120 | 0.324 | 0.115 | 0.319 |
| Pacific | Dummy, =1 if residing in a Pacific state | 0.130 | 0.336 | 0.122 | 0.327 |
| Afr. Amer | Dummy, =1 if African-American ethnicity | 0.078 | 0.268 | 0.114 | 0.317 |
| Hispanic | Dummy, =1 if Hispanic ethnicity | 0.103 | 0.304 | 0.083 | 0.276 |
| Asian | Dummy, =1 if Asian-American ethnicity | 0.038 | 0.192 | 0.040 | 0.195 |
| Indian | Dummy, =1 if American Indian ethnicity | 0.011 | 0.104 | 0.012 | 0.109 |
| N of obs | | 54687 | | 46045 | |

Merged Outgoing Rotation Group data of the Current Population Survey are available online at `http://www.nber.org/morg/annual/`.

Table F.2: Residential Home Prices, Windsor, Canada, 1987: Descriptive statistics

| Variable | Description | Mean | St.Dev | Minimum | Maximum |
|---|---|---|---|---|---|
| lnp | Log of sale price (dependent variable) | 11.059 | 0.372 | 10.127 | 12.155 |
| lnlot | Lot size of the property in square feet | 8.467 | 0.398 | 7.409 | 9.693 |
| bdms3 | Dummy, =1 if house has 3 bedrooms | 0.551 | 0.498 | 0 | 1 |
| bdms4 | Dummy, =1 if house has 4 or more bedrooms | 0.196 | 0.397 | 0 | 1 |
| fb2 | Dummy, =1 if house has 2 bathrooms | 0.244 | 0.430 | 0 | 1 |
| fb3 | Dummy, =1 if house has 3 or more bathrooms | 0.020 | 0.141 | 0 | 1 |
| sty | Number of floors, excluding basement | 1.808 | 0.868 | 1 | 4 |
| drv | Dummy, =1 if house has a driveway | 0.859 | 0.348 | 0 | 1 |
| rec | Dummy, =1 if house has a recreational room | 0.178 | 0.383 | 0 | 1 |
| ffin | Dummy, =1 if house has a full finished basement | 0.350 | 0.477 | 0 | 1 |
| ghw | Dummy, =1 if house uses gas for hot water heating | 0.046 | 0.209 | 0 | 1 |
| ca | Dummy, =1 if house has central air conditioning | 0.317 | 0.466 | 0 | 1 |
| gar1 | Dummy, =1 if house has a garage for 1 car | 0.231 | 0.422 | 0 | 1 |
| gar2 | Dummy, =1 if house has a garage for 2 or more cars | 0.220 | 0.414 | 0 | 1 |
| reg | Dummy, =1 if house in Riverside or South Windsor | 0.234 | 0.424 | 0 | 1 |
| N of obs | | 546 | | | |

Source: Derived from the dataset used in Anglin and Gencay (1996). Data are available online at http://qed.econ.queensu.ca/jae/1996-v11.6/anglin-gencay/.

Table F.3: Speeding tickets, Massachusetts, 1987: Descriptive statistics

| Variable | Definition | Mean | St.Dev | Min | Max |
|---|---|---|---|---|---|
| Amount | Fine amount (in $) | 122.03 | 56.25 | 3.00 | 725.00 |
| ln(Amount) | log of Amount | 4.707 | 0.438 | 1.099 | 6.586 |
| Mph over | Miles per hour over the speed limit | 17.08 | 5.79 | 1 | 75 |
| ln(Mph over) | log of mph over | 2.783 | 0.333 | 0 | 4.317 |
| Afr.American | 1 if the driver is African American | 0.051 | 0.219 | 0 | 1 |
| Hispanic | 1 if the driver is Hispanic | 0.047 | 0.211 | 0 | 1 |
| Female | 1 if the driver is female | 0.332 | 0.471 | 0 | 1 |
| ln(Age) | log of age (in years) | 3.442 | 0.366 | 2.485 | 4.585 |
| OutTown | 1 if out of town driver | 0.847 | 0.360 | 0 | 1 |
| OutState | 1 if out of state driver | 0.221 | 0.415 | 0 | 1 |
| ln(CourtDist) | log of distance to court (in miles) | 2.886 | 1.298 | 1.609 | 8.529 |
| ln(Pvalue.pc) | log property value per capita | 11.165 | 0.499 | 9.828 | 13.580 |
| OR | 1 if a tax increase rejected via override refe | 0.026 | 0.160 | 0 | 1 |
| SP | 1 if the officer is state police | 0.445 | 0.497 | 0 | 1 |
| N of obs | | 31674 | | | |

Source: Derived from the dataset used in Makowsky and Stratmann (2009); variable names have been slightly changed. The data are available online at http://www.aeaweb.org/issue.php?journal=AER&volume=99&issue=1.

Table F.4: Trade creation and trade diversion, 1960-2000: Descriptive statistics

| Variable | Definition | Mean | St.Dev | Min | Max |
|---|---|---|---|---|---|
| lnimportij | log of bilateral imports from j to i (current US dollars) | 11.14 | 3.27 | -3.54 | 21.01 |
| Trade creation dummy variables | | | | | |
| tc.nafta | 1 = both countries are NAFTA members | 0.000 | 0.019 | 0 | 1 |
| tc.eu | 1 = both countries are EU members | 0.016 | 0.124 | 0 | 1 |
| tc.efta | 1 = both countries are EFTA members | 0.006 | 0.077 | 0 | 1 |
| mx.eea | 1 = both countries are EEA members | 0.012 | 0.108 | 0 | 1 |
| tc.caricom | 1 = both countries are CARICOM members | 0.001 | 0.030 | 0 | 1 |
| tc.ap | 1 = both countries are Andean Pact members | 0.004 | 0.061 | 0 | 1 |
| tc.mercosur | 1 = both countries are MERCOSUR members | 0.001 | 0.025 | 0 | 1 |
| tc.asean | 1 = both countries are ASEAN FTA members | 0.001 | 0.027 | 0 | 1 |
| tc.certain | 1 = both countries are ANZCERTA members | 0.000 | 0.015 | 0 | 1 |
| tc.apec | 1 = both countries are APEC members | 0.013 | 0.114 | 0 | 1 |
| tc.laia | 1 = both countries are LAIA members | 0.021 | 0.143 | 0 | 1 |
| tc.cacm | 1 = both countries are CACM members | 0.004 | 0.064 | 0 | 1 |
| tc.bilateralPTA | 1 = both countries are in a joint BPTA | 0.005 | 0.070 | 0 | 1 |
| Trade diversion dummy variables | | | | | |
| td.nafta | 1 = only one country is a NAFTA member | 0.033 | 0.178 | 0 | 1 |
| td.eu | 1 = only one country is an EU member | 0.262 | 0.440 | 0 | 1 |
| td.efta | 1 = only one country is an EFTA member | 0.160 | 0.367 | 0 | 1 |
| td.eea | 1 = only one country is an EEA member | 0.107 | 0.310 | 0 | 1 |
| td.caricom | 1 = only one country is a CARICOM member | 0.047 | 0.211 | 0 | 1 |
| td.ap | 1 = only one country is an Andean Pact member | 0.091 | 0.287 | 0 | 1 |
| td.mercosur | 1 = only one country is a MERCOSUR member | 0.026 | 0.159 | 0 | 1 |
| td.asean | 1 = only one country is an ASEAN FTA member | 0.033 | 0.178 | 0 | 1 |
| td.anzcerta | 1 = only one country is an ANZCERTA member | 0.032 | 0.176 | 0 | 1 |
| td.apec | 1 = only one country is an APEC member | 0.134 | 0.340 | 0 | 1 |
| td.laia | 1 = only one country is a LAIA member | 0.200 | 0.400 | 0 | 1 |
| td.cacm | 1 = only one country is a CACM member | 0.083 | 0.276 | 0 | 1 |
| td.bilateralPTA | 1 = only one country is a member of a BPTA | 0.170 | 0.376 | 0 | 1 |
| Non-PTA control variables | | | | | |
| lpgdpij | Sum over (i,j) of log nominal GDP | 6.58 | 3.06 | -4.29 | 17.65 |
| lpgdppcij | Sum over (i,j) of log real GDP per capita | 17.38 | 1.41 | 12.24 | 20.85 |
| ldist | Log of bilateral distance | 8.23 | 0.81 | 4.40 | 9.42 |
| sachsij | Sum over (i,j) of the Sachs-Warner trade policy index | 1.22 | 0.70 | 0 | 2 |
| vola3 | St.dev. of the volatility in the bilateral exchange rate | 4.83 | 7.32 | 0 | 97.61 |
| floatij | Sum over (i,j) of dummy: 1 if floating exchange | 0.71 | 0.73 | 0 | 2 |
| cu | 1 = a common currency union | 0.009 | 0.097 | 0 | 1 |
| adifsecschool25 | Abs. log difference in years of secondary schooling | 1.13 | 0.91 | 0.00 | 5.86 |
| adifdensity | Abs. log difference in population density | 1.70 | 1.32 | 0.00 | 8.24 |
| adifgdppc | Abs. log difference of real GDP per capita | 1.25 | 0.91 | 0.00 | 4.07 |
| border | 1 = common land border | 0.033 | 0.180 | 0 | 1 |
| islandij | Sum over (i,j) of dummy: 1 if island | 0.272 | 0.487 | 0 | 2 |
| landlockij | Sum over (i,j) of dummy: 1 if landlocked | 0.228 | 0.448 | 0 | 2 |
| lpareaij | Sum over(i,j) of log surface area | 24.69 | 2.89 | 11.82 | 32.08 |

<div align="right">(continued)</div>

| Variable | Definition | Mean | St.Dev | Min | Max |
|----------|-----------|------|--------|-----|-----|
| lremoteij | Remoteness | 17.94 | 0.35 | 16.97 | 19.03 |
| colony | 1 = one country was a former colony of the other | 0.028 | 0.165 | 0 | 1 |
| comcol | 1 = common colonizer | 0.064 | 0.244 | 0 | 1 |
| comlang | 1 = common language | 0.235 | 0.424 | 0 | 1 |
| N of obs | | 37983 | | | |

Source: Derived from the dataset used in Eicher et al. (2012); variable names have been slightly changed. "Remoteness" is defined as the sum of the log of the average distance, weighed by relative GDP, of each country from all trading partners. The data are available online at `http://qed.econ.queensu.ca/jae/2012-v27.2 /eicher-henn-papageorgiou/`.

Table F.5: Capital Assets Price Model, 1960:1 - 2012:12: Descriptive statistics

| Variable | Definition | Mean | St.Dev | Min | Max |
|---|---|---|---|---|---|
| $R_{11} - R_f$ | Excess return in portfolio with ME:1 and B/M:1 | 0.207 | 8.058 | -34.81 | 39.36 |
| $R_{12} - R_f$ | Excess return in portfolio with ME:1 and B/M:2 | 0.715 | 6.904 | -31.53 | 38.19 |
| $R_{13} - R_f$ | Excess return in portfolio with ME:1 and B/M:3 | 0.771 | 6.004 | -29.29 | 27.55 |
| $R_{14} - R_f$ | Excess return in portfolio with ME:1 and B/M:4 | 0.949 | 5.663 | -29.49 | 27.20 |
| $R_{15} - R_f$ | Excess return in portfolio with ME:1 and B/M:5 | 1.083 | 6.108 | -29.35 | 32.85 |
| $R_{21} - Rf$ | Excess return in portfolio with ME:2 and B/M:1 | 0.384 | 7.224 | -33.31 | 26.91 |
| $R_{22} - R_f$ | Excess return in portfolio with ME:2 and B/M:2 | 0.664 | 6.010 | -32.14 | 25.54 |
| $R_{23} - R_f$ | Excess return in portfolio with ME:2 and B/M:3 | 0.850 | 5.461 | -28.36 | 25.76 |
| $R_{24} - R_f$ | Excess return in portfolio with ME:2 and B/M:4 | 0.896 | 5.319 | -26.64 | 26.76 |
| $R_{25} - R_f$ | Excess return in portfolio with ME:2 and B/M:5 | 0.963 | 6.024 | -29.44 | 29.46 |
| $R_{31} - R_f$ | Excess return in portfolio with ME:3 and B/M:1 | 0.438 | 6.673 | -30.23 | 24.01 |
| $R_{32} - R_f$ | Excess return in portfolio with ME:3 and B/M:2 | 0.717 | 5.474 | -29.69 | 24.45 |
| $R_{33} - R_f$ | Excess return in portfolio with ME:3 and B/M:3 | 0.713 | 5.045 | -25.07 | 21.36 |
| $R_{34} - R_f$ | Excess return in portfolio with ME:3 and B/M:4 | 0.828 | 4.957 | -23.42 | 22.82 |
| $R_{35} - R_f$ | Excess return in portfolio with ME:3 and B/M:5 | 0.973 | 5.521 | -26.77 | 28.62 |
| $R_{41} - R_f$ | Excess return in portfolio with ME:4 and B/M:1 | 0.537 | 5.934 | -26.54 | 25.24 |
| $R_{42} - R_f$ | Excess return in portfolio with ME:4 and B/M:2 | 0.531 | 5.185 | -29.43 | 19.87 |
| $R_{43} - R_f$ | Excess return in portfolio with ME:4 and B/M:3 | 0.675 | 5.070 | -26.24 | 23.43 |
| $R_{44} - R_f$ | Excess return in portfolio with ME:4 and B/M:4 | 0.797 | 4.845 | -21.11 | 23.74 |
| $R_{45} - R_f$ | Excess return in portfolio with ME:4 and B/M:5 | 0.778 | 5.527 | -24.44 | 27.32 |
| $R_{51} - R_f$ | Excess return in portfolio with ME:5 and B/M:1 | 0.403 | 4.748 | -22.24 | 21.84 |
| $R_{52} - R_f$ | Excess return in portfolio with ME:5 and B/M:2 | 0.476 | 4.472 | -22.96 | 16.11 |
| $R_{53} - R_f$ | Excess return in portfolio with ME:5 and B/M:3 | 0.479 | 4.386 | -22.42 | 18.12 |
| $R_{54} - R_f$ | Excess return in portfolio with ME:5 and B/M:4 | 0.507 | 4.381 | -19.40 | 19.18 |
| $R_{55} - R_f$ | Excess return in portfolio with ME:5 and B/M:5 | 0.583 | 5.038 | -19.59 | 17.56 |
| $R_m - R_f$ | Excess return in the overall market | 0.458 | 4.490 | -23.24 | 16.10 |
| $F_{SMB}$ | Difference in returns of portfolios of stocks with small and big market equity (ME) | 0.219 | 3.067 | -16.39 | 22.00 |
| $F_{HML}$ | Difference in returns of portfolios of stocks with high and low B/M ratios (B/M) | 0.390 | 2.844 | -12.60 | 13.84 |
| N of obs | | 635 | | | |

## G   Applications: Further Results

Table G.1: Log-wage equations, female wage and salaried workers, MORG 1998

| | OLS | | GTL | | $\frac{\hat{\beta}_{OLS}-\hat{\beta}_{GTL}}{\hat{\beta}_{GTL}}$ | $\frac{SE_{GTL}}{SE_{OLS}}$ |
| | Estimate | Stan.Err. | Estimate | Stan.Err. | | |
|---|---|---|---|---|---|---|
| Years of Schooling | 0.110 | 0.001 | 0.113 | 0.001 | -0.022 | 0.960 |
| $\text{Age}-16$ | 0.347 | 0.007 | 0.344 | 0.007 | 0.007 | 0.944 |
| $(\text{Age}-16)^2$ | -0.055 | 0.001 | -0.055 | 0.001 | 0.013 | 0.952 |
| MidAtla | 0.052 | 0.009 | 0.057 | 0.009 | -0.095 | 0.960 |
| EastNoC | -0.009 | 0.009 | -0.011 | 0.008 | -0.116 | 0.950 |
| WestNoC | -0.129 | 0.009 | -0.132 | 0.009 | -0.022 | 0.948 |
| SouthAtl | -0.048 | 0.009 | -0.053 | 0.008 | -0.087 | 0.952 |
| EastSoC | -0.147 | 0.011 | -0.154 | 0.011 | -0.045 | 0.943 |
| WestSoC | -0.116 | 0.010 | -0.122 | 0.009 | -0.050 | 0.951 |
| Mountain | -0.096 | 0.009 | -0.097 | 0.009 | -0.008 | 0.953 |
| Pacific | 0.074 | 0.009 | 0.075 | 0.009 | -0.009 | 0.960 |
| Afr. American | -0.056 | 0.006 | -0.059 | 0.006 | -0.058 | 0.952 |
| Hispanic | -0.110 | 0.008 | -0.107 | 0.007 | 0.021 | 0.953 |
| Asian | -0.068 | 0.011 | -0.061 | 0.010 | 0.111 | 0.976 |
| Indian | -0.048 | 0.018 | -0.051 | 0.017 | -0.069 | 0.954 |
| Intercept | 0.515 | 0.016 | 0.488 | 0.015 | | |
| $\sigma$ | | | 0.236 | 0.002 | | |
| $\alpha$ | | | 0.003 | 0.004 | | |
| $\delta$ | | | 0.013 | 0.003 | | |
| logL | -26051.5 | | -25242.4 | | | |
| (Absolute) Average | | | | | 0.049 | 0.953 |

Dependent variable: Log of hourly wage. Number of observations = 46045. Skewness and kurtosis of OLS residuals equal $-0.19$ and $4.89$; the Jarque-Bera test of normality of the OLS residuals has a $p$-value of less than 0.001. The Wald test of the GTL estimates of $(\alpha, delta)$ equals 1339.9, rejecting normality with a $p$-value of less than 0.001. The LM test equals 670.4 with a $p$-value of less than 0.001 and with 52 range violations. The Vuong test that compares OLS and GTL equals $-11.80$ in favor of the GTL model with a $p$-value of less than 0.001.

Table G.2: Trade creation and diversion, 1960-2000: Slopes of control variables

| | OLS | | GTL | | $\frac{\hat{\beta}_{OLS}-\hat{\beta}_{GTL}}{\hat{\beta}_{GTL}}$ | $\frac{SE_{GTL}}{SE_{OLS}}$ |
| | Estimate | Stan.Err. | Estimate | Stan.Err. | | |
|---|---|---|---|---|---|---|
| lprdctgdpij | 0.942 | 0.012 | 0.904 | 0.012 | 0.043 | 0.954 |
| lprdctgdpp j | 0.277 | 0.027 | 0.268 | 0.026 | 0.034 | 0.948 |
| ldist | -1.078 | 0.034 | -1.014 | 0.031 | 0.063 | 0.915 |
| sachsij | 0.214 | 0.029 | 0.148 | 0.027 | 0.449 | 0.936 |
| vola3 | 0.000 | 0.002 | 0.000 | 0.001 | -1.280 | 0.882 |
| floatij | 0.097 | 0.020 | 0.050 | 0.017 | 0.923 | 0.849 |
| cu | 1.212 | 0.197 | 1.198 | 0.199 | 0.012 | 1.011 |
| diffsecsc 25 | 0.038 | 0.025 | 0.026 | 0.023 | 0.447 | 0.930 |
| diffdensity | 0.126 | 0.014 | 0.104 | 0.013 | 0.205 | 0.905 |
| diffgdppc | 0.060 | 0.028 | 0.054 | 0.026 | 0.119 | 0.960 |
| border | 0.404 | 0.152 | 0.386 | 0.130 | 0.047 | 0.855 |
| islandij | -0.222 | 0.050 | -0.214 | 0.047 | 0.039 | 0.931 |
| landlockij | -0.271 | 0.046 | -0.263 | 0.044 | 0.032 | 0.951 |
| lprdctareaij | -0.083 | 0.010 | -0.072 | 0.009 | 0.147 | 0.955 |
| lprdctremo j | 1.327 | 0.086 | 1.366 | 0.080 | -0.029 | 0.936 |
| colony | 1.123 | 0.124 | 1.089 | 0.104 | 0.031 | 0.841 |
| comcol | 0.556 | 0.095 | 0.561 | 0.090 | -0.007 | 0.939 |
| comlang | 0.277 | 0.051 | 0.258 | 0.046 | 0.074 | 0.904 |

Notes: See Table 8.

Table G.3: Trade creation and diversion, 1960-2000: Adding continent import-export pair dummies

| | OLS | | GTL | | $\frac{\hat{\beta}_{OLS}-\hat{\beta}_{GTL}}{\hat{\beta}_{GTL}}$ | $\frac{SE_{GTL}}{SE_{OLS}}$ |
| | Estimate | Stan.Err. | Estimate | Stan.Err. | | |
|---|---|---|---|---|---|---|
| Trade creation dummy variables | | | | | | |
| tc.nafta | -0.353 | 0.350 | -0.011 | 0.328 | 32.493 | 0.937 |
| tc.eu | 0.985 | 0.125 | 0.783 | 0.108 | 0.257 | 0.861 |
| tc.efta | 1.332 | 0.140 | 1.161 | 0.125 | 0.147 | 0.893 |
| tc.eea | 0.292 | 0.088 | 0.350 | 0.075 | -0.168 | 0.848 |
| tc.caricom | 1.950 | 0.484 | 1.791 | 0.424 | 0.089 | 0.876 |
| tc.ap | 0.867 | 0.190 | 0.831 | 0.173 | 0.043 | 0.912 |
| tc.mercosur | 1.009 | 0.305 | 1.010 | 0.324 | -0.001 | 1.063 |
| tc.asean | 0.435 | 0.216 | 0.503 | 0.183 | -0.135 | 0.843 |
| tc.anzcerta | -0.694 | 0.313 | -0.435 | 0.259 | 0.598 | 0.828 |
| tc.apec | 1.417 | 0.100 | 1.136 | 0.090 | 0.247 | 0.899 |
| tc.laia | -0.389 | 0.179 | -0.664 | 0.171 | -0.414 | 0.956 |
| tc.cacm | 1.636 | 0.193 | 1.541 | 0.175 | 0.062 | 0.905 |
| tc.bilateralPTA | 0.158 | 0.097 | 0.126 | 0.092 | 0.249 | 0.950 |

(continued)

| | OLS | | GTL | | $\frac{\hat{\beta}_{OLS}-\hat{\beta}_{GTL}}{\hat{\beta}_{GTL}}$ | $\frac{SE_{GTL}}{SE_{OLS}}$ |
|---|---|---|---|---|---|---|
| | Estimate | Stan.Err. | Estimate | Stan.Err. | | |
| Trade diversion dummy variables | | | | | | |
| td.nafta | 0.154 | 0.076 | 0.121 | 0.065 | 0.279 | 0.845 |
| td.eu | 0.790 | 0.057 | 0.625 | 0.052 | 0.265 | 0.903 |
| td.efta | 0.510 | 0.063 | 0.402 | 0.057 | 0.268 | 0.900 |
| td.eea | -0.249 | 0.047 | -0.149 | 0.042 | 0.674 | 0.897 |
| td.caricom | -0.752 | 0.107 | -0.692 | 0.102 | 0.088 | 0.951 |
| td.ap | 0.061 | 0.072 | 0.057 | 0.064 | 0.062 | 0.886 |
| td.mercosur | -0.013 | 0.072 | -0.042 | 0.065 | -0.687 | 0.896 |
| td.asean | 0.435 | 0.070 | 0.395 | 0.061 | 0.100 | 0.878 |
| td.anzcerta | -0.394 | 0.094 | -0.299 | 0.084 | 0.315 | 0.893 |
| td.apec | 0.383 | 0.050 | 0.271 | 0.044 | 0.414 | 0.888 |
| td.laia | -0.870 | 0.085 | -0.852 | 0.082 | 0.022 | 0.967 |
| td.cacm | -0.323 | 0.090 | -0.233 | 0.086 | 0.386 | 0.958 |
| td.bilateralPTA | -0.237 | 0.051 | -0.219 | 0.043 | 0.080 | 0.847 |
| Non-PTA control variables | | | | | | |
| lprdctgdpij | 0.944 | 0.014 | 0.911 | 0.013 | 0.036 | 0.938 |
| lprdctgdpp j | 0.321 | 0.033 | 0.318 | 0.031 | 0.010 | 0.939 |
| ldist | -0.990 | 0.051 | -0.909 | 0.044 | 0.090 | 0.869 |
| sachsij | 0.242 | 0.029 | 0.193 | 0.027 | 0.256 | 0.915 |
| vola3 | 0.000 | 0.002 | 0.000 | 0.001 | -2.109 | 0.877 |
| floatij | 0.092 | 0.020 | 0.035 | 0.017 | 1.635 | 0.841 |
| cu | 1.052 | 0.217 | 0.951 | 0.245 | 0.105 | 1.129 |
| diffsecsc 25 | 0.027 | 0.026 | 0.003 | 0.024 | 7.690 | 0.910 |
| diffdensity | 0.087 | 0.014 | 0.066 | 0.013 | 0.316 | 0.915 |
| diffgdppc | 0.091 | 0.029 | 0.084 | 0.028 | 0.080 | 0.954 |
| border | 0.454 | 0.145 | 0.380 | 0.123 | 0.194 | 0.850 |
| islandij | -0.161 | 0.051 | -0.145 | 0.047 | 0.109 | 0.927 |
| landlockij | -0.274 | 0.045 | -0.275 | 0.041 | -0.004 | 0.929 |
| lprdctareaij | -0.078 | 0.012 | -0.076 | 0.011 | 0.025 | 0.941 |
| lprdctremo j | 1.309 | 0.118 | 1.193 | 0.111 | 0.097 | 0.936 |
| colony | 1.097 | 0.118 | 1.093 | 0.102 | 0.004 | 0.862 |
| comcol | 0.511 | 0.099 | 0.442 | 0.093 | 0.156 | 0.933 |
| comlang | 0.262 | 0.052 | 0.230 | 0.048 | 0.138 | 0.915 |
| $\sigma$ | | | 0.777 | 0.014 | | |
| $\alpha$ | | | -0.093 | 0.009 | | |
| $\delta$ | | | 0.136 | 0.005 | | |
| log Likelihood | -74136.2 | | -71867.5 | | | |
| (Absolute) Average | | | | | 1.006 | 0.915 |

Dependent variable: Log of bilateral imports. The model also includes control variables (reported in Table G.2 in the Appendix) and time dummy variables (not reported). Number of observations = 37983. Skewness and kurtosis of OLS residuals equal $-0.71$ and $4.90$; the Jarque-Bera test of normality of the OLS residuals has a $p$-value of less than 0.001. The Wald test of the GTL estimates of $(\alpha, \delta)$ equals 1482.7, rejecting normality with a $p$-value of less than 0.001. The Vuong test that compares OLS and GTL equals $-25.03$ in favor of the GTL model with a $p$-value of less than 0.001. The GTL estimates imply $\kappa_3 = -2.56$ and $\kappa_4 = 61.72$.

# References

Anglin, P. M., Gencay, R., 1996. Semiparametric estimation of a hedonic price function. Journal of Applied Econometrics 11(6), 633–648.

Blattberg, R., Sargent, T., 1971. Regression with non-gaussian stable disturbances: some sampling results. Econometrica 39(3), 501–510.

Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance 52(1), 25–46.

Carroll, R., Fan, J., Gijbels, I., Wand, M., 1997. Generalized partially linear single-index models. Journal of the American Statistical Association 92(438), 477–489.

Chen, E. H., Dixon, W. J., 1972. Estimators for the linear regression model based on Winsorized observations. Journal of the American Statistical Association 67(339), 664–671.

Chen, L.-A., Welsh, A. H., Chan, W., 2001. Estimators for the linear regression model based on Winsorized observations. Statistica Sinica 11(1), 147–172.

Eicher, T. S., Henn, C., Papageorgiou, C., 2012. Trade creation and diversion revisited: accounting for model uncertainty and natural trading partner effects. Journal of Applied Econometrics 27(2), 296–321.

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33(1), 3–56.

Fama, E. F., French, K. R., 2004. The capital asset pricing model: Theory and evidence. Journal of Economic Perspectives 18(3), 25–46.

Fernancez, C., Steel, M. F., 1999. Multivariate Student-t regression models: Pitfalls and inference. Biometrika 86(1), 153–167.

Fernandez, C., Steel, M. F. J., 1998. On bayesian modeling of fat tails and skewness. Journal of the American Statistical Association 93(441), 359–371.

Ferreira, J., Steel, M., 2006. A constructive representation of univariate skewed distributions. Journal of the American Statistical Association 101(474), 823–829.

Freimer, M., Mudholkar, G. S., Kollia, G., Lin, C. T., 1988. A study of the generalized tukey lambda family. Communications in Statistics: Theory and Methods 17, 3547–3567.

Ghosh, S., Yamarik, S., 2004. Are regional trading arrangement trade creating? An application of extreme bounds analysis. Journal of International Economics 63(2), 369–395.

Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood method: Thoery. Econometrica 52(3), 681–700.

Hallin, M., Swan, Y., Verdebout, T., Veredas, D., 2011. Rank-based testing in linear models with stable errors. Journal of Nonparametric Statistics 23(2), 305–320.

Hallin, M., Swan, Y., Verdebout, T., Veredas, D., 2013. One-step R-estimation in linear models with stable errors. Journal of Econometrics 172(2), 195–204.

Harvey, A. C., Sucarrat, G., 2014. Evaluating density forecasts with applications to financial risk management. Computational Statistics and Data Analysis 76(SI), 320–338.

Haupt, H., Schnurbus, J., Tschernig, R., 2010. On nonparametric estimation of a hedonic price function. Journal of Applied Econometrics 25, 894–901.

He, X., Jureckova, J., Koenker, R., Portnoy, S., 1990. Tail behavior of regression estimators and their breakdown points. Econometrica 58(5), 1195–1214.

Horowitz, J. L., 1998. Semiparametric Methods in Econometrics. Springer-Verlag, New York.

Huber, P. J., 1964. Robust estimation of a location parameter. Annals of Mathematical Statistics 35, 73101.

Huber, P. J., 1981. Robust Statistics. Wiley, New York.

Huber, P. J., Ronchetti, E. M., 2009. Robust Statistics. Wiley, New York, 2nd edition.

Ichimura, H., 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single index models. Journal of Econometrics 38(1-2), 195–204.

Jureckova, J., Koenker, R., Portnoy, S., 2001. Tail behavior of the least-squares estimator. Statistics and Probability Letters 55(4), 377–384.

Karian, Z. A., Dudewicz, E. J., McDonald, P., 1996. The extended Generalized Lambda Distribution system for fitting distributions to data: History, completion of theory, tables, applications, the 'final word' on moment fits. Communications in Statistics–Simulation 25(3), 611–642.

Koenker, R., 2005. Quantile Regression. Cambridge University Press, New York.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46(1), 33–50.

Koenker, R., Portnoy, S., 1987. L-Estimation for linear models. Journal of the American Statistical Association 82(399), 851–857.

Komunjer, I., 2007. Asymmetric power distribution: theory and applications to risk measurement. Journal of Applied Econometrics 22(5), 891–921.

Li, Q., Racine, J. S., 2007. Nonparametric Econometrics. Princeton University Press, Princeton.

Makowsky, M. D., Stratmann, T., 2009. Political economy at any speed: What determines traffic citations? American Economic Review 99(1), 509–527.

Mikosch, T., de Vries, C. G., 2013. Heavy tails of OLS. Journal of Econometrics 172(2), 205–221.

Newey, W. K., McFadden, D., 1994. Large Sample Estimation and Hypothesis Testing, volume IV, chapter 36, pp. 2111–2245. Elsevier Science B.V.

Nolan, J. P., Ojeda-Revah, D., 2013. Linear and nonlinear regression with stable errors. Journal of Econometrics 172, 186–194.

Parmeter, C. F., Henderson, D. J., Kumbhakar, S. C., 2007. Nonparametric estimation of a hedonic price function. Journal of Applied Econometrics 22, 695–699.

Powell, J., Stock, J., Stoker, T. M., 1989. Semiparametric estimation of index coefficients. Econometrica 51(2), 1403–1430.

Ramberg, J., Schmeiser, B., 1974. An approximate method for generating asymmetric random variables. Communications ACM 17, 78–82.

Samorodnitsky, G., Rachev, S. T., Kurz-Kim, J.-R., Stoyanov, S. V., 2007. Asymptotic distribution of unbiased linear estimators in the presence of heavy-tailed stochastic regressors and residuals. Probability and Mathematical Statistics 27(2), 275–302.

Subramanianl, A., Wei, S.-J., 2007. The WTO promotes trade, strongly but unevenly. Journal of International Economics 72(1), 151–175.

Vijverberg, C.-P. C., Vijverberg, W. P., 2012. Pregibit: A family of discrete choice models, IZA Discussion Paper 6359.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50(1), 1–25.

White, H., 1984. Asymptotic Theory for Econometrians. Academic Press, San Diego, California.

Wooldridge, J. M., 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press, Cambridge, Massachusetts.

Yale, C., Forsythe, A. B., 1976. Winsorized regression. Technometrics 18(3), 291–300.