ROBUSTNESS OF BINARY CHOICE MODELS TO CONDITIONAL HETEROSCEDASTICITY

By

Tim Ginker and Offer Lieberman

November 17, 2016

RESEARCH INSTITUTE FOR ECONOMETRICS

DISCUSSION PAPER NO. 6-16

DEPARTMENT OF ECONOMICS
BAR-ILAN UNIVERSITY
RAMAT-GAN 5290002, ISRAEL

http://econ.biu.ac.il/en/node/2473

Robustness of Binary Choice Models to Conditional Heteroscedasticity

Tim Ginker* and Offer Lieberman†

Revised, November 17, 2016

Abstract

We show that when the true data generating process of a large class of binary choice models contains conditional heteroscedasticity, predictions based on the misspecified MLE in which conditional heteroscedasticity is ignored, are unaffected by the misspecification.

Key words and phrases: Conditional heteroscedasticity; Misspecified Models; Probit.

JEL Classification: C22

*Bar-Ilan University.

†Bar-Ilan University. Support from Israel Science Foundation grant No. 1082-14 and from the Sapir Center in Tel Aviv University are gratefully acknowledged. Correspondence to: Department of Economics and Research Institute for Econometrics (RIE), Bar-Ilan University, Ramat Gan 52900, Israel. E-mail: offer.lieberman@gmail.com

1 Introduction

In recent years qualitative response models have become popular in time series analysis. In the financial context, in particular, they have been applied in connection with the sign of asset return forecasts, as these may lead to profitable speculative positions and correct hedging decisions. See, among others, Levich (2011). While, *apriori*, this class of models seems to be less informative than continuous response models, Leung *et. al.*'s (2000) detailed comparative study between binary and continuous response models revealed that the former outperform the latter in its ability to generate trading profits.

Sign forecastability has been assumed to be driven mainly by conditional mean dynamics of the underlying process in most recent studies on this issue. For instance, Nyberg (2011) examined the ability of the binary dependent dynamic probit model to predict the direction of monthly excess stock returns, extending Kauppi and Saikkonen's (2008) model. He concluded that in terms of out-of-sample performance, binary models can be useful in asset allocation decisions, especially when the mean dynamics resemble an error correction specification.

Binary choice models were shown to be useful in the context of decomposition - type models. For instance, Anatolyev and Gospodinov (2010) expressed the financial asset return as a product of its sign and its absolute value. The two components were modeled separately as a copula before a joint forecast was constructed. Earlier, Rydberg and Shephard (2003) specified the stock return as a product of two binary variables, defining the returns direction and market activity, and multiplied by a process which defines the size of a price change.

Previous findings on modeling conditional heteroscedasticity in qualitative response models showed that the volatility parameters are statistically significant and may have a good explanatory power. Specifically, Dueker (1999) modeled the discrete changes in the bank prime lending rate by a dynamic ordered probit with Markov-switching conditional heteroscedasticity. His results indicated that conditional heteroscedasticity plays an important role in explaining the data. Broseta (2000) reported a good fit for a learning model in which the latent residuals were allowed to follow an ARCH(1) process (Engle (1982)). Hausman and Lo (1992) estimated a model for stock price changes with heteroscedastic ordered probit by dividing the price changes into eight intervals.

In the same line of literature, Christoffersen and Diebold (2006) suggested that the volatility and other high - order conditional moments may produce sign dependence, drawing the theoretical connection between the asset return volatility dynamics and its sign forecastability. From their study it follows that directional forecasts could be inferred from the volatility dynamics even when the conditional mean is constant. There are numerous other studies that have documented the strong dependence of asset returns volatility. For surveys of the empirical evidence and volatility modeling in finance, the reader is referred to Mikosch *et. al.* (2009) and Bollerslev *et. al.* (1992).

To this end, we consider in this paper the model

$$y_t = 1\{x_t'\gamma + \varepsilon_t > 0\},\ \ t = 1, ..., n, \tag{1}$$

where $1\{\cdot\}$ is the indicator function which takes the value of unity if the condition in the brackets is satisfied and zero, otherwise, x_t is a $K \times 1$ vector of explanatory variables which are assumed to be ergodic stationary, γ is a $K \times 1$ vector of unknown parameters, and for all t and s, conditional on $\mathcal{F}_{t-1}$ and x_s, $\varepsilon_t \sim F(0, \sigma_t^2)$, where $\mathcal{F}_t$ is the increasing sequence of σ-fields generated by $\{\varepsilon_j\}_{j=1}^t$ and F is a symmetric CDF on which conditions are given in Assumption A below. The conditional variance, σ_t^2, is merely assumed to satisfy some very mild regularity conditions so that the class of heteroscedasticity models allowed is very general and includes in it, as a special case, the prominent GARCH(p,q) specification (Bollerslev (1986)). For this model, under the classical assumptions including a fixed σ_t^2, $\forall t$, the main workhorse for estimating this model is undoubtedly the probit maximum likelihood estimator (MLE), if F is normal, or by Logit, if F is logistic, although other alternatives exist, such as Horowitz's (1993) semiparametric estimator. It is well known that under these restrictive assumptions (i.e., which include homoscedasticity of ε_t), the MLE is consistent and asymptotically efficient. However, when the true data generating process follows (1) but is misspecified to have homoscedastic ε_t, the MLE will no longer be consistent. See for instance, Greene (2012, p. 693) and Yatchew and Griliches (1985) - the latter developed an approximation for the probit MLE bias in the presence of a simple heteroscedasticity form in a cross sectional context. We show in this paper that this misspecification will result in a positive scaling effect on the asymptotic mean of the MLE. This form of inconsistency under the general setting has not been known hitherto. The implication is that, surprisingly, the MLE - based predictions will be unaffected by the misspecification. In other words, even if conditional heteroscedasticity of a general form will be ignored and the model will be estimated by the MLE, the predictions based on the (wrong) estimator will be unaffected. This result is of importance and practical relevance because it has been widely acknowledged that the volatility of asset returns varies across time.

Our main Theorems corroborate some of the simulation results reported by Munizaga *et. al.* (2000), which revealed the remarkable robustness of the misspecified Probit and Logit model - based predictions to conditional heteroscedasticity. Moreover, we show that t-tests can be based on the MLE with reference to the standard normal distribution in spite of the misspecification.

The main results of the paper are given in the following Section. Simulations are reported in Section 3 and final remarks are provided in Section 4.

2 Main Results

By $\mathcal{F}_{t-1}$ we denote the σ-field generated by $\varepsilon_{t-1}, \varepsilon_{t-2}, \cdots$. We shall make the following assumption.

Assumption A

1. The data generating function is given by $y_t = 1\{x_t'\gamma + \varepsilon_t > 0\}$.

2. For all t and s, conditional on $\mathcal{F}_{t-1}$ and x_s, ε_t has a zero mean, conditional variance σ_t^2, $0 < \sigma_t < \infty$ and cumulative distribution function (CDF) F. The CDF F is smooth and strictly monotonic with a bounded density f which has $\mathbb{R}$ as its support and which is symmetric. In addition, $F(\nu)$ is concave for $\nu > 0$.

3. The true parameter vector γ_0 is an element of the interior of a convex parameter space $\Gamma \subset \mathbb{R}^K$.

4. The $K \times 1$ vector x_t is finite, strictly stationary and ergodic, and is not contained in any linear subspace of $\mathbb{R}^K$, $\forall t$.

5. The process $\{\sigma_t\}$ is strictly stationary and ergodic and independent of x_s, for all t and s.

We remark that Assumption A(2) holds for the normal and logistic distributions. The misspecified log-likelihood function in which conditional heteroscedasticity is ignored and σ_t is set to unity $\forall t$, is given by

$$\tilde{l}_n(\gamma) = \sum_{t=1}^{n} \tilde{l}_t(\gamma),$$

where

$$\tilde{l}_t(\gamma) = y_t \log(F^*(x_t'\gamma)) + (1 - y_t)\log(1 - F^*(x_t'\gamma)) \tag{2}$$

and F^* is the CDF of a random variable with a CDF F after it has been normalized to have mean zero and unit variance. Similarly, by f^* we denote the density corresponding to F^*. Let $\tilde{\gamma}_n = \arg\max_\Gamma \tilde{l}_n(\gamma)$. To emphasize, $\tilde{\gamma}_n$ is the maximizer of a misspecified log-likelihood function. By E_{γ_0} we denote an expectation taken under the true parameter value. The main result follows below.

Theorem 1 *Under Assumption A, there exists a finite and positive ρ which satisfies*

$$0 < \frac{1}{E_{\gamma_0}(\sigma_t)} \le \rho \le E_{\gamma_0}\left(\frac{1}{\sigma_t}\right) < \infty, \tag{3}$$

such that $\tilde{\gamma}_n \xrightarrow{p} \rho\gamma_0$.

The MLE of the correct likelihood[1], say $\hat{\gamma}_n$, is consistent. If, given $x_t = x$, the researcher wishes to base the predictive value, $\hat{y}_t$, of y_t, according to the rule $\hat{y}_t = 1\{F^*(x'\hat{\gamma}_n/\sigma_t) > 0.5\}$, then for large n, the rule is equivalent to $1\{x'\gamma_0 > 0\}$. Basing the prediction on $\tilde{\gamma}_n$ instead does not affect the result because for large n it is tantamount to

$$1\{\rho x'\gamma_0 > 0\} = 1\{x'\gamma_0 > 0\}.$$

In other words, the misspecified MLE-based prediction remains unaltered even though $\tilde{\gamma}_n$ is inconsistent. This result corroborates some of the simulation results of Munizaga *et. al.* (2000).

[1]By 'correct likelihood' it is meant that, among other things, σ_t is correctly specified. In general, it would be a function of a finite dimensional vector of parameters, as in the GARCH(p, q) process, for instance, and its parameters would have had to be estimated jointly with γ.

When the classical assumptions including homoscedasticity hold and the usual normalization, $\sigma_t = 1\ \forall t$, is imposed, it follows from (3) that $\rho = 1$, i.e., that the MLE based on the correct specification is consistent. Therefore, Theorem 1 is a generalization of the standard result.

Proof of Theorem 1: The proof can be made by verifying the conditions of Theorem 2.7 of Newey and McFadden (1994), the difference being that instead of the true parameter γ_0, we will show that the $\tilde{\gamma}_n$ converges to the value γ^* that uniquely maximizes $E_{\gamma_0}[\tilde{l}_n(\gamma)]$. To do so, we first note that for $\gamma \neq \gamma_0$,

$$E_{\gamma_0}\left(\left(x_t'\left(\gamma-\gamma_0\right)\right)^2\right) > 0.$$

As F is strictly monotonic, $x_t'\gamma \neq x_t'\gamma_0$ implies $F^*\left(x_t'\gamma\right) \neq F^*\left(x_t'\gamma_0\right)$ and that $1 - F^*\left(x_t'\gamma\right) \neq 1 - F^*\left(x_t'\gamma_0\right)$, so that $\tilde{l}_t\left(\gamma\right) \neq \tilde{l}_t\left(\gamma_0\right)$. Moreover, Assumption A(4) implies that $E_{\gamma_0}\left(\left|\tilde{l}_t\left(\gamma\right)\right|\right) < \infty$, $\forall\gamma \in \Gamma$ and thus, by Lemma 2.2 of Newey and McFadden (1994), identification is established. Secondly, Assumption A(2) on the concavity of F implies that $\log\left(F\left(\nu\right)\right)$ is also concave (see, for instance, Theorem 2.8 in De la Fuente (2000, p. 251)), which is sufficient for the concavity of $\tilde{l}_n\left(\gamma\right)$ (see, Newey and McFadden (1994), p. 2134). Third, as the process is ergodic stationary,

$$n^{-1}\tilde{l}_n(\gamma) \xrightarrow{a.s.} E_{\gamma_0}\left(\tilde{l}_t\left(\gamma\right)\right) = \tilde{l}_0(\gamma),$$

say, for all $\gamma \in \Gamma$. Therefore, the conditions of Theorem 2.7 in Newey and McFadden (1994) hold and we only need to find the unique value that maximizes $\tilde{l}_0(\gamma)$.

Under Assumption A, $\tilde{l}_0(\gamma)$ is a continuous, measurable and uniformly bounded function, so that by the Lebesgue Dominated Convergence Theorem,

$$
\begin{aligned}
&\left.\frac{\partial \tilde{l}_0(\gamma)}{\partial \gamma}\right|_{\gamma=\gamma^*} \\
&= E_{\gamma_0}\left(E_{\gamma_0}\left(\left.\left.\frac{\partial \tilde{l}_t(\gamma)}{\partial \gamma}\right|_{\gamma=\gamma^*}\right| \mathcal{F}_{t-1}, x_t = x\right)\right) \\
&= E_{\gamma_0}\left(E_{\gamma_0}\left(\left.\frac{y_t - F^*(x_t'\gamma^*)}{F^*(x_t'\gamma^*)(1-F^*(x_t'\gamma^*))} f^*(x_t'\gamma^*)x_t\right| \mathcal{F}_{t-1}, x_t = x\right)\right) \\
&= E_{\gamma_0}\left(\frac{f^*(x'\gamma^*)x}{F^*(x'\gamma^*)(1-F^*(x'\gamma^*))}\right. \qquad (4) \\
&\times \left. E_{\gamma_0}\left(\left.\left[F^*\left(\frac{x'\gamma_0}{\sigma_t}\right) - F^*(x'\gamma^*)\right]\left[1\left\{x'\gamma_0 > 0\right\} + 1\left\{x'\gamma_0 \leq 0\right\}\right]\right| x : x'\gamma_0 = a\right)\right),
\end{aligned}
$$

where the expectation in the last line is conditional on the value of x such that $x'\gamma_0 = a$, for some $a \in \mathbb{R}$. We proceed to treat the two parts of the integrals, corresponding to $1\{x'\gamma_0 > 0\}$ and $1\{x'\gamma_0 \leq 0\}$, separately. As f is assumed symmetric around zero and $F(\nu)$ is assumed concave for $\nu > 0$, it also holds that $F(1/\nu)$ is convex for $\nu > 0$ and that $F(\nu)$ is convex for $\nu < 0$ and therefore, under Assumption A,

$$
\begin{aligned}
F^*\left(\frac{a}{E_{\gamma_0}(\sigma_t)}\right) 1\{a > 0\} &\leq E_{\gamma_0}\left(\left. F^*\left(\frac{x'\gamma_0}{\sigma_t}\right) 1\{x'\gamma_0 > 0\}\right| x : x'\gamma_0 = a\right) \\
&\leq F^*\left(a E_{\gamma_0}\left(\frac{1}{\sigma_t}\right)\right) 1\{a > 0\} \qquad (5)
\end{aligned}
$$

and

$$
\begin{aligned}
F^*\left(\frac{a}{E_{\gamma_0}(\sigma_t)}\right) 1\{a < 0\} &\geq E_{\gamma_0}\left(\left. F^*\left(\frac{x'\gamma_0}{\sigma_t}\right) 1\{x'\gamma_0 < 0\}\right| x : x'\gamma_0 = a\right) \\
&\geq F^*\left(a E_{\gamma_0}\left(\frac{1}{\sigma_t}\right)\right) 1\{a < 0\}. \qquad (6)
\end{aligned}
$$

In view of (4), (5) and (6) and the fact that identification of the MLE is sufficient for the existence of a unique maximum, there exists a ρ satisfying (3) such that $\gamma^* = \rho\gamma_0$ and the theorem is proven. ■

Let

$$A_n(\gamma) = -\frac{1}{n}\frac{\partial^2 \tilde{l}_n(\gamma)}{\partial\gamma\partial\gamma'}$$

and

$$B_n(\gamma) = \frac{1}{n}\frac{\partial\tilde{l}_n(\gamma)}{\partial\gamma}\frac{\partial\tilde{l}_n(\gamma)}{\partial\gamma'}.$$

Theorem 2 *Under Assumption A and if y_t is strong mixing with mixing coefficients α_m satisfying $\sum_{m=1}^{\infty}\alpha_m^{\delta/(\delta+2)} < \infty$, for some $\delta > 0$: (i) $A_n(\tilde{\gamma}_n)$ converges in probability to a finite and nonsingular matrix*

$$A(\gamma^*) = -E_{\gamma_0}\left(\left.\frac{\partial^2\tilde{l}_t(\gamma)}{\partial\gamma\partial\gamma'}\right|_{\gamma=\gamma^*}\right).$$

(ii) The limit

$$B(\gamma^*) = E_{\gamma_0}\left(\left.\frac{1}{n}\frac{\partial\tilde{l}_n(\gamma)}{\partial\gamma}\frac{\partial\tilde{l}_n(\gamma)}{\partial\gamma'}\right|_{\gamma=\gamma^*}\right)$$

of $B_n(\gamma^*)$ exists and

$$\sqrt{n}\left.\frac{\partial\tilde{l}_n(\gamma)}{\partial\gamma}\right|_{\gamma=\gamma^*} \xrightarrow{d} N(0, B(\gamma^*)).$$

(iii) $\sqrt{n}(\tilde{\gamma}_n - \gamma^*) \xrightarrow{d} N(0, A^{-1}(\gamma^*)\, B^*(\gamma^*)\, A^{-1}(\gamma^*))$.

Proof of Theorem 2: By the Ergodic Theorem and from Example 1.2 of Newey and McFadden (1994), $A_n(\gamma)$ converges uniformly in probability to a nonstochastic matrix $A(\gamma)$, $\forall\gamma \in \Gamma$, and nonsingularity follows from Assumptions A(2) and A(4). Part (i) of the Theorem is established upon an application of Theorem 4.1.5 of Amemiya (1985) and Theorem 1 above. In view of (2),

$$\frac{\partial\tilde{l}_t(\gamma)}{\partial\gamma} = \frac{y_t - F^*(x_t'\gamma)}{F^*(x_t'\gamma)(1 - F^*(x_t'\gamma))}f^*(x_t'\gamma)x_t.$$

Under Assumptions A(2) and A(4), it is obvious that each of the elements of the vector $\left\{\partial\tilde{l}_t(\gamma)/\partial\gamma\right\}$ is a bounded sequence. The CLT conditions of

Theorem 18.5.3 of Ibragimov and Linnik (1971) are satisfied, hence part (ii) of the theorem is established. By the Mean Value Theorem,

$$0 = \frac{\partial \tilde{l}_n(\tilde{\gamma}_n)}{\partial \gamma} = \frac{\partial \tilde{l}_n(\gamma^*)}{\partial \gamma} + \frac{\partial^2 \tilde{l}_n(\bar{\gamma}_n)}{\partial \gamma \partial \gamma'} \left(\tilde{\gamma}_n - \gamma^*\right),$$

where $|\bar{\gamma}_n - \gamma^*| \leq |\tilde{\gamma}_n - \gamma^*|$. The remaining conditions of asymptotic normality of M-estimators (see, for instance, Theorem 4.1.3 of Amemiya (1985)) follow easily and so, the Theorem is proven. ■

Mikosch *et. al.* (2009, p. 63) showed that the GARCH(p, q) process is β-mixing with a geometric rate, which implies the condition stated in Theorem 2. We may apply the result of Theorem 2 to test the hypothesis $H_0 : \gamma_j = 0$ by using misspecified MLE-based t-statistic

$$t = \frac{\sqrt{n}\tilde{\gamma}_{n,j}}{\sqrt{\left(A_n(\tilde{\gamma}_{n,0})^{-1} B_n\left(\tilde{\gamma}_{n,0}\right) A_n(\tilde{\gamma}_{n,0})^{-1}\right)^{-1}_{j,j}}},$$

where $\tilde{\gamma}_{n,0}$ is the restricted MLE of γ. On an application of Theorem 2, $t \xrightarrow{d} N(0,1)$ under H_0.

3 Simulations

The goal of this section is to verify the theoretical results established in Theorem 1. In particular, the convergence and consistency properties of the probit and logit MLE when the error terms are generated by the GARCH(1,1) process. Wolfram Mathematica 10 is used in the simulations and parameter estimation. The simulated process is

$$r_t = \gamma_0 + \gamma_1 x_{1,t} + \gamma_2 x_{2,t} + \varepsilon_t,$$

$$\varepsilon_t = v_t \sigma_t, \ \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2,$$

$$y_t = 1\{r_t > 0\},$$

$$v_t \overset{iid}{\sim} N(0,1) \text{ or } Logistic(0, \frac{\sqrt{3}}{\pi}), \ \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} \overset{iid}{\sim} N\left(\begin{pmatrix} 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 1 & 0.3 \\ 0.3 & 3 \end{pmatrix}\right),$$

$$\gamma_0 = 0.3, \ \gamma_1 = 0.8, \ \gamma_2 = 0.9, \ \omega = 1, \ \alpha = 0.3, \ \beta = 0.2.$$

To avoid a possible separation problem the set of explanatory variables was built in a way that there is no dominant predictor. The process $\{\varepsilon_t\}$ is strictly stationary and ergodic, satisfying the conditions of Theorems 1 and 2 (see Bougerol and Picard (1992), Mikosch *et. al.* (2009)). In addition, the unconditional variance, σ_ε^2, is arbitrary chosen to be equal to 2. We generated 2000 samples, each with 250, 500, 1000, 2000 and 5000 observations. Tables 1- 4 summarize the misspecified probit and logit simulation results. In all cases, as n increases, the models' MLE converges to its limit and its variance decreases, as expected. As Theorem 1 suggests, the models' parameter estimates are inconsistent but converge to a positive scalar multiple of the real parameters. Tables 2 and 4 present the average probit and logit bias factor estimate $\hat{\rho}$, which, for large n, is approximately equal to 0.72 and 0.73 for the probit and logit models respectively. For the model under consideration, we obtain $\sigma_\varepsilon^{-1} = 0.707$ and $\hat{E}\left(\sigma_t^{-1}\right) \simeq 0.748$, which is consistent with the bounds given by (3).

4 Final Remarks

The results of Theorems 1 and 2 imply that for a rather general conditionally heteroscedastic process that include as a special case the prominent GARCH(p, q) process, the predictions which are based on the misspecified MLE are unaffected by the misspecification and a t-statistic for the hypothesis $H_0 : \gamma_j = 0$ can still be based on the estimator using the standard normal distribution. In particular, practitioners who wish to construct sign predictions of financial asset returns, for instance, can do so ignoring conditional heteroscedasticity.

References

Amemiya, T. (1985). *Advanced Econometrics.* Cambridge: Harvard University Press.

Anatolyev, S. and N. Gospodinov (2010). Modeling financial return dynamics via decomposition. *Journal of Business and Economic Statistics* 28, 232-245.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307-327.

Bollerslev, T., R. Y. Chou and K. F. Kroner (1992). ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics* 52, 5-59.

Bougerol, P. and N. Picard (1992). Strict stationarity of generalized autoregressive processes. *The Annals of Probability* 20, 1714-1730.

Broseta, B. (2000). Adaptive learning and equilibrium selection in experimental coordination games: An ARCH (1) approach. *Games and Economic Behavior* 32, 25-50.

Christoffersen, P. F. and F. X. Diebold (2006). Financial asset returns, direction of change, forecasting and volatility dynamics. *Management Science* 52, 1273-1287.

De La Fuente, A. (2000). *Mathematical Methods and Models for Economists.* Cambridge University Press.

Deuker, M. J. (1999). Conditional heteroskedasticity in qualitative response models of time series: A Gibbs-Sampling approach to the bank prime rate. *Journal of Business and Economic Statistics* 17, 466-472.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987-1008.

Greene, W. H. (2012). *Econometric analysis* (7th edition). NY: Prentice Hall.

Hausman, J. A. and A. W. Lo (1992). An ordered probit analysis of transaction stock prices. *Journal of Financial Economics* 31, 319-379.

Horowitz, J. (1993). Semiparametric estimation of a work-trip mode choice model. *Journal of Econometrics* 58, 49–70.

Ibragimov, I. A. and Y. V. Linnik (1971). *Independent and stationary sequences of random variables*. Wolters-Noorhoff, Groningen.

Kauppi, H. and P. Saikkonen (2008). Predicting U.S. recessions with dynamic binary response models. *Review of economics and statistics* 90, 777-791.

Lessafre, E. and H. Kaufmann (1992). Existence and uniqueness of the maximum likelihood estimator for a multivariate probit model. *Journal of the American Statistical Association* 87, 805-811.

Levich, R. M. (2001). *International Financial Markets*. McGraw-Hill/Irwin.

Leung, M. T., H. Daouk and A. S. Chen (2000). Forecasting stock indices: A comparison of classification and level estimation models. *International Journal of Forecasting* 16, 173-190.

Mikosch, T., J. P. Kreiß, R. A. Davis and G. A. Torben (2009). *Handbook of Financial Time Series*. Springer Berlin Heidelberg.

Munizaga, M. A., B. G. Heydecker and J. D. Ortúzar (2000). Representation of heteroskedasticity in discrete choice models. *Transportation Research Part B* 34, 219-240.

Newey, W. K and D. McFadden (1994). Large sample estimation and hypothesis testing. In R.F. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Vol. 4, North Holand.

Nyberg, H. (2011). Forecasting the direction of the U.S. stock market with dynamic binary probit models. *International Journal of Forecasting* 27, 561-578.

Rydberg, T. and N. Shephard (2003). Dynamics of trade-by-trade movements: Decomposition and models. *Journal of Financial Econometrics* 1, 2-25.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* 53, 1–16.

White, H. (2001). *Asymptotic Theory for Econometricians.* New York: Academic Press.

Yatchew, A. and Z. Griliches (1985). Specification error in probit models. *The Review of Economics and Statistics* 67, 134-139.

Table 1. Misspecified Probit model simulation results			
n	$\tilde{\gamma}_0$	$\tilde{\gamma}_1$	$\tilde{\gamma}_2$
250	0.222 (0.043)	0.597 (0.016)	0.67 (0.01)
500	0.219 (0.02)	0.585 (0.007)	0.658 (0.005)
1000	0.219 (0.009)	0.583 (0.004)	0.652 (0.002)
2000	0.218 (0.005)	0.576 (0.002)	0.65 (0.001)
5000	0.215 (0.002)	0.576 (0.0007)	0.649 (0.0004)

Note: the values represent the mean and variance (in brackets) of the misspecified probit MLE.

Table 2. Misspecified Probit MLE bias factor estimate ($\hat{\rho}$)					
	$n = 250$	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$
$\tilde{\gamma}_0/0.3$	0.739	0.731	0.731	0.727	0.718
$\tilde{\gamma}_1/0.8$	0.746	0.732	0.723	0.720	0.720
$\tilde{\gamma}_2/0.9$	0.744	0.731	0.725	0.723	0.721

Note: the values represent the probit MLE estimates from Table 1 divided by the real parameter values.

Table 3. Misspecified Logit model simulation results			
n	$\tilde{\gamma}_0$	$\tilde{\gamma}_1$	$\tilde{\gamma}_2$
250	0.229 (0.042)	0.605 (0.017)	0.68 (0.011)
500	0.22 (0.02)	0.592 (0.008)	0.667 (0.005)
1000	0.218 (0.009)	0.593 (0.004)	0.664 (0.002)
2000	0.219 (0.005)	0.587 (0.002)	0.661 (0.001)
5000	0.22 (0.002)	0.584 (0.0008)	0.659 (0.0005)

Note: the values represent the mean and variance (in brackets) of the misspecified Logit MLE.

Table 4. Misspecified Logit MLE bias estimate ($\hat{\rho}$)					
	$n = 250$	$n = 500$	$n = 1000$	$n = 2000$	$n = 5000$
$\tilde{\gamma}_0/0.3$	0.764	0.734	0.726	0.729	0.732
$\tilde{\gamma}_1/0.8$	0.758	0.74	0.741	0.734	0.731
$\tilde{\gamma}_2/0.9$	0.756	0.741	0.738	0.735	0.732

Note: the values represent the Logit MLE estimates from Table 3 divided by the real parameter values.