Journal of Business & Economic Statistics

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ubes20

Constrained Regression for Interval-valued Data

Gloria González-Rivera [a] & Wei Lin [a]

[a] Department of Economics , University of California , Riverside, Riverside , CA , 92521

Accepted author version posted online: 12 Jul 2013.

To cite this article: Journal of Business & Economic Statistics (2013): Constrained Regression for Interval-valued Data, Journal of Business & Economic Statistics, DOI: 10.1080/07350015.2013.818004

To link to this article: http://dx.doi.org/10.1080/07350015.2013.818004

Constrained Regression for Interval-valued Data

Gloria González-Rivera *
Wei Lin

Department of Economics
University of California, Riverside
Riverside, CA 92521

Abstract

Current regression models for interval-valued data do not guarantee that the predicted lower bound of the interval is always smaller than its upper bound. We propose a constrained regression model that preserves the natural order of the interval in all instances, either for in-sample fitted intervals or for interval forecasts. Within the framework of interval time series, we specify a general dynamic bivariate system for the upper and lower bounds of the intervals. By imposing the order of the interval bounds into the model, the bivariate probability density function of the errors becomes conditionally truncated. In this context, the OLS estimators of the parameters of the system are inconsistent. Estimation by maximum likelihood is possible but it is computationally burdensome due to the nonlinearity of the estimator when there is truncation. We propose a two-step procedure that combines maximum likelihood and least squares estimation, and a modified two-step procedure that combines maximum likelihood and minimum-distance estimation. In both instances, the estimators are consistent. However, when multicollinearity arises in the second step of the estimation, the modified two-step procedure is superior at identifying the model regardless of the severity of the truncation. Monte Carlo simulations show good finite sample properties of the proposed estimators. A comparison with the current methods in the literature shows that our proposed methods are superior by delivering smaller losses and better estimators (no bias and low mean squared errors) than those from competing approaches.We illustrate our approach with the daily interval of low/high SP500 returns and find that truncation is very severe during and after the financial crisis of 2008, so that OLS estimates should not be trusted and a modified two-step procedure should be implemented.

Key Words: Interval-valued Data, Inverse of the Mill's Ratio, Maximum Likelihood Estimation, Minimum Distance Estimator, Truncated Probability Density Function.

JEL Classification: C01, C32, C34.

*We would like to thank the participants in the International Symposium in Forecasting in Prague, the 2012 Joint Meetings of the American Statistical Association in San Diego, and the 2012 ICSA Applied Statistics Symposium in Boston. We are also grateful to the editor, Jonathan Wright, associate editor, and two anonymous referees whose constructive comments have greatly improved the manuscript. The usual caveat applies. The first author appreciates financial support from the UCR Academic Senate grants.

1 Introduction

With the advent of sophisticated information systems, data collection has become less costly and, as a result, massive data sets have been generated in many disciplines. Economics and business are not exceptions. For instance, financial data is available at very high frequencies for almost every asset that is transacted in a public market providing data sets with millions of observations. Marketing data sets offer high granularity about consumers and products characteristics. Environmental stations produce data sets that contain high and low frequency records of temperatures, atmospheric conditions, pollutants, etc. across many regions. Statistical institutes, like the Census Bureau, collect socioeconomic information about all individuals in a nation. These massive information data sets tend to be released in an aggregated format, either because of confidentiality reasons or because the interest of study is not the individual unit but a collective of units. In these cases, the researcher does not face classical data sets, i.e. $\{y_i\}$ for $i = 1, \cdots, n$ or $\{y_t\}$ for $t = 1, \cdots T$ where y_i or y_t are single values in the real line, but data aggregated in some fashion, like interval data $[y_l, y_u]$ that offers information on the lower and upper bound of the variable of interest. For example, information about income or net worth comes very often in interval format, or low and high prices of an asset in a given day, or daily temperature intervals, or low/high prices of electronic devices for several stores, etc.

Interval-valued data are also considered symbolic data sets. Within the symbolic approach (Billard and Diday, 2003, 2006), there are several proposals to fit a regression model to interval data. For a review, see Arroyo, González-Rivera and Maté (2011). The simplest approach (Billard and Diday, 2000) is to fit a regression model to the centers of the intervals of the dependent variable and of the regressors. Further approaches consider two separate regressions, one for the lower bound and another for the upper bound of the intervals, either with no constraints in the regression coefficients (Billard and Diday, 2002) or by constraining both regressions to share the same regression coefficients (Brito, 2007). In a similar line, Lima Neto and de Carvalho (2008) propose running two different regressions, one for the center and another for the range of the intervals, with

ACCEPTED MANUSCRIPT

no constraints. None of these approaches guarantees that the fitted values from the regressions will satisfy the natural order of an interval, i.e. $\hat{y}_l \leq \hat{y}_u$, for all observations. Lima Neto and de Carvalho (2010) impose non-negative constraints on the regression coefficients of the model for the range and solve a quadratic programming problem to find the least squares solution. However, for these constraints to be effective, the range regression must entertain only non-negative regressors (e.g., regressing the range of the dependent variable on the ranges of the regressors), which limit the usefulness of the model.

In this paper we propose a regression model, either for cross-sectional or time series data, that guarantees the natural order of the fitted interval bounds for all the observations in the sample, and for any potential interval forecast based on the model. Within the framework of interval time series (ITS), we specify a bivariate system for the lower and upper bounds of the time series. The observability restriction $y_{l,t} \leq y_{u,t}$ implies that the conditional probability density function of the errors is truncated. Under the assumption of bivariate normal errors, the amount of truncation will depend on the variance-covariance matrix of the errors and it will be time-varying because the truncation is a function of the difference between the conditional means of the lower and upper bounds. When the observability restriction is severe, i.e., the truncation of the bivariate density is substantial, not only the conditional expectations of the errors are different from zero but also the errors are correlated with the regressors, thus any least-squares estimation (linear or non linear) will fail to deliver consistent estimators of the parameters of the model. We propose a two-step estimation procedure, combining maximum likelihood and least squares estimation, that will deliver consistent estimators. The first step consists of modeling the range of the interval, which is distributed as a truncated normal density, to obtain maximum likelihood estimates of the inverse of the Mill's ratio $\hat{\lambda}_{t-1}$, which embodies the severity of the restriction. Only when the restriction is severe, the second step is necessary. This step consists of introducing $\hat{\lambda}_{t-1}$ in a least-squares regression to correct the selection bias imposed by the restriction. However, the estimation in the second step may be plagued with multicollinearity problems because in some instances $\hat{\lambda}_{t-1}$ is an almost linear function of the regressors. Since multicollinearity cannot be resolved by dropping some of the regressors, we propose a modified second step by implementing a minimum distance estimator that delivers consistent estimates of all parameters in the model. The advantage of the modified second step is that even when the observability restriction is not severe ($\hat{\lambda}_t \approx 0$ for most t), we are able to identify all parameters without much loss in efficiency.

As an illustration of the methods that we propose, we model the interval of daily low/high returns to the SP500 index before and after 2007. Before 2007, the daily interval exhibits very little volatility, but after 2007, volatility is the dominant characteristic due to the events of the financial crisis of 2008. These two periods have very different dynamics. We implement the modified two-step estimator and we find that in the stable period the observability restriction is not severe, so that simple OLS will suffice to estimate a dynamic system for the lower and upper bounds of the interval. In contrast, in the high volatility period the restriction is very severe, thus simple OLS estimates should not be trusted and the second step is necessary to guarantee the consistency of the estimators.

The modeling of the low/high interval is interesting in itself for several reasons. For instance, in technical analysis, trading strategies are based on the dynamics of an object, the ”candlestick”, which is composed of two intervals, the low/high and the open/close. In financial econometrics, the low/high interval also provides estimators of the volatility of asset returns, see Parkinson (1980), Yang and Zhang (2000), Alizadeh, Brandt, and Diebold (2002) among others. However, the most important reason for our interest in estimation and forecasting with interval-valued data lies on the fact that the only format available for some data sets is the interval format. Financial data sets are exceptional; they are very rich and information come in many formats, e.g., databases contain records of prices for every transaction in the market so that we could analyze prices at the highest and the lowest frequencies; there is an almost continuous measurement in the transaction price. But this is not always the case in other areas within economics or in other sciences. Some examples follow.

The US Energy Information Administration gather electricity prices for each state in US. Since there are so many factors affecting the prices of electricity, there is substantial variation across states and across localities in the same state. This agency provides average retail price at the state level in interval format, i.e. min/max price, which is more informative of the realities of this market. The US Department of Agriculture provides livestock prices also in interval format. The Livestock Marketing Information Center (Iowa State University) reports interval prices of several items, for instance, min/max daily beef prices. Though they compute a weighted price, this is not the price of a given transaction, so that the interval min/max contains more valuable information to the participants in the market. In the appraisal industry, the objective is to find a ”fair market price” for items, such as real estate, for which the market value cannot be observed directly unless the item is sold. It is standard practice in this industry to record min/max prices of similar items that have had a recent transaction so that the ”fair” market price, though non-observable, must be contained within such an interval. Even with financial datasets, it is interesting to note that bond market data is not as transparent as stock data and bond traders report the bid/ask interval of the transaction, in which the price is contained. In other fields different form economics, for instance medicine, we have databases with patient data recorded in interval format, the most indicative is blood pressure measurements i.e., diastolic and systolic pressure (low and high numbers respectively). In earth sciences, temperature records across locations also come in interval format, i.e. min/max temperature for a given location.

These examples show that the low/high interval of a variable is a common format that provides additional information beyond an average measurement, and in some cases, it is the only format available to the researcher. It should be noted that estimating and forecasting with low/high interval-valued data is different from estimating and forecasting two quantiles. The low/high bounds are extremes. In quantile regression, the loss function requires fixing the probability α associated with the quantile. If we wish to approximate the low/high interval with quantile regression, it seems natural to fix $\alpha = 0$ for estimation of the lower bound and $\alpha = 1$ for the estimation of the upper bound, but if the variables of analysis are defined in the domain $(-\infty, +\infty)$, these are also the values of the corresponding $(0, 1)$ quantiles. If our interest is any other quantile, e.g. the interquartile range $[Q_{0.25}, Q_{0.75}]$, and the data is available in a classic point-valued format, then quantile regression with monotonicity restrictions could be implemented as proposed by Chernozhukov *et al.* (2010).

We organize the paper as follows. In section 2, we provide the general framework and basic assumptions. In section 3, we present the two-step estimation procedure and develop its asymptotic properties. In section 4, we conduct extensive Monte Carlo simulations that show the finite sample properties of the two-step and modified two-step estimators. In section 5, we compare extensively our methods with those existing in the literature. In section 6, we illustrate the empirical aspects of our methods with the daily interval of low/high SP500 returns. In section 7, we conclude.

2 General Framework and Basic Assumptions

We introduce a general regression framework for interval-valued time series. The objective is the estimation of a parametric specification of the conditional mean of an interval-valued stochastic process. Generally, an interval is defined as follows:

Definition 1. *An interval* $[Y]$ *over a set* $(R, \leq)$ *is an ordered pair* $[Y_l, Y_u]$ *where* $Y_l, Y_u \in R$ *are the lower and upper bounds of the interval such that* $Y_l \leq Y_u$.

We can also define an interval random variable on a probability space (Ω, F, P) as the mapping $Y : F \rightarrow [Y_l, Y_u] \subset R$. In a time series framework, we further define an interval-valued stochastic process as a collection of interval random variables indexed by time, i.e. $\{Y_t\}$ for $t \in T$; and an interval-valued time series (ITS) as a realization $\{[y_{lt}, y_{ut}]\}_{t=1}^{T}$ of an interval-valued stochastic process.

We are interested in modeling the dynamics of the process $\{Y_t\} = \{[Y_{lt}, Y_{ut}]\}$ as a function of an information set that potentially includes not only the past history of the process, i.e. $Y^{t-1} = (Y_{t-1}, Y_{t-2}, \cdots, Y_0)$ but also any other exogenous random variables $X^t = (X_t, X_{t-1}, \cdots, X_0)$ where

ACCEPTED MANUSCRIPT

$X_t = (X_{1t}, X_{2t}, ..., X_{pt})$. In this context, we focus the modeling exercise on establishing the joint dynamics of the lower $\{Y_{lt}\}$ and upper $\{Y_{ut}\}$ bounds taking into account the natural ordering of the interval. Thus, a general data generating process is written as

$$Y_t \equiv \begin{bmatrix} Y_{lt} \\ Y_{ut} \end{bmatrix} = \begin{bmatrix} G_l(Y^{t-1}, X^t; \beta_l) \\ G_u(Y^{t-1}, X^t; \beta_u) \end{bmatrix} + \begin{bmatrix} \varepsilon_{lt} \\ \varepsilon_{ut} \end{bmatrix}, \quad \text{such that } Y_{lt} \leq Y_{ut} \tag{2.1}$$

where $G_l(\cdot)$, $G_u(\cdot)$ are differentiable functions, β_l , β_u are two $J \times 1$ parameter vectors, and $\epsilon_t \equiv$

$(\varepsilon_{lt}, \varepsilon_{ut})'$ is the error vector. The observability restriction $Y_{lt} \leq Y_{ut}$ will be imposed on the process.

The observability restriction in (2.1) is the key feature of the specification because it generates two important issues for the estimation of the model (2.1). First, the restriction $Y_{lt} \leq Y_{ut}$ implies a restriction on the distribution of the error vector. The errors now are restricted as follows,

$$\begin{aligned} G_l(Y^{t-1}, X^t; \beta_l) + \varepsilon_{lt} &\leq G_u(Y^{t-1}, X^t; \beta_u) + \varepsilon_{ut}, \\ \varepsilon_{ut} - \varepsilon_{lt} &\geq G_l(Y^{t-1}, X^t; \beta_l) - G_u(Y^{t-1}, X^t; \beta_u). \end{aligned} \tag{2.2}$$

The transformed observability restriction (2.2) implies that, conditioning on the information set $\Im_{t-1} \equiv (Y^{t-1}, X^t)$, the joint distribution of $(\varepsilon_{lt}, \varepsilon_{ut})$ is truncated from below. Figure 1 illustrates a truncated joint density of the errors. In the plane formed by the variables $(\varepsilon_{lt}, \varepsilon_{ut})$, the ellipse represents a contour of the joint density, and the 45° degree line $\varepsilon_{ut} = \varepsilon_{lt} + (G_l - G_u)$ is the truncation line, separating the shaded area, where $Y_{lt} \leq Y_{ut}$ holds, from the area where the restriction is violated.

[FIGURE 1]

From Figure 1, we observe that the feasible support for the errors will depend on the error variance-covariance matrix as well as any other parameters affecting the shape of the contours, and on the position of the truncation line, which is a function of the difference between the two conditional mean functions. Small dispersion of the errors together with a large difference, i.e. $G_l << G_u$ tend to mitigate the severity of the observability restriction because it reduces the probability of the errors falling below the truncation line to the point that the restriction might not longer be binding and it could be safely removed from the model. However, if the restriction is binding, it cannot be ignored in the model estimation because, on one hand, it may generated predicted values of Y_{lt} and Y_{ut} that do not follow the natural order of an interval, and on the other, it will affect the asymptotic properties of the estimators as we see next. By taking conditional expectations with respect to $\Im_{t-1}$ in (2.1),

$$E_{t-1}(Y_{lt}|Y_{lt} \leq Y_{ut}) = G_l(Y^{t-1}, X^t; \beta_l) + E_{t-1}(\varepsilon_{lt}|\varepsilon_{ut} - \varepsilon_{lt} \geq G_l - G_u),$$

$$E_{t-1}(Y_{ut}|Y_{lt} \leq Y_{ut}) = G_u(Y^{t-1}, X^t; \beta_u) + E_{t-1}(\varepsilon_{ut}|\varepsilon_{ut} - \varepsilon_{lt} \geq G_l - G_u).$$

When the observability restriction is binding, the conditional expectations of the errors, which are $E_{t-1}(\varepsilon_{lt}|\varepsilon_{ut} - \varepsilon_{lt} \geq G_l - G_u)$ and $E_{t-1}(\varepsilon_{ut}|\varepsilon_{ut} - \varepsilon_{lt} \geq G_l - G_u)$, will not be zero and furthermore, they will depend on the regressors of the model through the functions $G_l(\cdot)$ and $G_u(\cdot)$. Thus, any least-squares estimation (linear or nonlinear) will fail to deliver consistent estimators for the model.

Before introducing our estimation procedures, we need to state some basic assumptions on (2.1).

Assumption 1. *(Weak Stationarity) The interval-valued stochastic process* $\{Y_t\} = \{Y_{lt}, Y_{ut}\}$ *is covariance stationary, which means that the lower* $\{Y_{lt}\}$ *and upper* $\{Y_{ut}\}$ *processes are themselves covariance-stationary. We also require covariance stationarity in the regressors* $X_t \equiv (X_{1t}, ..., X_{pt})'$.

This assumption allows estimators with standard asymptotic properties. The proposed methods will also apply to non-stationary data but the properties of the estimators will be non-standard.

Assumption 2. *(Exogeneity) The regressors* (Y^{t-1}, X^t) *are strictly exogenous variables i.e.,* $E(\epsilon_t|Y^{t-1}, X^t) = 0$

This assumption is standard in regression analysis to protect the estimators against endogeneity bias. In our context, the objective is to analyze the dynamics of $\{Y_{ut}\}$ and $\{Y_{lt}\}$ as a system. For instance, in a VAR system, the right hand side of the system will have lags of $\{Y_{ut}\}$ and $\{Y_{lt}\}$. If we were to introduce additional regressors X_t, we could proceed in several ways, either expanding the VAR system to include X_t as another element of the system, or considering only predetermined regressors, i.e. $X_{t-1}, X_{t-2}, ...$, or requiring the weak exogeneity of X_t. By proceeding in either way, we will focus exclusively on the endogeneity generated by the binding observability restriction, that is, when $E_{t-1}(\varepsilon_{lt} | \varepsilon_{ut} - \varepsilon_{lt} \geq G_l - G_u)$ and $E_{t-1}(\varepsilon_{ut} | \varepsilon_{ut} - \varepsilon_{lt} \geq G_l - G_u)$, are not zero.

Assumption 3. *(Conditional Independence)* $(X_T, ..., X_{t+1})$ *and* Y^t *are conditional independent given* X^t*, i.e.* $(X_T, ..., X_{t+1}) \perp Y^t | X^t$.

This assumption relates to the previous one in the sense that it opens the system of $\{Y_{ut}\}$ and $\{Y_{lt}\}$ to the effect of other regressors which are not explicitly modeled within the system. For instance, in a VAR framework, if we were to model jointly $\{Y_{ut}\}$, $\{Y_{lt}\}$, and X_t, this assumption will not be needed. But because we focus only on the dynamics of $\{Y_{ut}\}$ and $\{Y_{lt}\}$, we need to assume that Y_t does not Granger-cause X_t to avoid biased and potentially inconsistent estimators.

Assumption 4. *(Normality) The error terms* $\epsilon_t \equiv (\varepsilon_{lt}, \varepsilon_{ut})$ *are* i.i.d. *bivariate normal random variables with joint density* $f(\epsilon_t) = (2\pi)^{-1} |\Sigma|^{-1/2} \exp\{-\epsilon_t' \Sigma^{-1} \epsilon_t / 2\}$ *with the* 2×2 *variance-covariance matrix* $\Sigma = [\sigma_l^2 \quad \rho\sigma_l\sigma_u \,;\, \rho\sigma_l\sigma_u \quad \sigma_u^2]$.

This assumption may seem restrictive but it provides at least a quasi-maximum likelihood approach to the estimation of $\{Y_{ut}\}$ and $\{Y_{lt}\}$. If the observability restriction is not binding, estimation by maximum likelihood under normality or by least squares, will produce consistent but inefficient estimators. If heteroscedasticity is present, the estimators are still consistent but we would need to implement a heteroscedasticity-consistent estimator of the variance for a correct inference. If we were to assume any other density, and again running the risk of a false assumption, we would not be sure whether QMLE results hold (Newey and Steigerwald, 1997). If the observability restriction is binding, bivariate normality implies that the distribution of the errors is conditionally truncated normal with conditional heteroscedasticity. Our estimation procedures take care of the heteroscedasticity, and since we are modeling extremes, low and high, the density of these variables cannot be symmetric, thus the truncation takes care of the asymmetry. Furthermore, the simulations presented in Sections 4 and 5 show that our estimators are very robust to misspecification of the density when there are relevant dynamics in the conditional means of $\{Y_{ut}\}$ and $\{Y_{lt}\}$.

The potential misspecification of the regressor λ_{t-1} seems to affect mainly the estimation of the constant but we will show that the estimation of the system generates good fitted intervals with substantially smaller losses than other competing methods.

3 Estimation

Given the implications of the observability restriction for a least squares estimator of the parameters in (2.1), it is natural to think that a full information estimator, like maximum likelihood (ML), will be better suited to guarantee consistency. In this section, we will introduce the conditional log-likelihood function of a sample y^T in order to underline the contribution of the restriction to the estimation. However, our main objective is to develop a two-step estimation procedure that delivers consistent estimators but it is easier to implement and it overcomes some of the limitations of the ML estimator.

3.1 Conditional log-likelihood function

For a sample of size T, $y^T \equiv (y_T, \cdots, y_1)$ and $x^T \equiv (x_T, \cdots, x_1)$, and for a fixed initial value y^0, let $f_Y\left(y^T \left| x^T; \theta\right.\right)$ be the joint conditional density of y^T, where $\theta \in \Theta$ is an open subset of R^K. The conditional likelihood $\ell(y^T, \theta)$ of y^T is $f_{Y^T}(y^T|x^T;\theta)$ if $y_{lt} \leq y_{ut}$ and 0 otherwise. It follows that

$$\ell(\mathbf{y}, \theta) = f_{Y^T}(y^T | y_l^T \leq y_u^T, x^T; \theta) \times \Pr(y_l^T \leq y_u^T | x^T; \theta) = \prod_{t=1}^{T} \frac{f_{Y_t}(y_t | y^{t-1}, x^t; \theta)}{\Pr(y_{lt} \leq y_{ut} | y^{t-1}, x^t; \theta)} \tag{3.1}$$

where f_{Y_t} is the density of Y_t conditional on the information (Y^{t-1}, X^t). In (3.1), we have also called assumption 3. Under assumption 4, the conditional log-likelihood function of a sample y^T is

$$\begin{aligned} L(y^T, \theta) &= -\log 2\pi - \frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2T}\sum_{t=1}^{T}\left[y_t - G(y^{t-1}, x^t; \beta)\right]' \mathbf{\Sigma}^{-1}\left[y_t - G(y^{t-1}, x^t; \beta)\right] \\ &\quad -\frac{1}{T}\sum_{t=1}^{T}\log R_t(y^{t-1}, x^t; \theta). \end{aligned} \tag{3.2}$$

where $\theta = (\beta, \Sigma)$; and $R_t(y^{t-1}, x^t; \theta)$ is defined as

$$R_t(y^{t-1}, x^t; \theta) \equiv \Pr(y_{lt} \leq y_{ut} | y^{t-1}, x^t; \theta) = 1 - \Phi\left(\frac{G_l(y^{t-1}, x^t; \beta_l) - G_u(y^{t-1}, x^t; \beta_u)}{\sqrt{\sigma_u^2 + \sigma_l^2 - 2\rho\sigma_u\sigma_l}}\right)$$

in which $\Phi(\cdot)$ is the standard normal cumulative distribution function.

The maximum likelihood estimator $\widehat{\theta}_{ML}$ is the maximizer of (3.2). This estimator will be highly nonlinear, even for a linear system as in (2.1), because of the contribution of the observability restriction term $R_t(y^{t-1}, x^t; \theta)$ to the log-likelihood function. $R_t(y^{t-1}, x^t; \theta)$ provides the probability mass that is left in the joint density after the truncation takes place. It is easily seen that $0 \leq R_t(y^{t-1}, x^t; \theta) \leq 1$. If the restriction is not binding, $R_t(y^{t-1}, x^t; \theta) = 1$ for all t, and its contribution to the log-likelihood function is zero.[1] In this case the restriction is redundant and it can be removed from the specification of the model. On the other hand, if the observability restriction is binding i.e. $R_t(y^{t-1}, x^t; \theta) < 1$ for some t, it must be taken into account in the estimation of the model. Ignoring the restriction will result in the inconsistency of ML estimator. In theory, the ML estimator has obvious advantage. If the true distribution of ε_t is normal as in assumption 4, under certain regularity conditions, the ML estimator $\widehat{\theta}_{ML}$ is consistent and asymptotically normal.[2] However in practice, given the nonlinearity of the ML estimator induced by the observability restriction, we should expect multiple local maxima in the log-likelihood function leading to multiple solutions and non-trivial convergence problems in the maximization algorithm. Thus, the consistency of ML estimator will depend on a good guess of the initial value of the parameters. For these reasons, we propose a two-step procedure that combines maximum likelihood and least squares estimation, that it is easy to implement and will deliver consistent estimators of the parameters of the model.

[1] A sufficient and necessary condition for a non-binding restriction is $\frac{G_l(y^{t-1}, x^t; \beta_l) - G_u(y^{t-1}, x^t; \beta_u)}{\sqrt{\sigma_u^2 + \sigma_l^2 - 2\rho\sigma_u\sigma_l}} \ll 0$ for all t.

[2] Regularity conditions that guarantee the consistency and asymptotic normality of ML estimator $\widehat{\theta}_{ML}$ are in Amemiya (1985, Theorems 4.1.1 and 4.1.3) and in White (1994,Theorem 4.6).

3.2 Two-Step Estimation: General Remarks

Given the popularity of VAR models, we will consider the process (2.1) to follow a linear autoregressive specification of order p. However, the two-step procedure to be described next will be also applicable to nonlinear models by properly choosing a nonlinear estimation technique in the second step.

The interval autoregressive model, IAR(p), is described as follows

$$\begin{bmatrix} y_{l,t} \\ y_{u,t} \end{bmatrix} = \begin{bmatrix} \beta_{lc} \\ \beta_{uc} \end{bmatrix} + \sum_{j=1}^{p} \begin{bmatrix} \beta_{11}^{(j)} & \beta_{12}^{(j)} \\ \beta_{21}^{(j)} & \beta_{22}^{(j)} \end{bmatrix} \begin{bmatrix} y_{l,t-j} \\ y_{u,t-j} \end{bmatrix} + \begin{bmatrix} \varepsilon_{lt} \\ \varepsilon_{ut} \end{bmatrix}$$

with observability restriction $y_{lt} \leq y_{ut}$, and an error term ϵ_t that is bivariate normal *i.i.d.*. Conditioning on the information set $\Im_{t-1} = (y_{t-1}, ..., y_{t-p}, ...)$, the conditional mean of the IAR(p) process is

$$E_{t-1}(y_{lt}|y_{ut} \geq y_{lt}) = \beta_{lc} + \sum_{j=1}^{p} \beta_{11}^{(j)} y_{l,t-j} + \sum_{j=1}^{p} \beta_{12}^{(j)} y_{u,t-j} + E_{t-1}(\varepsilon_{lt}|y_{ut} \geq y_{lt})$$

$$E_{t-1}(y_{ut}|y_{ut} \geq y_{lt}) = \beta_{uc} + \sum_{j=1}^{p} \beta_{21}^{(j)} y_{l,t-j} + \sum_{j=1}^{p} \beta_{22}^{(j)} y_{u,t-j} + E_{t-1}(\varepsilon_{ut}|y_{ut} \geq y_{lt})$$

Under the normality assumption 4, we derive the conditional expectation of the errors (see web appendix), which are $E_{t-1}(\varepsilon_{lt}|y_{ut} \geq y_{lt}) = C_l \lambda_{t-1}$ and $E_{t-1}(\varepsilon_{ut}|y_{ut} \geq y_{lt}) = C_u \lambda_{t-1}$, where $C_l = -(\sigma_l^2 - \rho\sigma_u\sigma_l)/\sigma_m$, $C_u = (\sigma_u^2 - \rho\sigma_u\sigma_l)/\sigma_m$, $\sigma_m^2 = \sigma_u^2 + \sigma_l^2 - 2\rho\sigma_l\sigma_u$, and

$$\lambda_{t-1} = \frac{\phi(\Delta(y^{t-1}, \Delta\beta)/\sigma_m)}{1 - \Phi(\Delta(y^{t-1}, \Delta\beta)/\sigma_m)} \tag{3.3}$$

$$\Delta(y^{t-1}, \Delta\beta) \equiv G_l - G_u = \Delta\beta_c + \sum_{j=1}^{p} \Delta\beta_1^{(j)} y_{l,t-j} + \sum_{j=1}^{p} \Delta\beta_2^{(j)} y_{u,t-j}$$

$$\Delta\beta = (\beta_{lc} - \beta_{uc}, \beta_{11}^{(1)} - \beta_{21}^{(1)}, \beta_{12}^{(1)} - \beta_{22}^{(1)}, \cdots, \beta_{11}^{(p)} - \beta_{21}^{(p)}, \beta_{12}^{(p)} - \beta_{22}^{(p)}). \tag{3.4}$$

Therefore, the regression models can be explicitly written as

$$y_{lt} = \beta_{lc} + \sum_{j=1}^{p} \beta_{11}^{(j)} y_{l,t-j} + \sum_{j=1}^{p} \beta_{12}^{(j)} y_{u,t-j} + C_l \lambda_{t-1} + v_{lt} \tag{3.5}$$

$$y_{ut} = \beta_{uc} + \sum_{j=1}^{p} \beta_{21}^{(j)} y_{l,t-j} + \sum_{j=1}^{p} \beta_{22}^{(j)} y_{u,t-j} + C_u \lambda_{t-1} + v_{ut} \tag{3.6}$$

where now $v_{lt} = \varepsilon_{lt} - C_l \lambda_{t-1}$ and $v_{ut} = \varepsilon_{ut} - C_u \lambda_{t-1}$ are martingale difference sequences with respect to $\Im_{t-1}$, i.e. $E_{t-1}(v_{lt}|y_{ut} \geq y_{lt}) = 0$ and $E_{t-1}(v_{ut}|y_{ut} \geq y_{lt}) = 0$.

Two remarks are in order. First, since $C_u - C_l = \sigma_m$, we need $\sigma_m > 0$ to be strictly positive for C_u and C_l to be well defined. This implies that the specific case $\sigma_u^2 = \sigma_l^2$ and $\rho = 1$ must be ruled out. This could happen when the interval $[\varepsilon_{lt}, \varepsilon_{ut}]$ is degenerate and collapses to a single value. Secondly, λ_{t-1} is the inverse of the Mill's ratio and embodies the severity of the observability restriction. When the restriction is non-binding $R_t(y^{t-1}, x^t; \theta) = 1$ for all t, which implies that $\lambda_{t-1} = 0$ for all t.

Based on regressions (3.5) and (3.6), the two-step estimation strategy consists of estimating λ_{t-1} first and assessing how binding the observability restriction is. The second step is only meaningful when the restriction is binding. In this case, we proceed to plug in $\hat{\lambda}_{t-1}$ in (3.5) and (3.6) and perform least squares. The proposed two-step estimation strategy resembles Heckman's (1979) two-step procedure for sample selection models. However, there are important conceptual differences. In Heckman's, the selection mechanism (the first step) includes the full sample of observations, e.g. women who participate and who do not in the labor market, and the regression model (the second step) includes a partial sample, those for which the dependent variable of interest is observed, e.g. the wage of those women who work. In our strategy, we carry the same sample in both steps because those observations that violate the observability restriction will never be observed. Hence, from the start, our first step will focus on a truncated normal regression that arises very naturally when we model the range of the interval, and from which we will estimate λ_{t-1}. Our second step is analogous to Heckman's in that the objective is to correct the selection bias of the least squares estimator in the regression of interest. However, Heckman's bias is inconsequential when the error terms of the selection equation and of the regression of interest are uncorrelated. In our second step, even if the errors of the lower and upper bound regressions are uncorrelated, the inconsistency of the least squares estimator will remain when the observability restriction is binding and is omitted in the second-step regression.

ACCEPTED MANUSCRIPT

3.3 Two-step Estimation: The First Step

Our objective is to estimate λ_{t-1}. To this end, we model the range of the interval $\Delta y_t = y_{ut} - y_{lt}$, which according to the IAR(p) model will exhibit the following dynamics

$$y_{ut} - y_{lt} = -\left(\Delta\beta_c + \sum_{j=1}^{p} \Delta\beta_1^{(j)} y_{l,t-j} + \sum_{j=1}^{p} \Delta\beta_2^{(j)} y_{u,t-j}\right) + \Delta\varepsilon_t \tag{3.7}$$

Under normality assumption 4, and imposing the observability restriction, the difference of the two error terms, $\Delta\varepsilon_t$, follows a truncated normal distribution. Thus, the conditional density of Δy_t is,

$$f(\Delta y_t|\Delta y_t \geq 0, y^{t-1}; \Delta\beta, \sigma_m) = \frac{1}{\sigma_m} \frac{\phi(\Delta y_t/\sigma_m + \Delta(y^{t-1}, \Delta\beta)/\sigma_m)}{1 - \Phi(\Delta(y^{t-1}, \Delta\beta)/\sigma_m)}. \tag{3.8}$$

Based on (3.8), we can construct the log-likelihood function of a sample of T observations $\Delta\mathbf{y}$

$$T^{-1}L(\Delta\mathbf{y}; \Delta\beta, \sigma_m) = \frac{1}{T}\sum_{t=1}^{T} \log f(\Delta y_t|\Delta y_t \geq 0, y^{t-1}; \Delta\beta, \sigma_m) \tag{3.9}$$

to obtain the maximum likelihood estimators $\widehat{\Delta\beta}$ and $\widehat{\sigma}_m$ as the $\arg\max_{\Delta\beta,\sigma_m}[T^{-1}L(\Delta\mathbf{y}; \Delta\beta, \sigma_m)]$. The ML estimators will be plugged in (3.3) to finally obtain $\widehat{\lambda}_{t-1}$.

There are two advantages in modeling the range of the interval. The number of estimated parameters is reduced from $2(1 + 2p) + 3$ in the full ML estimation (3.2) to $1 + 2p + 1$ in (3.9). More importantly, for the truncated normal regression, there is a unique solution to the maximization problem so that the ML estimator is the global maximizer of the likelihood function. Consistency and asymptotic normality of the ML estimators and $\hat{\lambda}_{t-1}$ are easily established. We add the following assumption

Assumption 5. *(Mixing Conditions) The interval-valued stochastic process $\{Y_t\} = \{Y_{lt}, Y_{ut}\}$ is either (a) ϕ-mixing of size $-r/(2r-1)$, $r \geq 1$ or (b) α-mixing of size $-r/(r-1)$, such that $E|Y_{lt}|^{r+\delta} < \Delta < \infty$ and $E|Y_{ut}|^{r+\delta} < \Delta < \infty$ for some $\delta > 0$ for all t.*

Theorem 1. *(Consistency and Asymptotic Normality of the first-step ML Estimator) Let $\theta^* \equiv (\Delta\beta/\sigma_m, \sigma_m) \equiv (\Delta\beta^*, \sigma_m)$ be a $1 \times (2p+2)$ parameter vector corresponding to model (3.7). Under assumptions 1 – 5, the maximum likelihood estimator $\widehat{\theta}^*$ has the following properties,*

(a) $\widehat{\theta}^*_{ML}$ converges to the true value θ^*_0 in probability, i.e., $\widehat{\theta}^*_{ML} \xrightarrow{p} \theta^*_0$;

(b) $\widehat{\theta}^*_{ML}$ is asymptotically normally distributed, i.e. $\sqrt{T}(\widehat{\theta}^*_{ML} - \theta^*_0) \xrightarrow{d} N(0, \mathbf{V})$, where the asymptotic covariance matrix is $\mathbf{V} = -\operatorname{plim}_{T\to\infty}[E(\partial^2 L/\partial\theta^*\partial\theta^{*\prime}|_{\theta^*_0})]^{-1}$.

The truncated normal regression model has been extensively studied for cross-sectional data. Tobin (1958) proposed the maximum likelihood estimator and Amemiya (1973) proved its consistency and asymptotic normality. Orme (1989), Orme and Ruud (2002) proved that the solution to the likelihood equations is unique and that there is a global maximizer of the log-likelihood function. The proofs of the asymptotic properties in Amemiya (1973) are directly applicable to time series data by strengthening the moment conditions. With assumption 5, we replace the Kolmogorov's strong law of large numbers and Liapounov's central limit theorem for non-identically distributed random variables in Amemiya (1973) with McLeish(1974)'s strong law of large numbers (Theorem 2.10) and Wooldridge-White (1988)'s central limit theorem for mixing processes (Corollary 3.1) to guarantee that Theorem 1 holds. The asymptotic properties of the estimator of the inverse of the Mill's ratio follow as a corollary of Theorem 1 because $\lambda(\cdot)$ is a continuous and differentiable function with respect to θ^*.

Corollary 1. *(Consistency and Asymptotic Normality of the Inverse of the Mill's Ratio) The estimator of the inverse of the Mill's ratio* $\widehat{\Lambda} \equiv (\hat{\lambda}_0, \cdots, \hat{\lambda}_{T-1})$ *has the following properties*

(a) $\lambda(y^t, \widehat{\Delta\beta}^*_{ML})$ *converges in probability to the true* $\lambda(y^t, \Delta\beta^*_0)$, *i.e.,* $\lambda(y^t, \widehat{\Delta\beta}^*_{\mathrm{ML}}) \xrightarrow{p} \lambda(y^t, \Delta\beta^*_0)$;

(b) $\widehat{\Lambda}$ *is asymptotically normally distributed, i.e.,* $\sqrt{T}(\widehat{\Lambda} - \Lambda) \xrightarrow{d} N(\mathbf{0}, \mathbf{S}_0)$, *where the asymptotic covariance matrix* $\mathbf{S}_0 = \mathbf{J}(\Delta\beta^*_0)\mathbf{V}_{\Delta\beta^*_0}\mathbf{J}(\Delta\beta^*_0)'$ *and* $\mathbf{V}_{\Delta\beta^*_0}$ *is the asymptotic covariance matrix of* $\sqrt{T}(\widehat{\Delta\beta}^*_{ML} - \Delta\beta^*_0)$, *a leading principal minor of matrix* $\mathbf{V}$. *The t-th row of matrix* $\mathbf{J}(\Delta\beta^*_0)$ *is,* $j_t = \lambda(y^{t-1}, \Delta\beta^*_0)[\lambda(y^{t-1}, \Delta\beta^*_0) - \mathbf{z}_{t-1}\Delta\beta^*_0]\mathbf{z}_{t-1}$, *and vector* $\mathbf{z}_{t-1}$ *is* $(1, y_{l,t-1}, y_{u,t-1}, \cdots, y_{l,t-p}, y_{u,t-p})$.

3.4 Two-step Estimation: The Second Step

We plug the estimate $\widehat{\lambda}_{t-1}$ in the regressions (3.5) and (3.6) to obtain the feasible model. We need to redefine the new error terms in the feasible regressions as u_{lt} and u_{ut}, which have two sources of variation, one coming from the λ estimator, and the other coming from the error term in the infeasible regression, i.e. $u_{lt} = C_l(\lambda_{t-1} - \widehat{\lambda}_{t-1}) + v_{lt}$ and $u_{ut} = C_u(\lambda_{t-1} - \widehat{\lambda}_{t-1}) + v_{ut}$. As a result, the error term of the feasible regression will be heteroscedastic. Writing the feasible regressions in matrix form

$$\mathbf{y}_l = \widehat{\mathbf{H}}\gamma_l + \mathbf{u}_l, \quad \mathbf{y}_u = \widehat{\mathbf{H}}\gamma_u + \mathbf{u}_u \tag{3.10}$$

where

$$\begin{aligned}
&\mathbf{y}_l = (y_{l,1}, \cdots, y_{l,T})', \quad && \mathbf{y}_u = (y_{u,1}, \cdots, y_{u,T})', \\
&\gamma_l = (\beta_l, C_l), && \gamma_u = (\beta_u, C_u), \\
&\mathbf{u}_l = C_l(\Lambda - \widehat{\Lambda}) + \mathbf{v}_l, && \mathbf{u}_u = C_u(\Lambda - \widehat{\Lambda}) + \mathbf{v}_u, \\
&\widehat{\Lambda} = (\widehat{\lambda}_0, \cdots, \widehat{\lambda}_{T-1})', && \widehat{\mathbf{H}} = (\mathbf{Z}, \widehat{\Lambda}),
\end{aligned}$$

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_0 \\ \vdots \\ \mathbf{z}_{T-1} \end{bmatrix} = \begin{bmatrix} 1 & y_{l,0} & y_{u,0} & \cdots & y_{l,1-p} & y_{u,1-p} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & y_{l,T-1} & y_{u,T-1} & \cdots & y_{l,T-p} & y_{u,T-p} \end{bmatrix}.$$

The least squares estimators of the parameters γ_l and γ_u are

$$\widehat{\gamma}_l = (\widehat{\mathbf{H}}'\widehat{\mathbf{H}})^{-1}\widehat{\mathbf{H}}'\mathbf{y}_l, \qquad \widehat{\gamma}_u = (\widehat{\mathbf{H}}'\widehat{\mathbf{H}})^{-1}\widehat{\mathbf{H}}'\mathbf{y}_u \tag{3.11}$$

The next theorem establishes the asymptotic properties of the two-step estimators $\widehat{\gamma}_l$ and $\widehat{\gamma}_u$.

Theorem 2. *(Consistency and asymptotic normality of the second step OLS estimator) Under the following assumptions,*

(i) $\text{plim}_{T\to\infty} \mathbf{H}'\mathbf{H}/T = \mathbf{B}^{-1}$*, which is nonsingular;*

(ii) $\mathbf{H}'\mathbf{J}(\Delta\beta^*)/T$ *converges uniformly in probability to the matrix function* $\mathbf{Q}(\Delta\beta^*)$*;* $\mathbf{J}'(\Delta\beta^*)\mathbf{J}(\Delta\beta^*)/T$ *is bounded uniformly in probability at least in a neighborhood of true value* $\Delta\beta_0^*$*;*

(iii) $E|h_{t-1,i}v_{lt}|^2 < \infty$, $E|h_{t-1,i}v_{ut}|^2 < \infty$*, and* $E|j_{t-1,i}v_{lt}|^2 < \infty$ *for all* t *and* $i = 1, \cdots, 2p+2$*;*

(iv) $\mathbf{\Psi}_{l,T} \equiv \text{var}(T^{-1/2}\mathbf{H}'\mathbf{v}_l) \xrightarrow{p} \mathbf{\Psi}_l$ *and* $\mathbf{\Psi}_{u,T} \equiv \text{var}(T^{-1/2}\mathbf{H}'\mathbf{v}_u) \xrightarrow{p} \mathbf{\Psi}_u$*, and* $\mathbf{\Psi}_l$*,* $\mathbf{\Psi}_u$ *are finite and positive definite;*

Then, the two-step estimators $\widehat{\gamma}_l$ *and* $\widehat{\gamma}_u$

(a) converge to their true values in probability,

(b) with asymptotic normal distributions, i.e., $\sqrt{T}(\widehat{\gamma}_l - \gamma_l) \xrightarrow{d} N(\mathbf{0}, \mathbf{B}\mathbf{\Xi}_l\mathbf{B}')$*, and* $\sqrt{T}(\widehat{\gamma}_u - \gamma_u) \xrightarrow{d} N(\mathbf{0}, \mathbf{B}\mathbf{\Xi}_u\mathbf{B}')$*, where* $\mathbf{B} \equiv \text{plim}_{T\to\infty}(\widehat{\mathbf{H}}'\widehat{\mathbf{H}}/T)^{-1} = \text{plim}_{T\to\infty}(\mathbf{H}'\mathbf{H}/T)^{-1}$*, and*

$$\mathbf{\Xi}_l = \mathbf{\Psi}_l + C_l^2\mathbf{Q}_0'\mathbf{S}_0\mathbf{Q}_0 + \mathbf{M}_{l0} + \mathbf{M}_{l0}' \tag{3.12}$$

$$\mathbf{\Xi}_u = \mathbf{\Psi}_u + C_u^2\mathbf{Q}_0'\mathbf{S}_0\mathbf{Q}_0 + \mathbf{M}_{u0} + \mathbf{M}_{u0}' \tag{3.13}$$

ACCEPTED MANUSCRIPT

with

$$\mathbf{Q}_0 = \text{plim}_{T\to\infty}\,\mathbf{H}'\mathbf{J}(\Delta\beta_0^*)/T, \qquad \mathbf{S}_0 = \mathbf{J}(\Delta\beta_0^*)\mathbf{V}_{\Delta\beta_0^*}\mathbf{J}(\Delta\beta_0^*)',$$

$$\mathbf{M}_{l0} = \text{plim}_{T\to\infty}\,E\left(\mathbf{H}'\mathbf{v}_l(\Lambda - \widehat{\Lambda})'\mathbf{H}C_l\right)/T, \quad \mathbf{M}_{u0} = \text{plim}_{T\to\infty}\,E\left(\mathbf{H}'\mathbf{v}_u(\Lambda - \widehat{\Lambda})'\mathbf{H}C_u\right)/T.$$

In equations (3.12) and (3.13), the first terms $\mathbf{\Psi}_l$ and $\mathbf{\Psi}_u$ are the variance-covariance matrices of the errors v_{lt} and v_{ut} respectively, if Λ were observable. The second term $\mathbf{Q}_0'\mathbf{S}_0\mathbf{Q}_0$ captures the uncertainty induced by the estimates of $\widehat{\Lambda}$. The last two terms, $\mathbf{M}_{l0}$ and $\mathbf{M}_{u0}$, capture the covariances between the error terms v_{lt} and v_{ut} with $\widehat{\Lambda}$. Although v_{lt} and v_{ut} are martingale difference sequences, they are correlated with λ_{t+i} for $i = 0, 1, \cdots, T-t$. This is a further difference with Heckman's two-step estimator. In Heckman's covariance matrix, the matrix $\mathbf{M}_0$ is zero because in a cross-sectional setting the error v is uncorrelated with the inverse of the Mill's ratio. Since the asymptotic variance-covariance matrices in (3.12) and (3.13) capture the heteroscedasticity induced by the observability restriction together with the time dependence induced by $\widehat{\Lambda}$, Newey and West (1994)'s HAC variance-covariance matrix estimator should suffice to estimate $\mathbf{B}\mathbf{\Xi}_l\mathbf{B}$ and $\mathbf{B}\mathbf{\Xi}_u\mathbf{B}$ consistently. We also estimate the unconditional variances σ_l^2 and σ_u^2 of the respective errors ε_{lt} and ε_{ut} and their correlation coefficient ρ by implementing a simple method of moments (see web appendix).

3.5 Two-step Estimation: Implementation Issues

The implementation of the two-step estimator may be subject to multicollinearity, and consequently the parameters γ_l and γ_u in the second step, equations (3.10), may not be precisely estimated or, in extreme cases, they may not be identified at all. There are two reasons for multicollinearity. First, the functional form (3.3) of the inverse of the Mill's ratio $\lambda(\cdot)$ is nearly linear over a wide range of its argument $\Delta(y^{t-1}, \Delta\beta)/\sigma_m$ so that the estimated regressor $\widehat{\Lambda}$ is almost collinear with the regressors in $\mathbf{Z}$. This multicollinearity issues cannot be resolved by just dropping some of the regressors because the inclusion of $\widehat{\Lambda}$ is necessary to guarantee the consistency of the estimators $\hat{\beta}_l$ and $\hat{\beta}_u$.

The second reason pertains to those cases in which the observability condition is not binding.

When the observability condition is not binding, the population value of $\lambda(\cdot)$ is zero. Within a sample, we will observe values close to zero and very small variance in $\hat{\lambda}_t$. The direct consequence is that C_l and C_u are not identifiable. In the simulation section, we will discuss cases in which this problem is severe.

For these two reasons, we propose a *modified* second step estimator that overcomes the identification problem of C_l and C_u, and in addition, provides a direct identification of the unconditional variances σ_l^2 and σ_u^2 of the respective structural errors ε_{lt} and ε_{ut} and their correlation coefficient ρ.

3.6 Two-step Estimation: A Modified Two-step Estimator

The first step of the estimation is identical to that explained in section 3.3 , from which we obtain the estimates $\widehat{\boldsymbol{\Lambda}}$ and $\widehat{\sigma}_m$. In the second step, we exploit the relationships among C_l, C_u, σ_u^2, and σ_l^2, i.e.,

$$C_u + C_l = [\sigma_u^2 - \sigma_l^2]/\sigma_m \quad \text{and} \quad C_u - C_l = \sigma_m. \tag{3.14}$$

If σ_l^2, σ_u^2 and σ_m were known, the system of equations (3.14) would have a unique solution, and C_l and C_u will be uniquely identified. By writing σ_u^2 and σ_l^2 as functions of C_l and C_u, i.e. $\sigma_u^2(C_u)$ and $\sigma_l^2(C_l)$, we propose the following minimum distance estimator, which permits identifying C_l and C_u,

$$(\widetilde{C}_l, \widetilde{C}_u) = \underset{(C_l, C_u)}{\arg\min}\{C_u + C_l - [\sigma_u^2(C_u) - \sigma_l^2(C_l)]/\widehat{\sigma}_m\}^2, \quad \text{such that} \quad C_u - C_l = \widehat{\sigma}_m. \tag{3.15}$$

Our first task is to find $\sigma_u^2(C_u)$ and $\sigma_l^2(C_l)$. In order to do so, observe that the *unconditional* variance σ_u^2 and σ_l^2 of the error terms ε_{ut} and ε_{lt} can be written as follows

$$\begin{aligned}\sigma_l^2 &= \operatorname{var}(\varepsilon_{lt}) = \operatorname{var}(E(\varepsilon_{lt}|\Delta\varepsilon_t \geq \Delta(y^{t-1}; \Delta\beta))) + E(\operatorname{var}(\varepsilon_{l,t}|\Delta\varepsilon_t \geq \Delta(y^{t-1}; \Delta\beta))) \\ &= C_l^2 \operatorname{var}(\lambda_{t-1}) + E(\operatorname{var}(v_{lt}|y^{t-1})).\end{aligned} \tag{3.16}$$

Similarly, $\sigma_u^2 = C_u^2 \operatorname{var}(\lambda_{t-1}) + E(\operatorname{var}(v_{ut}|y^{t-1}))$, and

$$\sigma_m^2 = \operatorname{var}(\Delta\varepsilon_t) = \sigma_m^2 \operatorname{var}(\lambda_{t-1}) + E(\operatorname{var}(\Delta v_t|y^{t-1})), \tag{3.17}$$

with $\Delta v_t = v_{ut} - v_{lt} = \Delta y + z_{t-1}\Delta\beta - \sigma_m\lambda_{t-1}$ by subtracting (3.5) and (3.6), and $\Delta\beta$ defined by (3.4).

From (3.17), we have $\operatorname{var}(\lambda_{t-1}) = 1 - E[\operatorname{var}(\Delta v_t|y^{t-1})]/\sigma_m^2$, so that we need consistent estimators for the population moments $E(\mathrm{var}(v_{lt}|y^{t-1}))$, $E(\mathrm{var}(v_{ut}|y^{t-1}))$ and $E(\mathrm{var}(\Delta v_t|y^{t-1}))$ to obtain $\sigma_u^2(C_u)$ and $\sigma_l^2(C_l)$ as functions of sample information. The following Proposition 1 guarantees that this is the case. First, let us call

$$\widehat{\Delta v_t} = \Delta y_t + \mathbf{z}_{t-1}\widehat{\Delta\beta} - \widehat{\sigma}_m\widehat{\lambda}_{t-1}, \tag{3.18}$$

$$\widehat{u}_{lt} = y_{lt} - \mathbf{z}_{t-1}\beta_l(C_l) - C_l\widehat{\lambda}_{t-1}, \tag{3.19}$$

$$\widehat{u}_{ut} = y_{ut} - \mathbf{z}_{t-1}\beta_u(C_u) - C_u\widehat{\lambda}_{t-1}, \tag{3.20}$$

where $\widehat{\Delta\beta}$ and $\widehat{\lambda}_{t-1}$ are the estimates from the first step, and $\beta_l(C_l)$ and $\beta_u(C_u)$ are the concentrated OLS estimates of β in (3.10), i.e.

$$\beta_l(C_l) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{Y}_l - C_l\widehat{\mathbf{\Lambda}}), \qquad \beta_u(C_u) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{Y}_u - C_u\widehat{\mathbf{\Lambda}}), \tag{3.21}$$

Proposition 1. *Under assumptions 1 to 5 and for ϕ- or α-mixing sequences v_{lt} and v_{ut} with at least finite second moments, we have that $\sum_{t=1}^T \widehat{\Delta v}_t^2/T \xrightarrow{p} E(\mathrm{var}(\Delta v_t|y^{t-1}))$, $\sum_{t=1}^T \widehat{u}_{lt}^2/T \xrightarrow{p} E(\mathrm{var}(v_{lt}|y^{t-1}))$, $\sum_{t=1}^T \widehat{u}_{ut}^2/T \xrightarrow{p} E(\mathrm{var}(v_{ut}|y^{t-1}))$, and therefore, $\widehat{\sigma}_l^2(C_l) \equiv C_l^2(1 - \sum_{t=1}^T \widehat{\Delta v}_t^2/T\widehat{\sigma}_m^2) + \sum_{t=1}^T \widehat{u}_{lt}^2/T \xrightarrow{p} \sigma_l^2$ and $\widehat{\sigma}_u^2(C_u) \equiv C_u^2(1 - \sum_{t=1}^T \widehat{\Delta v}_t^2/T\widehat{\sigma}_m^2) + \sum_{t=1}^T \widehat{u}_{ut}^2/T \xrightarrow{p} \sigma_u^2$.*

The implementation of the minimum distance estimator in (3.15) is described in Figure 2.

[FIGURE 2]

We proceed as follows:

1. pick any point (C_l^*, C_u^*) on the line $C_u = \hat{\sigma}_m + C_l$;
2. compute the corresponding concentrated $\beta_l(C_l^*)$ and $\beta_u(C_u^*)$ as in (3.21);
3. compute the corresponding residuals $\widehat{u}_{lt}$, $\widehat{u}_{ut}$, and $\widehat{\Delta v}_t$ as in (3.19), (3.20), and (3.18) respectively;
4. calculate the intercept $(\sigma_u^2(C_u^*) - \sigma_l^2(C_l^*))/\widehat{\sigma}_m$ to obtain the point (C_l^*, C_u^{**}) on the line $C_u = [\sigma_u^2(C_u^*) - \sigma_l^2(C_l^*)]/\widehat{\sigma}_m + C_l$;
5. assess the distance $(C_u^* - C_u^{**})^2$;
6. go back to 1. Repeat until the distance function (3.15) is minimized by the minimizer $(\widetilde{C}_l, \widetilde{C}_u)$.

Given the optimal solution $(\widetilde{C}_l, \widetilde{C}_u)$, the estimators of the parameters β of the original model are readily available as well as the variance-covariance matrix of the errors ε_{lt} and ε_{ut}, i.e.

$$\begin{aligned}&\widetilde{\beta}_l = \beta_l(\widetilde{C}_l) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(Y_l - \widetilde{C}_l\widehat{\Lambda}), \quad \widetilde{\beta}_u = \beta_u(\widetilde{C}_u) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(Y_u - \widetilde{C}_u\widehat{\Lambda})\\ &\widetilde{\sigma}_l^2 = \sigma_l^2(\widetilde{C}_l),\ \widetilde{\sigma}_u^2 = \sigma_u^2(\widetilde{C}_u),\ \widetilde{\rho} = \frac{\widetilde{\sigma}_m^2 - \widetilde{\sigma}_l^2 - \widetilde{\sigma}_u^2}{-2\widetilde{\sigma}_l\widetilde{\sigma}_u}\end{aligned} \qquad (3.22)$$

Theorem 3. *(Consistency of Modified Two-step Estimator) The modified two-step estimator* $(\widetilde{C}_l, \widetilde{C}_u)$ *and those defined in* (3.22) *converge in probability to the true values of the parameters.*

In order to prove Theorem 3, which states the consistency of estimates $\widetilde{C}_l$ and $\widetilde{C}_u$ in (3.15), we only need to verify the assumptions stated in Theorem 7.3.2 in Mittelhammer *et. al.* (2000) that guarantee the consistency of extremum estimators.[3] Proposition 1 shows that the restricted objective function in (3.15) converges in probability to that provided in (3.14). In addition, since the system of equations (3.14) has a unique solution and the restricted objective function (3.15) is a continuous and convex function in C_l and C_u, it is uniquely minimized at the true values of C_l and C_u.

4 Simulation

We perform Monte Carlo simulations to assess the finite sample performance of the two proposed estimation strategies: the two-step and modified two-step estimators; and compare these estimators with a naive OLS estimator that does not take into account the observability restriction.

The data generating process (DGP) is specified as an IAR(1)

$$\begin{bmatrix} y_{l,t} \\ y_{u,t} \end{bmatrix} = \begin{bmatrix} \beta_{lc} \\ \beta_{uc} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} y_{l,t-1} \\ y_{u,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{l,t} \\ \varepsilon_{u,t} \end{bmatrix}, \quad \text{such that} \quad y_{u,t} \geq y_{l,t} \qquad (4.1)$$

and with an error term that is bivariate normally distributed $\epsilon_t \equiv (\varepsilon_{l,t}, \varepsilon_{u,t})' \sim N(\mathbf{0}, \mathbf{\Sigma})$.

The interval time series $\{[y_{l,t}, y_{u,t}]\}_{t=1}^{T}$ is generated sequentially to guarantee that the bounds are not crossing each other i.e. $y_{l,t} > y_{u,t}$. We proceed as follows. Given the interval datum $[y_{l,t-1}, y_{u,t-1}]$ at time $t-1$, we draw error terms $\epsilon_t = [\varepsilon_{l,t}, \varepsilon_{u,t}]$ from the bivariate normal density and calculate

[3] See Newey and MacFadden (1994, pp. 2133-34) for the proof. The four assumptions are (a) $m(\theta, \mathbf{Y}, \mathbf{X})$ converges uniformly in probability to a function of θ, say $m_0(\theta)$; (b) $m_0(\theta)$ is continuous in θ; (c) $m_0(\theta)$ is uniquely maximized at the true value θ_0; and (d) the parameter space Ω is compact.

ACCEPTED MANUSCRIPT

$[y_{l,t}, y_{u,t}]$ for time t. If a cross-over happens (i.e. $y_{l,t} > y_{u,t}$), we draw another pair of error terms until the observability restriction $y_{l,t} \leq y_{u,t}$ is met. In doing so, we guarantee that the errors ϵ_t are truncated bivariate normally distributed, and that the truncation varies across time because it depends on the past interval-valued data $[y_{t-1}]$ as well as on the assumed parameters β's in the IAR(1) DGP.

We have designed eight different specifications as follows:

	Binding Cases				Non-binding Cases			
Parameters	B-1	B-2	B-3	B-4	NB-1	NB-2	NB-3	NB-4
β_{lc}	0	0	0	0	-2	-2	-2	-2
β_{uc}	0	0	0	0	2	2	2	2
β_{11}	0.8	0.8	0.1	0.1	0.8	0.8	0.1	0.1
β_{12}	0.1	0.1	0.05	0.05	0.1	0.1	0.05	0.05
β_{21}	0.1	0.1	0.05	0.05	0.1	0.1	0.05	0.05
β_{22}	0.8	0.8	0.1	0.1	0.8	0.8	0.1	0.1
C_l	$-1/\sqrt{2}$	-1.4564	$-1/\sqrt{2}$	-1.4564	$-1/\sqrt{2}$	-1.4564	$-1/\sqrt{2}$	-1.4564
C_u	$1/\sqrt{2}$	-0.3479	$1/\sqrt{2}$	-0.3479	$1/\sqrt{2}$	-0.3479	$1/\sqrt{2}$	-0.3479
σ_l^2	1	3	1	3	1	3	1	3
σ_u^2	1	1	1	1	1	1	1	1
ρ	0	0.8	0	0.8	0	0.8	0	0.8
Sample Size	250, 2000	250, 2000	250, 2000	250, 2000	250, 2000	250, 2000	250, 2000	250, 2000
Number of Simulation = 1000								

We have simulated a block of four DGPs where the observability restriction is binding and another block of four DGPs where it is not. Since the observability restriction for the IAR(1) implies that $\Delta\epsilon_t/\sigma_m \geq \Delta(y_{t-1}, \theta^*)$. The right hand side of the inequality will determine whether the observability restriction is binding or not. We guarantee that the observability restriction is not binding when $\Delta(y_{t-1}, \theta^*) = \Delta\beta_c^* + \Delta\beta_1^* y_{l,t-1} + \Delta\beta_2^* y_{u,t-1} \ll 0$. Otherwise, the restriction could be mildly or severely binding depending upon the choices of the parameters of the DGP. In our simulations, we fix the parameters in $\Delta\beta_1^*$ and $\Delta\beta_2^*$ and play with the intercept $\Delta\beta_c^*$ to allow the restriction to be binding or not. For the four cases, B-1 to B-4, $\beta_{lc} - \beta_{uc} = 0$, so that the observability restriction becomes binding; and for the four cases, NB-1 to NB-4, $\beta_{lc} - \beta_{uc} = -4$, so that the restriction is not severely binding. Within each block, we simulate two IAR(1) DGPs, one with high persistence and another with low persistence; and for each one we assume two different variance-covariance matrix $\mathbf{\Sigma}$ for the errors, one with uncorrelated errors and another with highly correlated errors. For each DGP, we also run small and large sample experiments ($T = 250$ and 2000) with 1000 replications per DGP. Due to space constraints, we report here our results for only four cases, B-2 and B-4 in Table 1 and NB-2 and NB-4 in Table 2; the results for the remaining cases are in the web appendix. These are our findings for all eight cases:

[TABLES 1-2]

1. When the observability restriction is binding (Cases B-1 to B-4), the mean values of the OLS estimates are quite far from the true values, as we expected. OLS estimators are not consistent due to the correlation of the regressors with the errors. When the restriction is not severely binding (Cases NB-1 to NB-4), the mean values of the OLS estimates are very close to the true values. In this case, λ_{t-1} is very close to zero, so that the endogeneity problem does not arise.

2. When we implement the two-step estimation, the main issue that we face is identification of the model whether or not the restriction is binding. If the restriction is binding but λ_{t-1} is almost linear in the regressors of the model, multicollinearity arises (Cases B-3 and B-4). The problem is more severe when there is low persistence in the model and the errors are correlated (Case B-4). Only when λ_{t-1} exhibits substantial variation (Cases B-1 and B-2), we do not face a problem with the identification of the model and the mean values of the two-step estimates are very close to the true values. If the restriction is not binding, we expect severe multicollinearity. In Cases NB-1 to NB-4, the RMSE's of $\widehat{C_l}$ and $\widehat{C_u}$ explode regardless of the persistence of the model and the sample size. When there is low persistence in the model (Cases NB-3 and NB-4), the RMSE's of $\widehat{\beta}_{l,c}$ and $\widehat{\beta}_{u,c}$ also explode because the nearly-zero regressor λ_{t-1} is highly collinear with the constant terms.

3. Modified two-step estimation resolves very nicely the identification problem whether the observability restriction is binding or not. If it is binding (Cases B-1 to B-4), the estimators are consistent whether there is low or high persistence and whether the errors are or not correlated. The modified two-step estimates are very close to the true values and their standard errors are smaller than those of the two-step estimates, even in those cases where the model is well-identified

ACCEPTED MANUSCRIPT

(Cases B-1 and B-2). If the restriction is not binding (Cases NB-1 to NB-4) and thus redundant, the OLS estimator is consistent and efficient but the modified two-step estimator does not seem to be less efficient as the RMSE's of the modified two-step estimates are very close to those of the OLS estimates.

In practice, we do not know *a priori* whether the restriction is binding. In the first step, we assess the severity of the restriction by testing whether $\lambda_t = 0$. In the second step, we gather further information about the value of the restriction because when it is binding, the OLS estimates should be substantially different from the two-step estimates. In addition, the regressor $\widehat{\lambda}_{t-1}$ should be statistically significant. Since multicollinearity affects the significance of $\widehat{\lambda}_{t-1}$, we strongly recommend running the modified two-step estimator and assessing the differences with the OLS estimator.

5 Comparison with Existing Approaches

We compare our two-step (TS) and modified two-step (MTS) estimators with those proposed in the current literature. We implement the approach of Lima Neto and De Carvalho (2008, 2010), henceforth LNC, and we also estimate a location-scale model from which we construct interval estimates.

For an interval-valued time series $\{Y_t\} = \{[Y_{lt}, Y_{ut}]\}$, we obtain the time series of the centers, i.e., $y_{ct} = (y_{lt} + y_{ut})/2$, and of the radius, i.e., $y_{rt} = (y_{ut} - y_{lt})/2$. LNC estimate the following system

$$y_{ct} = \beta_0^c + \beta_1^c y_{c,t-1} + \cdots + \beta_p^c y_{c,t-p} + \epsilon_t^c \quad (5.1)$$

$$y_{rt} = \beta_0^r + \beta_1^r y_{r,t-1} + \cdots + \beta_p^r y_{r,t-p} + \epsilon_t^r. \quad (5.2)$$

Their center/range method (CRM) estimates each equation by least squares and their constrained center/range method (CCRM) imposes the restriction $\beta_j^r \geq 0$, $j = 0, \ldots, p$ on the equation of the radius to ensure that $\hat{y}_{rt} \geq 0$ and, therefore, $\hat{y}_{lt} \leq \hat{y}_{ut}$. Then, the equation of the center is estimated by least squares and the constrained equation of the radius by adapting Lawson and Hanson's algorithm.

Before we proceed with the comparison among methodologies, it is very important to underline the implications of the LNC system of center/radius equations for the system of lower/upper bound equations. It is easy to transform the center/radius vector to the lower/upper bound vector by defining the 2×2 matrix $M = [1/2 \quad 1/2\,; \,-1/2 \quad 1/2]$ such that $[y_{ct} \quad y_{rt}]' = M[y_{lt} \quad y_{ut}]'$. Hence, the system (5.1) and (5.2) is transformed into a system of lower/upper bounds equations as follows,

$$\begin{bmatrix} y_{lt} \\ y_{ut} \end{bmatrix} = \begin{bmatrix} \beta_0^c - \beta_0^r \\ \beta_0^c + \beta_0^r \end{bmatrix} + \sum_{l=1}^{p} \begin{bmatrix} (\beta_l^c + \beta_l^r)/2 & (\beta_l^c - \beta_l^r)/2 \\ (\beta_l^c - \beta_l^r)/2 & (\beta_l^c + \beta_l^r)/2 \end{bmatrix} \begin{bmatrix} y_{lt-1} \\ y_{ut-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t^c - \epsilon_t^r \\ \epsilon_t^c + \epsilon_t^r \end{bmatrix}, \tag{5.3}$$

which is extremely restrictive because, for each of the p coefficient matrices, the diagonal elements must be identical, equal to $(\beta_l^c + \beta_l^r)/2$, as well as the off-diagonal elements, equal to $(\beta_l^c - \beta_l^r)/2$. In the unlikely case that these restrictions hold, LNC and our approach will deliver the same results.

The second set of comparisons is with a location-scale model [4] applied to the time series of centers. We estimate a GARCH(1,1) model, i.e., $y_{ct} = \mu_c + \sigma_t \zeta_t$ with $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$, where the i.i.d. standardized error ζ_t follows a standard normal (GARCH-N) or Student-t (GARCH-T) density with ν degrees of freedom. Based on this model, we construct $(1-\alpha)$-probability intervals, which will depend on the density assumptions on ζ_t, i.e., $[\hat{y}_{lt}, \hat{y}_{ut}]_\alpha = \left[\overline{y}_{ct} - z_{\frac{\alpha}{2}} \hat{\sigma}_t, \overline{y}_{ct} + z_{\frac{\alpha}{2}} \hat{\sigma}_t\right]$, and $[\hat{y}_{lt}, \hat{y}_{ut}]_\alpha = \left[\overline{y}_{ct} - t_{\hat{\nu},\frac{\alpha}{2}} \hat{\sigma}_t \sqrt{\hat{\nu}/(\hat{\nu}-2)}, \overline{y}_{ct} + t_{\hat{\nu},\frac{\alpha}{2}} \hat{\sigma}_t \sqrt{\hat{\nu}/(\hat{\nu}-2)}\right]$. Since the original data $[y_{lt}, y_{ut}]$ are the observed extreme values of the process at time t, we will stretch the estimated interval $[\hat{y}_{lt}, \hat{y}_{ut}]_\alpha$ to cover as much as 99% or 99.5% probability, so that $\hat{y}_{lt}$ and $\hat{y}_{ut}$ are far away into the tails of the distribution.

We simulate data from four DGPs, which are characterized by whether the observability restriction is binding or not, whether there is high or low persistence in the dynamics of the conditional mean, and whether the errors of the model are drawn from a bivariate normal density or from a bivariate Student-t density with five degrees of freedom. The four DGPs are:

[4] We are grateful to a referee who suggested the 5-parameter location-scale model as a classical benchmark

	β_{0l}	β_{0u}	β_{11}	β_{12}	β_{21}	β_{22}	σ_l^2	σ_u^2	ρ	Notes*
DGP1	0	0	-0.8	0.1	-0.1	0.8	3	1	-0.8	B,H,N,T5
DGP2	0	0	-0.1	0.05	-0.05	0.1	3	1	-0.8	B,L,N,T5
DGP3	-2	2	-0.8	0.1	-0.1	0.8	3	1	-0.8	NB,H,N,T5
DGP4	-2	2	-0.1	0.05	-0.05	0.1	3	1	-0.8	NB,L,N,T5

* B: binding observability restriction; NB: non-binding;
H: high persistence; L: low persistence; N: normal errors; T5: Student-*t* errors

5.1 In-Sample Evaluation Criteria : Loss Functions

For every DGP, we generate 1000 samples and evaluate the performance of each estimation method according to: (i) Root Mean Squared Error (RMSE) for upper and lower bounds, (ii) Coverage (CR) and Efficiency Rates (ER) of the estimated intervals (Rodrigues and Salish, 2011), (iii) Multivariate Loss Functions (MLF) for the vector of lower and upper bounds (Komunjer and Owyang, 2011), and (iv) Mean Distance Error (MDE) between the fitted and actual intervals (Arroyo *et al.*, 2010).

For a sample of size T, let us call $\hat{y}_t = [\hat{y}_{lt}, \hat{y}_{ut}]$ the fitted values of the corresponding interval $y_t = [y_{lt}, y_{ut}]$ obtained by each methodology. These are the definitions of the four criteria:

(i) RMSE: $RMSE_l = \sqrt{\sum_{t=1}^{T}(\hat{y}_{lt} - y_{lt})^2/T}$ and $RMSE_u = \sqrt{\sum_{t=1}^{T}(\hat{y}_{ut} - y_{ut})^2/T}$;

(ii) CR and ER: $CR = \frac{1}{T}\sum_{t=1}^{T} w(y_t \cap \hat{y}_t)/w(y_t)$, $ER = \frac{1}{T}\sum_{t=1}^{T} w(y_t \cap \hat{y}_t)/w(\hat{y}_t)$, where $y_t \cap \hat{y}_t$ is the intersection of actual and fitted intervals, and $w(\cdot)$ is the width of the interval. The coverage rate (CR) is the average proportion of the actual interval covered by the fitted interval, and the efficiency rate (ER) is the average proportion of the fitted interval covered by the actual interval. Both rates are between zero and one and a large rate means a better fit. Given an actual interval, a wide fitted interval implies a large coverage rate but a low efficiency rate, on the contrary, a tight fitted interval implies a low coverage rate but a high efficiency rate. Therefore, we take into account the potential trade-off between the two rates by calculating an average of the two, i.e., $(CR + ER)/2$.

(iii) MLF: We implement the following multivariate loss function $L_p(\tau, \mathbf{e}) \equiv \left(\| \mathbf{e} \|_p + \tau'\mathbf{e}\right) \| \mathbf{e} \|_p^{p-1}$ where $\| \cdot \|_p$ is the l_p-norm, τ is two-dimensional parameter vector bounded by the unit ball $\mathcal{B}_q$ in $\mathbb{R}^2$ with l_q-norm (where p and q satisfy $1/p + 1/q = 1$), and $\mathbf{e} = (e_l, e_u)$ is the bivariate residual interval $(\hat{y}_{lt} - y_{lt}, \hat{y}_{ut} - y_{ut})$. We consider two norms, $p = 1$ and $p = 2$ and their corresponding τ parameter vectors within the unit balls $\mathcal{B}_\infty$ and $\mathcal{B}_2$ respectively, $MLF_1 = \int_{\tau \in \mathcal{B}_\infty} (|e_l| + |e_u| + \tau_1 e_l + \tau_2 e_u) d\tau$, $MLF_2 = \int_{\tau \in \mathcal{B}_2} \left[e_l^2 + e_u^2 + (\tau_1 e_l + \tau_2 e_u)(e_l^2 + e_u^2)^{1/2} \right] d\tau$.

(iv) MDE: Let $D^q(\hat{y}_t, y_t)$ be a distance measure of order q between the fitted and the actual intervals, the mean distance error is defined as $MDE^q(\{\hat{y}_t\}, \{y_t\}) = [\sum_{t=1}^{T} D^q(\hat{y}_t, y_t)/T]^{1/q}$. We consider $q = 1$ and $q = 2$, with a distance measure such as $D_(\hat{y}_t, y_t) = \frac{1}{\sqrt{2}}[(\hat{y}_{lt} - y_{lt})^2 + (\hat{y}_{ut} - y_{ut})^2]^{1/2}$.

In Tables 3 and4, we report the values of the four aforementioned evaluation criteria for DGP1 and DGP3 respectively. Results for DGP2 and DGP4 are available in the web appendix.

[TABLES 3-4]

The numbers in boldface correspond to the minimum losses when we consider the functions RMSE, MLF, and MDE, and to the maximum rates when we consider the weighted CR/ER rates. In each table, we provide two scenarios: in the upper panel, the DGP is simulated with multivariate normal errors so that our methods TS and MTS perform under the correct distributional assumption, and in the lower panel, the DGP is simulated with multivariate Student-*t* errors to assess the performance of TS and MTS under density misspecification. These are our findings for the four DGPs considered:

1. Across the four DGPs, TS and MTS exhibit superior performance over the other methods.

2. Across methods, TS and MTS are superior to CCRM and CRM, and these are far better than the GARCH models. The classical methodology embedded in normal or fat-tail location-scale models is by far the worst performer across all evaluation functions and it is very inefficient on delivering an acceptable fitted interval as the efficiency rates (ER) shows.

3. With misspecified Student-*t* errors, the losses across all methods are larger than those under correct error specification, which is expected, nevertheless TS and MTS provide the smallest loss.

4. Across DGPs, DGP1 and DGP3, which have high persistence in the conditional mean, have the smallest losses, and in particular, TS and MTS deliver unmatched performance even in the cases of misspecified distributional assumptions.

5. DGP2 and DGP4 have low persistence in the conditional mean. In these two specifications, the coefficients are all very close to zero, thus, in these cases the constraints imposed by CCRM and CRM are not so restrictive and, as a consequence, the performance of CCRM and CRM is close to that of TS and MTS, but the performance of the location-scale models is still far behind the other methods.

6. Only for DGP4 with low persistence in mean and non-binding observability restriction, the performance of all methods is roughly equivalent, which is expected as all constraints are relaxed.

In summary, when the researcher faces an interval-valued data set, a priori, she does not know the persistence of the data and whether the observability restriction is or is not binding, thus, it is advisable to start the estimation of the model by implementing TS or/and MTS. If there is high persistence in the conditional mean, even if the observability restriction is non-binding, it pays off to implement TS and MTS as the losses are substantially smaller than those from the competing methodologies. In addition, the implementation of a location-scale model also entails the choice of distributional assumptions, which is subject to misspecification issues.

5.2 In-Sample Evaluation Criteria: Mean Estimates, Bias, and MSE

We compare the mean estimates of the parameters in the conditional mean delivered by TS and MTS with those provided by CCRM and CRM. As before, we consider four DGPs with correctly specified multivariate normal errors and with Student-*t* errors to assess the effect of density misspecification.

For DGP1 and DGP3, we present the simulation results in Table 5 for the case of multivariate normal errors and in Table 6 for multivariate Student-*t* errors (5 degrees of freedom). Similar tables for DGP2 and DGP4 can be found in the web appendix. The numbers in boldface are the best estimates, the lowest bias and the lowest mean-square error.

[TABLES 5-6]

For normal errors, when the restriction is binding and there is high persistence (DGP1), CCRM and CRM perform very badly. The mean estimates have a large bias and frequently the wrong sign. On the contrary, TS and MTS deliver unbiased estimates with the lowest mean-square error. When the process has low persistence (DGP2), the best estimation method is MTS, which delivers unbiased estimates. TS suffers from the multicollinearity problem explained above and thus it is not recommended if our interest is understanding the dynamics of the conditional mean. CCRM and CRM estimates are not recommended either because of their large bias. In DGP3 and DGP4, the observability restriction is non-binding but the results are very similar. When the process has high persistence (DGP3), either TS or MTS deliver unbiased estimates with the lowest mean-square error, and CCRM and CRM generate highly biased estimates. When the process has low persistence (DGP4), MTS is the best performer because it takes care of the multicollinearity problem and delivers unbiased estimates.

For Student-*t* errors, when the observability restriction is binding and there is high persistence (DGP1), the best performer is TS followed by MTS as they provide estimates with the lowest biases and capture the right dynamics. On the other hand, CCRM and CRM do not capture the persistence in the conditional mean and their estimates are highly biased. A common problem to these four methods is that the estimates of the constants are very biased. However, in TS and MTS, these biases are somehow compensated by the estimates of the coefficients corresponding to the regressor λ_{t-1} so that the overall estimation generates good fitted intervals with substantially lower losses than those generated by CCRM and CRM as we have seen in Table 3 (lower panel). Thus, the misspecification of the multivariate density does not seem to affect greatly the performance of TS and MTS. When the process has low persistence (DGP2), no method seems to deliver overall unbiased estimates, and the problem of the estimation of the constant is severe. Note that the design of low persistence with binding observability restriction (DGP2) represents the worst scenario because, by construction, the intervals are very tight; the specification of the conditional means deliver very small values around zero, so that the regressor λ_{t-1} carries all the weight to estimate fitted intervals with the right order. Yet TS delivers the smallest losses. In DGP3 and DGP4, the observability restriction is non-binding. When the process has high persistence (DGP3), TS and MTS are superior performers, they deliver unbiased estimates with the lowest mean-square error. CCRM and CRM produce highly biased estimates. When the process has low persistence (DGP4), MTS is the best performer overall.

In summary, evaluating the estimation performance of the four methods, we reach similar conclusions as those when we evaluate their goodness of fit. Even under misspecification of the multivariate density of the errors, if there is high persistence in the conditional mean, whether the observability restriction is binding or not, TS and MTS are superior estimation techniques. If the persistence is low and the observability restriction is non-binding, we recommend MTS, even with a misspecified density. Only when the persistence is low and the observability restriction is binding, the misspecification of the density may play a role on estimating the right dynamics but yet TS and MTS are not dominated by the competing methods and they offer the advantage of preserving the natural order of an interval.

6 Empirical Illustration: SP500 Low/High Return Interval

We highlight the most important aspects of our methodology with the interval time series of the daily low/high returns to the SP500 index. The returns are computed with respect to the closing price of the previous day, that is, $r_{ht} = (P_{high,t} - P_{close,t-1})/P_{close,t-1}$ and $r_{lt} = (P_{low,t} - P_{close,t-1})/P_{close,t-1}$, where $P_{high,t}$ and $P_{low,t}$ are the highest and lowest price in the trading day t, and $P_{close,t-1}$ is the closing price in the previous day $t-1$. Our sample runs from January 1st, 2004 to April 29th, 2011. We have split the sample into two periods that have very different dynamics so that we can showcase the role of the observability condition in the modeling exercise. The first period goes from January 1st, 2004 to January 1st, 2007; we call it the 'stable period' because is characterized by very low volatility. In contrast, the second period that goes from January 1st, 2007 to April 29th, 2011 is the 'unstable period' because of the high volatility associated with the great panic of the 2008 financial crisis. For both periods, we plot the time series of low/high returns interval in Figure 3a and Figure 3b.

[FIGURE 3a] [FIGURE 3b]

In the stable period, both low and high returns exhibit low volatilities ($\sigma_l^2 = 0.1726$ and $\sigma_u^2 = 0.1609$), varying within a range of $[-2\%, 2\%]$, whereas in the unstable period, the two time series vary within a wider range of $[-5\%, 5\%]$, and exceptionally, in the last months of 2008, moving within a range of -10% and 10%, thus producing a much higher volatile environment ($\sigma_l^2 = 1.6539$ and $\sigma_u^2 = 1.3347$). The unstable period is dominated by a tremendous volatility shock, which is not present in the stable period. The correlation of low and high returns is 0.5797 in the stable period, which is larger than the correlation of 0.2982 in the unstable period.

Due to space constraints, we offer here a summary of the estimation results and we report specific details in several tables posted in the web appendix. We run an unrestricted IAR(p) system and select the optimal lags by minimizing the BIC. In the stable period, the optimal number of lags is 2, and in the unstable period is 5. We implement the first-step of the estimation by modeling the range of the interval time series $\Delta r_t = r_{ut} - r_{lt}$ as in (3.7). By maximizing the log-likelihood function based on a truncated normal density (3.8), we obtain the estimates $\widehat{\theta}^*$ for the stable and unstable periods. We observe that the correlation between the range and lagged lower-bound returns is negative, while the correlation between the range and lagged upper-bound returns is positive; however the magnitude of the effect of the lower-bound returns is dominant, which implies that, on average, the range will narrow when there is an upward movement in both bounds.

Based on the estimates $\widehat{\theta}^*$, we produce an estimate of the inverse of Mill's ratio λ_{t-1}, which characterizes the severity of the observability restriction. We plot the estimated time series $\widehat{\lambda}_t$ in Figures 4a and Figure 4b together with a 95% confidence interval.

[FIGURE 4a] [FIGURE 4b]

In the stable period, the values of $\widehat{\lambda}_t$ are very small, between 0 and 0.070; the mean is 0.027 and the standard deviation 0.014. This indicates that λ_t is practically zero, thus the observability restriction is not binding. In contrast, in the unstable period, $\widehat{\lambda}_t$ oscillate between 0 and 0.684

ACCEPTED MANUSCRIPT

with mean 0.213 and standard deviation 0.181; these values imply that the relevant portion of the function $\widehat{\lambda}_t$ is not entirely linear in the regressors. In the unstable period, there are a few regions where $\widehat{\lambda}_t$ is very close to zero; this happens mainly in the highly volatile period of the end of 2008, when the range of the interval is very large, so that the observability restriction is less binding than in the rest of the sample.

With the estimated $\widehat{\lambda}_{t-1}$, we implement the second step of the estimation. We calculate the second-step estimator by running the feasible regressions (3.10), and the modified second-step minimum distance estimator by solving the problem in (3.15). We also implement a stationary block bootstrap procedure (details reported in the web appendix) to obtain the standard errors of the modified second-step estimator because the analytical expression of the standard errors will be difficult to obtain as we carry three sources of uncertainty, i.e. $\hat{\Lambda}$ and $\hat{\sigma}_m$ in the first step, the estimates $(\widetilde{C}_l, \widetilde{C}_u)$ in the modified second step, and the idiosyncratic uncertainty of the errors in the IAR system. Bootstrap is a common practice to overcome the difficulties of the estimation of asymptotic variances in various contexts, see Efron (1979), Buchinsky (1995), Ledoit, Santa-Clara, and Wolf (2003), and Goncalves and White (2005), among others. The optimal block size for the stable period is around 2 and for the unstable period 53. This large difference in the block size can be interpreted as the existence of larger persistence in the IAR system of the unstable period than in the stable period.

For the stable period, the estimated $\widehat{\lambda}_t$ in the first step already suggest that the observability restriction is not binding, thus OLS should suffice. However, in the estimation tables we also report the estimates from the two-step and modified two-step estimation procedures to underline the presence of multicollinearity caused by $\widehat{\lambda}_t$ being almost zero. We note that the OLS estimates and the modified minimum-distance estimates are almost identical, and that there is not loss of efficiency by implementing the modified estimator. This is what we expect when the restriction is not binding. Furthermore, the two-step estimator is less reliable, the estimates are different from the OLS estimates, even changing signs, and their standard errors are large as a consequence of

ACCEPTED MANUSCRIPT

the induced multicollinearity. For the unstable period, we know that $\widehat{\lambda}_t$ is large and different from zero, thus the observability restriction is binding in a substantial part of the sample. As expected, the modified two-step estimates are different from the OLS estimates, more so in the regression for the lower bound. We note that the estimates associated with $\widehat{\lambda}_t$, though with the right signs, are barely significant in the the two-step estimation because of some mild multicollinearity, which is corrected in the modified two-step estimation.

The severity of observability restriction is better illustrated in Figures 5a and 5b.

[FIGURE 5a] [FIGURE 5b]

The ellipses are contours of the bivariate normal probability density of the errors with different confidence levels (from 50% to 99%). The contours are drawn according to the estimates produced by the modified two-step estimation procedure. The 45-degree lines indicate the role of the observability restrictions for each time t (see Figure 1), so that the area of the density below the line is truncated. Observe that in the stable period, Figure 5a, the contours are smaller than those in the unstable period, Figure 5b, because of smaller variances. In the stable period, the lines corresponding to the observability restriction are clustered outside the 99% contour level, so that the truncation is minimal; however for the unstable period, the truncation of the bivariate density is large, mainly in the direction of the south-east quadrant, indicating the severity of the observability restriction.

Finally, we compare the performance of the different estimation techniques by considering the same loss functions as in Section 5, i.e. RMSE, CR & ER, MLF, and MDE, which is reported in Table 7.

[TABLE 7]

The upper panel shows the results for the unstable period (2007-2011) when the observability restriction is binding, and the lower panel for the stable period (2004-2007) when the observability restriction is non-binding. Overall and across panels, the estimation of a location-scale model, either with normal or Student-t errors, is not satisfactory, as the RMSE, MLF, and MDE losses are the largest among all methods. The location-scale model seems to provide slightly better CR & ER rates. In the unstable period, when the observability restriction is binding, TS and MTS provide the smaller losses; and in the stable period, when the observability restriction is non-binding, the losses of TS and MTS are equivalent to those of CCRM and CRM, as the restrictions become lax. The overall performance in both periods is consistent with that described for the simulated DGP2 and DGP4 in Section 5; these two DGPs contemplate low persistence in the conditional mean, which is what we found in the estimation of the low/high returns for the stable and unstable periods.

7 Conclusion

The analysis of interval-valued data has mainly focused on fitting classical regression models to the lower and upper bounds of the intervals but the natural order of the bounds has not been taken into account in the estimation of the regression. As a result, it is possible that for some observations the fitted lower bound could be larger than the fitted upper bound. In our analysis, we have constrained the regression such that a reversal of the bounds will never happen. The constraint is probabilistic in nature as the errors of the model come from a truncated bivariate probability density to guarantee the natural order of the interval. The truncation has several consequences for the estimation of the model. Even when the regression model is linear, an ML estimator will be non-linear and difficult to compute. If we were to apply OLS, the estimator would not be consistent because the truncation makes the error correlated with the regressors. To solve both predicaments, we have proposed a two-step estimation procedure, easy to implement, that delivers consistent estimators. It consists of a maximum likelihood estimator in the first step and either least-squares or minimum distance estimation in the second. The minimum distance estimator is a neat solution when there is substantial multicollinearity because identifies all parameters regardless of how large or small the truncation is.

We have shown that our estimators are superior over the existent approaches by examining several goodness-of-fit measures and concluded that, even when the observability restriction is non-binding, it pays off to implement our estimators because their losses are smaller than those from competing methods. We have also examined the bias and mean-squared error of our estimators with those of other methods and our conclusions remain unchanged. Even under misspecification of the multivariate density of the errors, when there are relevant dynamics in the conditional mean of the model, our estimators are superior. We have highlighted several empirical aspects of our methodology with the time series of the daily interval of low/high returns to the SP500 index and showed two instances, minimal and severe truncation, to underscore the value of implementing the proposed two-step estimator.

WEB APPENDIX

A1. (Section 3. Estimation)

Proof of the conditional expectation of the errors (section 3.2 of the article)

The observability restriction $y_{ut} \geq y_{lt}$ is equivalent to $\varepsilon_{ut} - \varepsilon_{lt} \geq \Delta(y^{t-1}, \Delta\beta)$, which linearly truncates the bivariate distribution of the errors. Now, we show the conditional mean, variance and covariance of the errors under the linear truncation. For the detailed proofs, please refer to Nath (1972). To easy the notation, we temporarily use Δ to denote $\Delta(y^{t-1}, \Delta\beta)$, and drop all the time subscripts t in the following expressions. Under Assumption 4 (normality of errors), we have,

$$m_{10} = E(\varepsilon_l|\varepsilon_u - \varepsilon_l \geq \Delta) = \frac{\rho\sigma_l\sigma_u - \sigma_l^2}{\sigma_m}\frac{\phi(\Delta/\sigma_m)}{1-\Phi(\Delta/\sigma_m)} \quad \text{(A.1)}$$

$$m_{01} = E(\varepsilon_u|\varepsilon_u - \varepsilon_l \geq \Delta) = \frac{\sigma_u^2 - \rho\sigma_l\sigma_u}{\sigma_m}\frac{\phi(\Delta/\sigma_m)}{1-\Phi(\Delta/\sigma_m)} \quad \text{(A.2)}$$

$$m_{20} = E(\varepsilon_l^2|\varepsilon_u - \varepsilon_l \geq \Delta) = \sigma_l^2 + \frac{\sigma_l^2(\rho\sigma_u - \sigma_l)^2}{\sigma_m^2}\frac{\Delta}{\sigma_m}\frac{\phi(\Delta/\sigma_m)}{1-\Phi(\Delta/\sigma_m)} \quad \text{(A.3)}$$

$$m_{02} = E(\varepsilon_u^2|\varepsilon_u - \varepsilon_l \geq \Delta) = \sigma_u^2 + \frac{\sigma_u^2(\sigma_u - \rho\sigma_l)^2}{\sigma_m^2}\frac{\Delta}{\sigma_m}\frac{\phi(\Delta/\sigma_m)}{1-\Phi(\Delta/\sigma_m)} \quad \text{(A.4)}$$

$$m_{11} = E(\varepsilon_l\varepsilon_u|\varepsilon_u - \varepsilon_l \geq \Delta) = \rho\sigma_l\sigma_u + \frac{\sigma_l\sigma_u(\rho\sigma_u - \sigma_l)(\sigma_u - \rho\sigma_l)}{\sigma_m^2}\frac{\Delta}{\sigma_m}\frac{\phi(\Delta/\sigma_m)}{1-\Phi(\Delta/\sigma_m)} \quad \text{(A.5)}$$

Note that the conditional means m_{10} and m_{01} in (A.1) and A.2 correspond to $E_{t-1}(\varepsilon_{lt}|y_{ut} \geq y_{lt})$ and $E_{t-1}(\varepsilon_{ut}|y_{ut} \geq y_{lt})$ respectively; and the conditional variances and covariance (A.3) – (A.5) are used to estimate unconditional variances and correlation coefficient σ_u^2, σ_l^2 and ρ consistently in the two-step estimation procedure.

Proofs of Theorem 2, and estimation of unconditional variances and correlation coefficient (section 3.4 of the article)

(a) Consistency

We only consider the regression of lower bounds y_{lt}, the same reasoning applies to the estimators for the upper bound. From the two-step estimator $\widehat{\gamma}_l$,

$$\begin{aligned}\widehat{\gamma}_l &= \left[\widehat{\beta}_l \quad \widehat{C}_l\right]' = \left(\widehat{\mathbf{H}}'\widehat{\mathbf{H}}\right)^{-1}\widehat{\mathbf{H}}'\mathbf{y}_l \\ &= (\widehat{\mathbf{H}}'\widehat{\mathbf{H}})^{-1}\widehat{\mathbf{H}}'(\widehat{\mathbf{H}}\gamma_l + \mathbf{u}_l) \\ &= \gamma_l + (\widehat{\mathbf{H}}'\widehat{\mathbf{H}})^{-1}\widehat{\mathbf{H}}'\mathbf{u}_l\end{aligned}$$

in which $\mathbf{u}_l \equiv C_l(\Lambda - \widehat{\Lambda}) + \mathbf{v}_l$. Note that,

$$\operatorname*{plim}_{T\to\infty}(\widehat{\mathbf{H}}'\widehat{\mathbf{H}})^{-1}\widehat{\mathbf{H}}'\mathbf{u}_l = \operatorname*{plim}_{T\to\infty}\left(\frac{\widehat{\mathbf{H}}'\widehat{\mathbf{H}}}{T}\right)^{-1}\operatorname*{plim}_{T\to\infty}\frac{\widehat{\mathbf{H}}'\mathbf{u}_l}{T}$$

Call $\mathbf{D} \equiv \widehat{\mathbf{H}} - \mathbf{H} = \iota' \otimes (\widehat{\Lambda} - \Lambda)$, defining the row vector $\iota' \equiv (0, \cdots, 0, 1)$ taking the value of 1 for the last element and 0 otherwise. Note that,

$$\begin{aligned}\Lambda - \widehat{\Lambda} &= -\mathbf{J}(\overline{\Delta\beta}^*)(\widehat{\Delta\beta}^* - \Delta\beta_0^*) \\ \mathbf{D} &= \iota' \otimes \mathbf{J}(\overline{\Delta\beta}^*)(\widehat{\Delta\beta}^* - \Delta\beta_0^*) \\ \mathbf{u}_l &= C_l\left(\Lambda - \widehat{\Lambda}\right) + \mathbf{v}_l \\ &= -C_l\mathbf{J}(\overline{\Delta\beta}^*)(\widehat{\Delta\beta}^* - \Delta\beta_0^*) + \mathbf{v}_l.\end{aligned}$$

Given assumptions (i) and (ii),

$$\begin{aligned}\frac{1}{T}\mathbf{H}'\mathbf{H} &= O_p(1) \\ \frac{1}{T}\mathbf{H}'\mathbf{J}(\overline{\Delta\beta}^*) &= O_p(1) \\ \frac{1}{T}\mathbf{J}'(\overline{\Delta\beta}^*)\mathbf{J}(\overline{\Delta\beta}^*) &= O_p(1)\end{aligned}$$

and by construction $\mathbf{v}_l$ is a martingale difference sequence with respect to information set $\Im_{t-1}$, and conditioning in the observability restriction $y_{lt} \leq y_{ut}$,

$$\begin{aligned}E(v_{lt}|\Im_{t-1}) &= E\{\varepsilon_{lt} - E[\varepsilon_{lt}|\Delta\varepsilon_t \geq \sigma_m\Delta(y^{t-1}, \Delta\beta^*)]|\Im_{t-1}\} \\ &= E(\varepsilon_{lt}|\Im_{t-1}) - C_l\lambda_{t-1} = 0\end{aligned}$$

and thus we have $E(\mathbf{h}_{t-1}v_{lt}|\Im_{t-1}) = 0$ for all t. Given assumptions (iii) and (iv), and by the central limit theorem for martingale difference sequence, we have

$$\begin{aligned}
\frac{1}{\sqrt{T}}\mathbf{J}'(\overline{\Delta\beta}^*)\mathbf{v}_l &= O_p(1) \\
\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{v}_l &\xrightarrow{d} N(0,\mathbf{\Psi}_l).
\end{aligned}$$

Then, we prove that

$$\begin{aligned}
\frac{\widehat{\mathbf{H}}'\widehat{\mathbf{H}}}{T} &= O_p(1), \\
\frac{\widehat{\mathbf{H}}'\mathbf{u}_L}{T} &= O_p(T^{-1/2}).
\end{aligned}$$

That is,

$$\begin{aligned}
\frac{1}{T}\widehat{\mathbf{H}}'\widehat{\mathbf{H}} &= \frac{1}{T}(\mathbf{H}'+\mathbf{D}')(\mathbf{H}+\mathbf{D}) \\
&= \frac{1}{T}\mathbf{H}'\mathbf{H} \\
&\quad +\frac{1}{T}\iota'\otimes\mathbf{H}'\mathbf{J}(\overline{\Delta\beta}^*)(\widehat{\Delta\beta}^*-\Delta\beta_0^*)+\left(\frac{1}{T}\iota'\otimes\mathbf{H}'\mathbf{J}(\overline{\Delta\beta}^*)(\widehat{\Delta\beta}^*-\Delta\beta_0^*)\right)' \\
&\quad +\iota\iota'\otimes(\widehat{\Delta\beta}^*-\Delta\beta_0^*)'\frac{\mathbf{J}'(\overline{\Delta\beta}^*)\mathbf{J}(\overline{\Delta\beta}^*)}{T}(\widehat{\Delta\beta}^*-\Delta\beta_0^*) \\
&= O_p(1)+O_p(T^{-1/2})+O_p(T^{-1}) \\
&\xrightarrow{p} \frac{1}{T}\mathbf{H}'\mathbf{H},
\end{aligned} \tag{A.6}$$

and

$$\begin{aligned}
\frac{\widehat{\mathbf{H}}'\mathbf{u}_l}{T} &= \frac{1}{T}(\mathbf{H}'+\mathbf{D}')[C_l(\Lambda-\widehat{\Lambda})+\mathbf{v}_l] \\
&= \frac{1}{T}\mathbf{H}'\mathbf{v}_L - C_l\frac{\mathbf{H}'\mathbf{J}(\overline{\Delta\beta}^*)}{T}(\widehat{\Delta\beta}^*-\Delta\beta_0^*) \\
&\quad -C_l\iota\otimes(\widehat{\Delta\beta}^*-\Delta\beta_0^*)'\frac{\mathbf{J}'(\overline{\Delta\beta}^*)\mathbf{J}(\overline{\Delta\beta}^*)}{T}(\widehat{\Delta\beta}^*-\Delta\beta_0^*) \\
&\quad +\iota\otimes(\widehat{\Delta\beta}^*-\Delta\beta_0^*)\frac{\mathbf{J}'(\overline{\Delta\beta}^*)\mathbf{v}_l}{T} \\
&= O_p(T^{-1/2})+O_p(T^{-1/2})+O_p(T^{-1})+O_p(T^{-1}) \\
&= O_p(T^{-1/2}) \\
&\xrightarrow{p} \frac{1}{T}\mathbf{H}'\mathbf{v}_l - C_l\frac{\mathbf{H}'\mathbf{J}(\overline{\Delta\beta}^*)}{T}(\widehat{\Delta\beta}^*-\Delta\beta_0^*)
\end{aligned} \tag{A.7}$$

Hence,

$$\widehat{\gamma}_l - \gamma_l = O_p(1)O_p(T^{-1/2})$$
$$= O_p(T^{-1/2})$$

Therefore, the two-step estimator $\widehat{\gamma}_L$ is consistent, i.e., $\text{plim}_{T\to\infty}\widehat{\gamma}_l = \gamma_l$

(b) Asymptotic Normality

Now we consider the asymptotic distribution of the two-step estimator,

$$\sqrt{T}(\widehat{\gamma}_l - \gamma_l) = \left(\frac{\widehat{\mathbf{H}}'\widehat{\mathbf{H}}}{T}\right)^{-1}\frac{\widehat{\mathbf{H}}'\mathbf{u}_l}{\sqrt{T}}$$

From equation (A.6) and (A.7), we have,

$$\sqrt{T}(\widehat{\gamma}_l - \gamma_l) \xrightarrow{p} \left(\frac{\mathbf{H}'\mathbf{H}}{T}\right)^{-1}\left(\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{v}_l - C_l\frac{\mathbf{H}'\mathbf{J}(\overline{\Delta\beta}^*)}{T}\sqrt{T}(\widehat{\Delta\beta}^* - \Delta\beta_0^*)\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{B}\mathbf{\Xi}_l\mathbf{B}')$$

where

$$\mathbf{B} = \underset{T\to\infty}{\text{plim}}\left(\frac{\mathbf{H}'\mathbf{H}}{T}\right)^{-1}$$

and

$$\begin{aligned}
&\text{var}\left(\frac{1}{\sqrt{T}}\mathbf{H}'\mathbf{v}_l + C_l\frac{\mathbf{H}'(\Lambda - \widehat{\Lambda})}{\sqrt{T}}\right)\\
= &\frac{1}{T}E(\mathbf{H}'\mathbf{v}_l\mathbf{v}_l'\mathbf{H}) + C_l^2E\left(\frac{\mathbf{H}'\mathbf{J}(\overline{\beta}^*)}{T}\widehat{\mathbf{S}}\frac{\mathbf{J}'(\overline{\beta}^*)\mathbf{H}}{T}\right) + E\left(\frac{\mathbf{H}'\mathbf{v}_l(\Lambda - \widehat{\Lambda})'\mathbf{H}C_l}{T}\right)\\
&+E\left(\frac{\mathbf{H}'(\Lambda - \widehat{\Lambda})\mathbf{v}_l'\mathbf{H}C_l}{T}\right)\\
= &\frac{1}{T}\sum_{t=1}^{T}E(\mathbf{h}_{t-1}\mathbf{h}_{t-1}'v_{lt}^2) + C_l^2\overline{\mathbf{Q}}'\widehat{\mathbf{S}}\overline{\mathbf{Q}} + \mathbf{M}_{lT} + \mathbf{M}_{lT}'\\
\xrightarrow{p} &\mathbf{\Psi}_l + C_l^2\mathbf{Q}_0'\mathbf{S}_0\mathbf{Q}_0 + \mathbf{M}_{l0} + \mathbf{M}_{l0}'\\
\equiv &\mathbf{\Xi}_l
\end{aligned}$$

where the second equality holds because $\mathbf{h}_{t-1}v_{Lt}$ is a martingale difference sequence.

(c) Estimation of Unconditional Variances and Correlation Coefficient $(\sigma_l^2, \sigma_u^2, \rho)$

In the two-step estimation procedure, $\Delta\beta$ and σ_m are consistently estimated in the first step. The first and second moments of ε_u and ε_l, conditioning on the observability restriction, can be written as $\widetilde{m}_{10}, \widetilde{m}_{01}, \widetilde{m}_{20}, \widetilde{m}_{02}, \widetilde{m}_{11}$, by plugging the estimates $\widehat{\Delta} \equiv \Delta(y^{t-1}, \widehat{\Delta\beta})$ and $\widehat{\sigma}_m$ into (A.1) – (A.5). Therefore, the parameters $(\sigma_l^2, \sigma_u^2, \rho)$ can be estimated by the simple method of moments as follows

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{lt}^{2} = \frac{1}{T}\sum_{t=1}^{T}\{\widetilde{m}_{20}(t) - [\widetilde{m}_{10}(t)]^{2}\}$$
$$\frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{ut}^{2} = \frac{1}{T}\sum_{t=1}^{T}\{\widetilde{m}_{02}(t) - [\widetilde{m}_{01}(t)]^{2}\}$$
$$\frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{lt}\widehat{u}_{ut} = \frac{1}{T}\sum_{t=1}^{T}\{\widetilde{m}_{11}(t) - [\widetilde{m}_{10}(t)\widetilde{m}_{01}(t)]\}$$

where $\widehat{u}_{lt}$ and $\widehat{u}_{ut}$ are the residuals of the second step regression,

$$\widehat{u}_{lt} = y_{lt} - \widehat{\beta}_{lc} - \sum_{j=1}^{p}\widehat{\beta}_{11}^{(j)}y_{l,t-j} - \sum_{j=1}^{p}\widehat{\beta}_{12}^{(j)}y_{u,t-j} - \widehat{C_l}\widehat{\lambda}_{t-1}$$
$$\widehat{u}_{ut} = y_{ut} - \widehat{\beta}_{uc} - \sum_{j=1}^{p}\widehat{\beta}_{21}^{(j)}y_{l,t-j} - \sum_{j=1}^{p}\widehat{\beta}_{22}^{(j)}y_{u,t-j} - \widehat{C_u}\widehat{\lambda}_{t-1}$$

Proof of Proposition 1 (section 3.6 of the article)

We need to prove (A.8) – (A.12) as follows,

$$\sum_{t=1}^{T}\widehat{\Delta v}_{t}^{2}/T \xrightarrow{p} E(\text{var}(\Delta v_t|y^{t-1})), \tag{A.8}$$
$$\sum_{t=1}^{T}\widehat{u}_{lt}^{2}/T \xrightarrow{p} E(\text{var}(v_{lt}|y^{t-1})), \tag{A.9}$$
$$\sum_{t=1}^{T}\widehat{u}_{ut}^{2}/T \xrightarrow{p} E(\text{var}(v_{ut}|y^{t-1})), \tag{A.10}$$
$$\widehat{\sigma}_{l}^{2}(C_l) \equiv C_l^2(1 - \sum_{t=1}^{T}\widehat{\Delta v}_{t}^{2}/T\widehat{\sigma}_{m}^{2}) + \sum_{t=1}^{T}\widehat{u}_{lt}^{2}/T \xrightarrow{p} \sigma_l^2, \tag{A.11}$$
$$\widehat{\sigma}_{u}^{2}(C_u) \equiv C_u^2(1 - \sum_{t=1}^{T}\widehat{\Delta v}_{t}^{2}/T\widehat{\sigma}_{m}^{2}) + \sum_{t=1}^{T}\widehat{u}_{ut}^{2}/T \xrightarrow{p} \sigma_u^2. \tag{A.12}$$

Note that we only need to prove (A.8), (A.9), and (A.11). Others can be proved similarly.

(a) Proof of (A.8):

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\widehat{\Delta v_t}^2 &= \frac{1}{T}\sum_{t=1}^{T}(\Delta y_t + z_{t-1}\widehat{\Delta\beta} - \widehat{\sigma}_m\widehat{\lambda}_{t-1})^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\left[\Delta v_t + z_{t-1}(\widehat{\Delta\beta} - \Delta\beta) + \widehat{\sigma}_m(\lambda_{t-1} - \widehat{\lambda}_{t-1}) + \lambda_{t-1}(\sigma_m - \widehat{\sigma}_m)\right]^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\Delta v_t^2 + \frac{1}{T}\sum_{t=1}^{T} z_{t-1}(\widehat{\Delta\beta} - \Delta\beta)(\widehat{\Delta\beta} - \Delta\beta)' z_{t-1}' \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\widehat{\sigma}_m^2(\lambda_{t-1} - \widehat{\lambda}_{t-1})^2 + \frac{1}{T}\sum_{t=1}^{T}\lambda_{t-1}^2(\sigma_m - \widehat{\sigma}_m)^2 \\
&\quad + \frac{2}{T}\sum_{t=1}^{T} z_{t-1}(\widehat{\Delta\beta} - \Delta\beta)\Delta v_t + \frac{2}{T}\sum_{t=1}^{T}\widehat{\sigma}_m(\lambda_{t-1} - \widehat{\lambda}_{t-1})\Delta v_t \\
&\quad + \frac{2}{T}\sum_{t=1}^{T}\lambda_{t-1}(\sigma_m - \widehat{\sigma}_m)\Delta v_t + \frac{2}{T}\sum_{t=1}^{T} z_{t-1}(\widehat{\Delta\beta} - \Delta\beta)(\lambda_{t-1} - \widehat{\lambda}_{t-1})\widehat{\sigma}_m \\
&\quad + \frac{2}{T}\sum_{t=1}^{T} z_{t-1}(\widehat{\Delta\beta} - \Delta\beta)\lambda_{t-1}(\sigma_m - \widehat{\sigma}_m) + \frac{2}{T}\sum_{t=1}^{T}\widehat{\sigma}_m(\lambda_{t-1} - \widehat{\lambda}_{t-1})\lambda_{t-1}(\sigma_m - \widehat{\sigma}_m)
\end{aligned}$$

In the above expression, the *first* term is

$$\frac{1}{T}\sum_{t=1}^{T}\Delta v_t^2 \xrightarrow{p} \frac{1}{T}\sum_{t=1}^{T}\mathrm{var}(\Delta v_t|y^{t-1}) \xrightarrow{p} E(\mathrm{var}(\Delta v_t|y^{t-1})). \tag{A.13}$$

In (A.13), the first convergence in probability is because of the Law of Large Numbers for mixing sequence. The second convergence in probability follows because of the ergodic theorem, since the assumptions on the stationarity and mixing properties of $\{Y_t\}$ imply its ergodicity. Therefore, we only need to prove that the rest of the terms in the summation converges to zero in probability. In the rest of the proof, we will be using the following property extensively $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$.

For the *second* term,

$$\text{vec}\left(\frac{1}{T}\sum_{t=1}^{T} z_{t-1}(\widehat{\Delta\beta}-\Delta\beta)(\widehat{\Delta\beta}-\Delta\beta)' z'_{t-1}\right) = \frac{1}{T}\sum_{t=1}^{T}(z'_{t-1}\otimes z_{t-1})\text{vec}\left((\widehat{\Delta\beta}-\Delta\beta)(\widehat{\Delta\beta}-\Delta\beta)'\right) \xrightarrow{p} 0$$

since

$$\frac{1}{T}\sum_{t=1}^{T}(z_{t-1}\otimes z_{t-1}) = O_p(1)$$

$$\widehat{\Delta\beta}-\Delta\beta \xrightarrow{p} 0$$

because of assumption (i) in Theorem 2 and result (a) in Theorem 1.

For the *third* term,

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{\sigma}_m^2(\lambda_{t-1}-\widehat{\lambda}_{t-1})^2 = \widehat{\sigma}_m^2\frac{1}{T}\sum_{t=1}^{T} j_{t-1}(\widehat{\Delta\beta}^* - \Delta\beta^*)(\widehat{\Delta\beta}^* - \Delta\beta^*)' j'_{t-1}$$

where j_{t-1} is the t-th row of Jacobian matrix $J(\Delta\beta^*)$, and therefore,

$$\begin{aligned} & \text{vec}\left(\frac{1}{T}\sum_{t=1}^{T} j_{t-1}(\widehat{\Delta\beta}^* - \Delta\beta^*)(\widehat{\Delta\beta}^* - \Delta\beta^*)' j'_{t-1}\right) \\ = \ & \frac{1}{T}\sum_{t=1}^{T} j_{t-1}\otimes j_{t-1}\text{vec}\left((\widehat{\Delta\beta}^* - \Delta\beta^*)(\widehat{\Delta\beta}^* - \Delta\beta^*)'\right) \\ \xrightarrow{p} \ & 0, \end{aligned}$$

given that

$$\frac{1}{T}\sum_{t=1}^{T}(j_{t-1}\otimes j_{t-1}) = O_p(1)$$

$$\widehat{\Delta\beta}^* - \Delta\beta^* \xrightarrow{p} 0$$

$$\widehat{\sigma}_m^2 \xrightarrow{p} \sigma_m^2$$

because of assumption (ii) in Theorem 2 and result (a) in Theorem 1.

The proofs for the rest terms are omitted here, since similar proof technique applies to the rest of the summation terms. Their convergence to zero in probability relies on assumptions in Theorem 2 and the results in Theorem 1. Therefore, we prove (A.8),

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{\Delta v_t}^2 \xrightarrow{p} \frac{1}{T}\sum_{t=1}^{T}\Delta v_t^2 \xrightarrow{p} E(\text{var}(\Delta v_t|y^{t-1})).$$

(b) Proof of (A.9).

Let β_l^c denote $\beta_l(C_l)$.

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{lt}^2 &= \frac{1}{T}\sum_{t=1}^{T}(y_{lt} - z_{t-1}\beta_l^c - C_l\widehat{\lambda}_{t-1})^2\\ &= \frac{1}{T}\sum_{t=1}^{T}\left[z_{t-1}(\beta_l - \beta_l^c) + C_l(\lambda_{t-1} - \widehat{\lambda}_{t-1}) + v_{lt}\right]^2\\ &= \frac{1}{T}\sum_{t=1}^{T}v_{lt}^2 + \frac{1}{T}\sum_{t=1}^{T}z_{t-1}(\beta_l - \beta_l^c)(\beta_l - \beta_l^c)'z_{t-1}' + \frac{1}{T}\sum_{t=1}^{T}C_l^2(\lambda_{t-1} - \widehat{\lambda}_{t-1})^2\\ &= \frac{1}{T}\sum_{t=1}^{T}z_{t-1}(\beta_l - \beta_l^c)v_{lt} + \frac{1}{T}\sum_{t=1}^{T}z_{t-1}(\beta_l - \beta_l^c)(\lambda_{t-1} - \widehat{\lambda}_{t-1})C_l + \frac{1}{T}\sum_{t=1}^{T}C_l(\lambda_{t-1} - \widehat{\lambda}_{t-1})v_{lt}.\end{aligned}$$

In the above expression, for the first term in the summation, we have

$$\frac{1}{T}\sum_{t=1}^{T}v_{lt}^2 \xrightarrow{p} \frac{1}{T}\sum_{t=1}^{T}\text{var}(v_{lt}|y^{t-1}) \xrightarrow{p} E(\text{var}(v_{lt}|y^{t-1}))$$

because of the Law of Large Numbers for mixing sequences and the ergodic theorem for $\{Y_t\}$. The rest of the terms in the summation converges to zero in probability by similar arguments as those in the proof of (A.8).

(c) Proof of (A.11).

Given σ_l^2, (A.8), (A.9), and the continuous mapping theorem, (A.11) holds.

A2. (Section 4. Simulation)

Table 1: Simulation Results for Case B-1 and Case B-3

(a) Simulation Results for Case B-1

	Small Sample Size ($T=250$)						Large Sample Size ($T=2000$)					
	OLS		Two-step		Modified Two-step		OLS		Two-step		Modified Two-step	
Parameters	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse
$\beta_{lc}=0$	−0.4957	0.5061	−0.0884	0.7240	−0.0241	0.1797	−0.4863	0.4875	−0.0141	0.2282	−0.0056	0.0653
$\beta_{uc}=0$	0.4979	0.5087	0.0938	0.6975	0.0295	0.1804	0.4854	0.4867	0.0226	0.2335	0.0029	0.0643
$\beta_{11}=0.8$	0.6645	0.1450	0.7632	0.1813	0.7792	0.0684	0.6793	0.1222	0.7950	0.0587	0.7971	0.0240
$\beta_{12}=0.1$	0.2204	0.1316	0.1215	0.1795	0.1065	0.0662	0.2191	0.1206	0.1036	0.0583	0.1015	0.0237
$\beta_{21}=0.1$	0.2191	0.1317	0.1208	0.1747	0.1052	0.0700	0.2191	0.1207	0.1059	0.0596	0.1009	0.0241
$\beta_{22}=0.8$	0.6637	0.1468	0.7607	0.1770	0.7767	0.0706	0.6788	0.1227	0.7920	0.0600	0.7968	0.0240
$C_l=-0.7071$			−0.5986	1.1124	−0.6992	0.0814			−0.6927	0.3218	−0.7052	0.0293
$C_u=0.7071$			0.6050	1.0984	0.6959	0.0776			0.6797	0.3352	0.7080	0.0278
$\sigma_l^2=1$	0.7918	0.2207	0.9811	0.1231	0.9844	0.1240	0.7965	0.2051	0.9957	0.0445	0.9956	0.0437
$\sigma_u^2=1$	0.7891	0.2232	0.9763	0.1206	0.9792	0.1200	0.7988	0.2028	0.9996	0.0430	0.9994	0.0423
$\rho=0$	0.2575	0.2653	0.0141	0.0875	0.0067	0.0884	0.2498	0.2509	−0.0009	0.0334	−0.0013	0.0328

(b) Simulation Results for Case B-3

	Small Sample Size ($T=250$)						Large Sample Size ($T=2000$)					
	OLS		Two-step		Modified Two-step		OLS		Two-step		Modified Two-step	
Parameters	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse
$\beta_{lc}=0$	−0.5715	0.5782	1.8196	198.5	−0.0080	0.2674	−0.5639	0.5648	−71.45	2069	−0.0004	0.0851
$\beta_{uc}=0$	0.5642	0.5711	−1.4060	147.8	0.0022	0.2598	0.5663	0.5672	−100.7	2825	0.0013	0.0843
$\beta_{11}=0.1$	0.0764	0.0744	0.1339	3.6156	0.0843	0.1197	0.0840	0.0300	0.0920	2.9502	0.0986	0.0410
$\beta_{12}=0.05$	0.0663	0.0745	0.0550	2.9242	0.0589	0.1221	0.0651	0.0295	−0.1036	3.3209	0.0507	0.0420
$\beta_{21}=0.05$	0.0646	0.0733	0.0493	2.9074	0.0569	0.1205	0.0659	0.0296	−0.0284	2.7397	0.0513	0.0403
$\beta_{22}=0.1$	0.0782	0.0750	0.0475	2.6718	0.0858	0.1189	0.0829	0.0302	0.0355	3.6375	0.0973	0.0412
$C_l=-1/\sqrt{2}$			−3.0891	230.0	−0.7011	0.1127			88.26	2576	−0.7061	0.0398
$C_u=1/\sqrt{2}$			1.3011	182.6	0.6987	0.1146			128.4	3510	0.7080	0.0400
$\sigma_l^2=1$	0.6806	0.3251	0.7884	0.2263	0.9939	0.1726	0.6852	0.3156	0.7954	0.2066	0.9987	0.0584
$\sigma_u^2=1$	0.6794	0.3262	0.7868	0.2273	0.9909	0.1713	0.6862	0.3145	0.7970	0.2050	1.0014	0.0586
$\rho=0$	0.4579	0.4613	0.2521	0.2600	0.0107	0.1300	0.4564	0.4568	0.2535	0.2545	0.0000	0.0454

Number of Simulation=1000, $1/\sqrt{2}\approx 0.7071$.

Table 2: Simulation Results for Case NB-1 and Case NB-3

(a) Simulation Results for Case NB-1

	Small Sample Size ($T=250$)						Large Sample Size ($T=2000$)					
	OLS		Two-step		Modified Two-step		OLS		Two-step		Modified Two-step	
Parameters	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse
$\beta_{lc}=-2$	−2.0890	0.4074	−2.1056	0.4345	−2.0883	0.4076	−2.0120	0.1551	−2.0135	0.1558	−2.0119	0.1551
$\beta_{uc}=2$	2.0825	0.3816	2.0907	0.4118	2.0819	0.3819	2.0263	0.1545	2.0265	0.1565	2.0262	0.1545
$\beta_{11}=0.8$	0.7842	0.0401	0.7830	0.0417	0.7843	0.0401	0.7978	0.0139	0.7977	0.0139	0.7978	0.0139
$\beta_{12}=0.1$	0.0975	0.0385	0.0988	0.0402	0.0975	0.0385	0.0995	0.0131	0.0996	0.0131	0.0995	0.0131
$\beta_{21}=0.1$	0.0983	0.0363	0.0989	0.0376	0.0983	0.0363	0.1006	0.0134	0.1006	0.0135	0.1006	0.0134
$\beta_{22}=0.8$	0.7850	0.0390	0.7844	0.0410	0.7850	0.0390	0.7967	0.0139	0.7967	0.0140	0.7967	0.0139
$C_l=-1/\sqrt{2}$			−161.5	9243	−0.7026	0.0553			2.8987	210.9267	−0.7067	0.0194
$C_u=1/\sqrt{2}$			−151.2	4418	0.7016	0.0541			1.1828	214.8670	0.7065	0.0189
$\sigma_l^2=1$	0.9921	0.0899	0.9842	0.0906	0.9823	0.0903	0.9997	0.0321	0.9987	0.0321	0.9984	0.0321
$\sigma_u^2=1$	0.9906	0.0885	0.9830	0.0893	0.9807	0.0891	0.9994	0.0300	0.9984	0.0300	0.9982	0.0300
$\rho=0$	−0.0004	0.0630	−0.0001	0.0632	−0.0066	0.0642	0.0002	0.0224	0.0002	0.0224	−0.0006	0.0225

(b) Simulation Results for Case NB-3

	Small Sample Size ($T=250$)						Large Sample Size ($T=2000$)					
	OLS		Two-step		Modified Two-step		OLS		Two-step		Modified Two-step	
Parameters	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse
$\beta_{lc}=-2$	−2.0165	0.1917	−1.7197	54.2693	−2.0115	0.1929	−2.0069	0.0722	−1.2760	28.02	−2.0021	0.0724
$\beta_{uc}=2$	2.0246	0.2021	2.6687	63.9572	2.0196	0.2033	2.0060	0.0688	0.6512	21.94	2.0011	0.0691
$\beta_{11}=0.1$	0.0945	0.0602	0.1137	0.6083	0.0948	0.0606	0.0991	0.0224	0.1031	0.2302	0.0994	0.0226
$\beta_{12}=0.05$	0.0517	0.0650	0.0402	0.7072	0.0514	0.0655	0.0508	0.0226	0.0450	0.2476	0.0504	0.0228
$\beta_{21}=0.05$	0.0498	0.0656	0.0385	0.8010	0.0495	0.0661	0.0504	0.0218	0.0406	0.2495	0.0500	0.0220
$\beta_{22}=0.1$	0.0897	0.0649	0.0779	0.6570	0.0900	0.0653	0.0989	0.0220	0.1080	0.2774	0.0992	0.0222
$C_l=-1/\sqrt{2}$			225.5	12212	−0.7022	0.0548			−169.4	6526	−0.7056	0.0196
$C_u=1/\sqrt{2}$			−185.8	15901	0.7038	0.0555			314.1	5409	0.7064	0.0199
$\sigma_l^2=1$	0.9843	0.0884	0.9839	0.0902	0.9817	0.0902	0.9911	0.0321	0.9972	0.0317	0.9970	0.0317
$\sigma_u^2=1$	0.9866	0.0881	0.9865	0.0901	0.9841	0.0898	0.9921	0.0320	0.9983	0.0319	0.9980	0.0319
$\rho=0$	0.0063	0.0598	−0.0014	0.0614	−0.0074	0.0626	0.0082	0.0243	0.0011	0.0236	0.0003	0.0236

Number of Simulation=1000

A3. (Section 5. Comparison with Existing Approaches)

Table 3: Methodology Evaluation for DGP2 (LOW persistence and BINDING observability restriction)

DGP2									
Multivariate Normal Distribution									
	RMSE		*CR & ER*			*MLF*		*MDE*	
	Lower	*Upper*	*CR*	*ER*	$\frac{CR+ER}{2}$	$p = 1$	$p = 2$	$q = 1$	$q = 2$
CCRM	1.0963	0.6885	0.7992	0.6583	0.7288	1.4169	1.6765	0.7869	0.9154
CRM	1.0954	0.6896	0.7989	0.6582	0.7285	1.4171	1.6760	0.7870	0.9153
TS	**1.0903**	**0.6809**	0.8025	0.6607	0.7316	**1.4062**	**1.6530**	**0.7801**	**0.9090**
MTS	**1.0906**	**0.6811**	0.8026	0.6606	0.7316	**1.4066**	**1.6539**	**0.7803**	**0.9092**
GARCH-N (99%)	1.1266	0.7353	0.8719	0.6061	0.7390	1.5261	1.8106	0.8437	0.9513
GARCH-N (99.5%)	1.1580	0.7825	0.8959	0.5826	**0.7392**	1.6093	1.9539	0.8848	0.9883
GARCH-T (99%)	1.1538	0.7756	0.8926	0.5859	**0.7393**	1.5975	1.9340	0.8790	0.9831
GARCH-T (99.5%)	1.2220	0.8736	0.9222	0.5497	0.7359	1.7631	2.2588	0.9600	1.0623
Multivariate Student's t Distribution ($\nu = 5$)									
	RMSE		*CR & ER*			*MLF*		*MDE*	
	Lower	*Upper*	*CR*	*ER*	$\frac{CR+ER}{2}$	$p = 1$	$p = 2$	$q = 1$	$q = 2$
CCRM	1.5667	0.9612	0.8074	0.6299	0.7187	1.8297	3.3843	1.0108	1.2998
CRM	**1.5655**	0.9627	0.8069	0.6297	0.7183	1.8303	**3.3834**	1.0110	**1.2997**
TS	**1.5583**	**0.9506**	0.8118	0.6330	**0.7224**	**1.8133**	**3.3377**	**0.9999**	**1.2909**
MTS	1.5718	**0.9606**	0.8105	0.6326	**0.7215**	**1.8205**	3.4105	**1.0034**	1.3027
GARCH-N (99%)	1.6448	1.0786	0.8976	0.5454	0.7215	2.1619	3.8773	1.1771	1.3909
GARCH-N (99.5%)	1.6967	1.1556	0.9154	0.5205	0.7179	2.3202	4.2235	1.2545	1.4516
GARCH-T (99%)	1.7500	1.2253	0.9261	0.5029	0.7145	2.4576	4.5762	1.3215	1.5107
GARCH-T (99.5%)	1.9388	1.4802	0.9511	0.4515	0.7013	2.9420	5.9745	1.5578	1.7249

Table 4: Methodology Evaluation for DGP4 (LOW persitence and NON-BINDING observability restriction)

DGP4									
Multivariate Normal Distribution									
	RMSE		*CR & ER*			*MLF*		*MDE*	
	Lower	*Upper*	*CR*	*ER*	$\frac{CR+ER}{2}$	$p = 1$	$p = 2$	$q = 1$	$q = 2$
CCRM	1.5523	0.9149	0.8544	0.7714	0.8129	1.9957	3.2475	1.0887	1.2742
CRM	1.5508	0.9168	0.8543	0.7713	0.8128	1.9959	3.2461	1.0887	1.2739
TS	**1.5448**	**0.9044**	0.8559	0.7728	**0.8144**	**1.9819**	**3.2052**	**1.0785**	**1.2658**
MTS	**1.5452**	**0.9047**	0.8559	0.7728	**0.8143**	**1.9824**	**3.2068**	**1.0788**	**1.2661**
GARCH-N (99%)	1.6669	1.0944	0.7208	0.8421	0.7814	2.1908	3.9774	1.1824	1.4101
GARCH-N (99.5%)	1.6217	1.0242	0.7591	0.8264	0.7928	2.1043	3.6798	1.1415	1.3563
GARCH-T (99%)	1.6594	1.0816	0.7276	0.8394	0.7835	2.1755	3.9249	1.1753	1.4007
GARCH-T (99.5%)	1.6000	0.9879	0.7817	0.8155	0.7986	2.0641	3.5370	1.1227	1.3297
Multivariate Student's t Distribution ($\nu = 5$)									
	RMSE		*CR & ER*			*MLF*		*MDE*	
	Lower	*Upper*	*CR*	*ER*	$\frac{CR+ER}{2}$	$p = 1$	$p = 2$	$q = 1$	$q = 2$
CCRM	1.8565	1.0996	0.8552	0.7584	0.8068	2.2450	4.6601	1.2265	1.5258
CRM	1.8548	1.1017	0.8551	0.7583	0.8067	2.2455	4.6582	1.2266	1.5255
TS	**1.8475**	**1.0870**	0.8572	0.7603	**0.8088**	**2.2266**	**4.5992**	**1.2131**	**1.5158**
MTS	**1.8480**	**1.0873**	0.8572	0.7603	**0.8087**	**2.2271**	**4.6015**	**1.2134**	**1.5162**
GARCH-N (99%)	1.9226	1.2020	0.7640	0.8142	0.7891	2.3202	5.1454	1.2626	1.6034
GARCH-N (99.5%)	1.8892	1.1479	0.8003	0.7951	0.7977	2.2663	4.8911	1.2377	1.5632
GARCH-T (99%)	1.8989	1.1546	0.7951	0.7978	0.7964	2.2774	4.9431	1.2432	1.5716
GARCH-T (99.5%)	1.8741	1.1119	0.8508	0.7591	0.8050	2.2674	4.7536	1.2398	1.5410

Table 5: Simulation Results of DGP2 and DGP4 with Multivariate Normal Errors

		DGP2 (*low persistence and binding O.R.*)						DGP4 (*low persistence and non-binding O.R.*)					
		b_{11}	b_{12}	b_{0L}	b_{21}	b_{22}	b_{0U}	b_{11}	b_{12}	b_{0L}	b_{21}	b_{22}	b_{0U}
	true value	-0.1	0.05	0	-0.05	0.1	0	-0.1	0.05	-2	-0.05	0.1	2
Mean	CCRM	-0.0516	-0.0588	-1.1988	-0.0588	-0.0516	0.8585	-0.0935	-0.0976	-1.8423	-0.0976	-0.0935	2.5035
	CRM	-0.0577	-0.0527	-1.2114	-0.0527	-0.0577	0.8710	**-0.1016**	-0.0895	-1.8780	-0.0895	-0.1016	2.5392
	TS	-0.0776	0.2367	26.5814	-0.0108	0.0674	-19.5935	-0.1671	-0.0905	-58.2921	**-0.0550**	0.1529	8.2993
	MTS	**-0.1005**	**0.0517**	**0.0072**	**-0.0499**	**0.0991**	**-0.0048**	**-0.1000**	**0.0527**	**-2.0082**	**-0.0505**	**0.0980**	**2.0045**
Bias^2	CCRM	0.0023	0.0118	1.4372	**0.0001**	0.0230	0.7370	**0.0000**	0.0218	0.0249	0.0023	0.0374	0.2536
	CRM	0.0018	0.0106	1.4675	**0.0000**	0.0249	0.7587	**0.0000**	0.0194	0.0149	0.0016	0.0406	0.2907
	TS	**0.0005**	0.0348	706.57	0.0015	0.0011	383.90	0.0045	0.0197	3168.8	**0.0000**	**0.0028**	39.681
	MTS	**0.0000**	**0.0000**	**0.0001**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0001**	**0.0000**	**0.0000**	**0.0000**
MSE	CCRM	**0.0025**	**0.0120**	1.4379	**0.0002**	0.0232	0.7374	**0.0002**	0.0219	0.0261	0.0024	0.0376	0.2545
	CRM	**0.0021**	**0.0107**	1.4686	**0.0002**	0.0252	0.7595	**0.0004**	0.0196	**0.0181**	0.0017	0.0410	0.2939
	TS	11.72	41.81	2.4e+06	4.641	10.58	4.5e+05	4.722	10.59	1.76e+06	0.6072	1.609	2.77e+04
	MTS	0.0051	**0.0118**	**0.0366**	0.0015	**0.0034**	**0.0112**	0.0017	**0.0044**	**0.0165**	**0.0005**	**0.0015**	**0.0055**

A4. (Section 6. Empirical Illustration:SP500 Low/High Return Interval)

Table 6: Simulation Results of DGP2 and DGP4 with Multivariate Student-*t* Errors

		DGP2 (*low persistence and binding O.R.*)						DGP4 (*low persistence and non-binding O.R.*)					
		b_{11}	b_{12}	b_{0L}	b_{21}	b_{22}	b_{0U}	b_{11}	b_{12}	b_{0L}	b_{21}	b_{22}	b_{0U}
	true value	-0.1	0.05	0	-0.05	0.1	0	-0.1	0.05	-2	-0.05	0.1	2
Mean	CCRM	-0.0613	-0.0685	-1.4316	-0.0685	-0.0613	1.0178	-0.0886	-0.0931	**-2.0091**	-0.0931	-0.0886	2.6154
	CRM	**-0.0671**	-0.0627	-1.4459	**-0.0627**	-0.0671	1.0322	**-0.0962**	-0.0855	**-2.0441**	-0.0855	-0.0962	2.6504
	TS	-0.1958	-0.2252	188.95	0.1240	0.3712	-4.9869	-0.1519	-0.0599	-22.722	**-0.0417**	**0.1063**	2.8518
	MTS	-0.2389	-0.0594	17.5335	0.0293	**0.1663**	-9.8989	**-0.1031**	**0.0475**	-1.8869	**-0.0487**	**0.1001**	**1.9402**
Bias^2	CCRM	0.0015	0.0140	2.0494	0.0003	0.0260	1.0360	0.0001	0.0205	**0.0001**	0.0019	0.0356	0.3787
	CRM	**0.0011**	0.0127	2.0906	**0.0002**	0.0279	1.0653	**0.0000**	0.0184	0.0019	0.0013	0.0385	0.4230
	TS	0.0092	0.0757	3.57e+04	0.0303	0.0735	24.8691	0.0027	0.0121	429.39	**0.0001**	**0.0000**	0.7256
	MTS	0.0193	**0.0120**	307.43	0.0063	**0.0044**	97.9887	**0.0000**	**0.0000**	0.0128	**0.0000**	**0.0000**	**0.0036**
MSE	CCRM	0.0017	0.0142	2.0508	0.0005	**0.0263**	1.0368	**0.0003**	0.0206	**0.0020**	0.0020	0.0358	0.3801
	CRM	**0.0015**	**0.0129**	2.0925	**0.0003**	0.0283	1.0667	**0.0004**	0.0185	0.0058	0.0014	0.0388	0.4267
	TS	70.43	197.8	8.0e+07	39.10	88.69	6.3e+07	4.7210	7.3450	1.5e+06	0.9674	2.0540	1.8e+05
	MTS	0.2874	0.7468	753.22	0.1046	0.2813	256.81	0.0021	**0.0061**	0.0488	**0.0007**	**0.0019**	**0.0146**

ACCEPTED MANUSCRIPT

Table 7: Descriptive Statistics for Stable and Unstable Period

	Stable Period				Unstable Period			
Statistics	low	high	center	radius	low	high	center	radius
Minimum	−1.9170	−0.0271	−0.9721	0.1195	−9.4210	−1.3010	−4.9190	0.1463
1st Quartile	−0.6570	0.1600	−0.2466	0.3092	−1.3330	0.1769	−0.4698	0.4671
Median	−0.3295	0.3944	0.0239	0.4188	−0.6204	0.5889	−0.0127	0.7106
3rd Quartile	−0.0817	0.6773	0.2660	0.5498	−0.1498	1.1620	0.4170	1.1040
Maximum	0.1375	2.3250	1.1610	1.1640	1.5050	11.9800	6.7410	5.6090
Mean	−0.4303	0.4655	0.0176	0.4479	−0.9593	0.8593	−0.0500	0.9093
Variance	0.1726	0.1609	0.1317	0.0351	1.6539	1.3347	0.9687	0.5256
Correlation	0.5797		−0.0530		0.2982		−0.1118	
Skewness	−1.1166	1.1397	−0.0501	0.8604	−2.6056	3.1601	−0.0715	2.7765
Kurtosis	3.8698	4.5562	2.8328	3.5695	13.2002	21.3206	9.1251	13.7251

Table 8: First Step Estimation for Stable Period

Truncated Normal Regression		
regressor	coefficient ($\Delta\beta^*$)	s.e.
const	−1.7655	0.1407
$r_{L,t-1}$	0.4315	0.1116
$r_{U,t-1}$	0.0201	0.1172
$r_{L,t-2}$	0.7010	0.1126
$r_{U,t-2}$	−0.2849	0.1150
σ_m	0.3699	0.0107
Time Span: 2004/1/1 – 2007/1/1		
Number of Observations: 756		

In the stationary block bootstrap, the block size follows a Geometric distribution with mean equal to b. To choose the optimal block size b, we follow the method proposed by Politis and White (2004) and Patton, Politis and White (2009). The optimal value of b minimizes the $MSE(\hat{\sigma}_b^2)$ with $\sigma_\infty^2 = \sum_{s=-\infty}^{\infty} R(s)$, where $R(s)$ is the auto-covariance function. This procedure considers only the bootstrapping for a scalar time series, however with interval time series we need to jointly bootstrap a 2×1 vector time series $\{(y_{lt}, y_{ut})\}_{t=0}^T$. We proceed by selecting separately the optimal block sizes b_l and b_u for the lower bound $\{y_{lt}\}_{t=0}^T$ and the upper bound $\{y_{ut}\}_{t=0}^T$ series respectively. Then, we use the average $(b_l + b_u)/2$ as the unified block size length to bootstrap the vector sequence $\{(y_{lt}, y_{ut})\}_{t=0}^T$.

ACCEPTED MANUSCRIPT

Table 9: First Step Estimation for Unstable Period

Truncated Normal Regression		
regressor	coefficient ($\Delta\beta^*$)	s.e.
const	0.0778	0.0731
$r_{L,t-1}$	0.3312	0.0396
$r_{U,t-1}$	0.0121	0.045
$r_{L,t-2}$	0.3364	0.0412
$r_{U,t-2}$	−0.1188	0.0451
$r_{L,t-3}$	0.2356	0.0424
$r_{U,t-3}$	0.0112	0.0448
$r_{L,t-4}$	0.2978	0.0426
$r_{U,t-4}$	−0.1404	0.0428
$r_{L,t-5}$	0.2874	0.0434
$r_{U,t-5}$	−0.1819	0.0404
σ_m	0.9534	0.0270
Time Span: 2007/1/1 – 2011/4/29		
Number of Observations: 1009		

Table 10: Block Sizes for Stationary Block Bootstrapping

Block Size	Stable Period	Unstable Period
b_l	1.8055	53.2275
b_u	2.5007	53.4241
$b = (b_l + b_u)/2$	2.1531	53.3258

We report the optimal block sizes for both periods in Table 10.

ACCEPTED MANUSCRIPT

Table 11: Lower/Upper Bound Regression Results of Three Models (Stable Period)

Lower Bound Regression (Dependent variable: r_{lt})						
	OLS		Two-step		Modified Two-step	
regressors	coefficient	s.e.	coefficient	s.e.	coefficient	s.e.
const	−0.3062	(0.0509)	−0.0729	(0.1400)	−0.2947	(0.0499)
r_{lt-1}	0.0586	(0.0451)	0.1400	(0.0568)	0.0626	(0.0461)
r_{ut-1}	−0.0283	(0.0470)	−0.0295	(0.0413)	−0.0283	(0.0444)
r_{lt-2}	0.1100	(0.0453)	0.2408	(0.0920)	0.1165	(0.0462)
r_{ut-2}	−0.0840	(0.0463)	−0.1410	(0.0606)	−0.0868	(0.0472)
$\widehat{\lambda}_{t-1}$			−4.1883	(2.3541)	−0.2069	(0.0221)
Degree of Freedom	749		748		749	
S.E. of Regression	0.4143		0.4139		0.4143	
Adjusted R^2	0.5208		0.5217		0.5147	
F-statistic	164.9		138.1		160.9	
Upper Bound Regression (Dependent variable: r_{ht})						
	OLS		Two-step		Modified Two-step	
regressors	coefficient	s.e.	coefficient	s.e.	coefficient	s.e.
const	0.3674	(0.0485)	0.8233	(0.1363)	0.3583	(0.0486)
r_{lt-1}	−0.0938	(0.0430)	0.0652	(0.0592)	−0.0970	(0.0430)
r_{ut-1}	−0.0209	(0.0448)	−0.0233	(0.0381)	−0.0209	(0.0436)
r_{lt-2}	−0.1377	(0.0432)	0.1179	(0.0898)	−0.1428	(0.0468)
r_{ut-2}	0.0163	(0.0441)	−0.0950	(0.0566)	0.0185	(0.0441)
$\widehat{\lambda}_{t-1}$			−8.1843	(2.1703)	0.1630	(0.0218)
Degree of Freedom	749		748		749	
S.E. of Regression	0.3954		0.393		0.3955	
Adjusted R^2	0.5846		0.5896		0.5802	
F-statistic	213.2		181.6		209.4	
σ_l^2	0.1708		0.4156		0.1730	
σ_u^2	0.1555		0.3934		0.1568	
ρ	0.6070		0.5848		0.5860	

Table 12: Lower/Upper Bound Regression Results of Three Models (Unstable Period)

Lower Bound Regression (Dependent variable: r_{lt})						
	OLS		Two-step		Modified Two-step	
regressors	coefficient	s.e.	coefficient	s.e.	coefficient	s.e.
const	−0.1281	(0.0649)	−0.0951	(0.1162)	0.1275	(0.1583)
r_{lt-1}	0.0434	(0.0425)	0.0456	(0.0568)	0.0601	(0.0611)
r_{ut-1}	0.0420	(0.0492)	0.0435	(0.0467)	0.0533	(0.0572)
r_{lt-2}	0.1025	(0.0443)	0.1055	(0.0660)	0.1260	(0.0545)
r_{ut-2}	−0.1075	(0.0486)	−0.1080	(0.0633)	−0.1109	(0.0583)
r_{lt-3}	0.1346	(0.0458)	0.1371	(0.0529)	0.1537	(0.0463)
r_{ut-3}	0.0094	(0.0480)	0.0102	(0.0785)	0.0159	(0.0894)
r_{lt-4}	0.1751	(0.0462)	0.1784	(0.0836)	0.2010	(0.0969)
r_{ut-4}	−0.1581	(0.0461)	−0.1595	(0.0939)	−0.1690	(0.1090)
r_{lt-5}	0.1202	(0.0469)	0.1230	(0.0582)	0.1423	(0.0615)
r_{ut-5}	−0.1100	(0.0436)	−0.1117	(0.0584)	−0.1232	(0.0594)
$\widehat{\lambda}_{t-1}$			−0.0874	(0.2263)	−0.6775	(0.1413)
Degree of Freedom	993		992		993	
S.E. of Regression	1.142		1.142		1.144	
Adjusted R^2	0.4953		0.4948		0.4646	
F-statistic	90.57		82.96		80.2	
Upper Bound Regression (Dependent variable: r_{ht})						
	OLS		Two-step		Modified Two-step	
regressors	coefficient	s.e.	coefficient	s.e.	coefficient	std. err.
const	0.1210	(0.0562)	−0.0469	(0.0896)	0.0169	(0.0875)
r_{lt-1}	−0.2563	(0.0369)	−0.2673	(0.0483)	−0.2631	(0.0637)
r_{ut-1}	0.0622	(0.0426)	0.0548	(0.0403)	0.0576	(0.0516)
r_{lt-2}	−0.1839	(0.0384)	−0.1994	(0.0522)	−0.1935	(0.0518)
r_{ut-2}	0.0191	(0.0422)	0.0214	(0.0719)	0.0205	(0.0725)
r_{lt-3}	−0.0866	(0.0397)	−0.0992	(0.0405)	−0.0944	(0.0474)
r_{ut-3}	0.0626	(0.0416)	0.0583	(0.0437)	0.0599	(0.0453)
r_{lt-4}	−0.0377	(0.0400)	−0.0548	(0.0537)	−0.0483	(0.0526)
r_{ut-4}	−0.0393	(0.0400)	−0.0321	(0.0723)	−0.0349	(0.0690)
r_{lt-5}	−0.0823	(0.0407)	−0.0968	(0.0574)	−0.0913	(0.0534)
r_{ut-5}	0.0339	(0.0378)	0.0425	(0.0404)	0.0392	(0.0427)
$\widehat{\lambda}_{t-1}$			0.4450	(0.2363)	0.2759	(0.0870)
Degree of Freedom	993		992		993	
S.E. of Regression	0.9898		0.9888		0.9885	
Adjusted R^2	0.5294		0.5304		0.5175	
F-statistic	103.7		95.49		98.88	
σ_l^2	1.2909		1.2080		1.3642	
σ_u^2	0.9699		0.9960		0.9813	
ρ	0.6780		0.5752		0.6208	

ACCEPTED MANUSCRIPT

References

Arroyo, J., González-Rivera, G, and Maté, C. (2011), "Forecasting with Interval and Histogram Data. Some Financial Applications," in *Handbook of Empirical Economics and Finance, A. Ullah and D. Giles (eds.)*. Chapman and Hall, pp. 247-280.

Amemiya, T. (1973), "Regression Analysis when the Dependent Variable Is Truncated Normal," *Econometrica*. Vol. 41, pp. 997-1016.

Billard, L., and Diday, E. (2003), "From the statistics of data to the statistics of knowledge: symbolic data analysis," *Journal of the American Statistical Association*. Vol. 98, pp. 470-487.

Billard, L., and Diday E. (2006), *Symbolic Data Analysis: Conceptual Statistics and Data Mining*, 1st ed., Wiley and Sons, Chichester.

Brito, P. (2007), "Modelling and analysing interval data," *Proceedings of the 30th Annual Conference of GfKl*. Springer, Berlin, pp. 197-208.

Chernozhukov, V., Fernandez-Val, I., and Galichon, A. (2010), "Quantile and Probability Curves Without Crossing," *Econometrica*, Vol. 78 (3), pp. 1093-1125.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, Vol. 7 (1), pp. 1-26.

Goncalves, S., and White, H. (1979), "Bootstrap Standard Error Estimates for Linear Regression," *Journal of the American Statistical Association*, Vol. 100 (471), pp. 970978.

Greene, W. (1990), "Multiple roots of the Tobit log-likelihood," *Journal of Econometrics*. Vol. 46, pp. 365-380.

Heckman, J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*. Vol. 47, pp. 153-161.

Jennrich, R. (1969), "Asymptotic Properties of Non-Linear Least Squares Estimators," *The Annals of Mathematical Statistics*. Vol. 40, pp. 633-643.

Komunjer, I., and Owyang, M. (2011), "Multivariate Forecast Evaluation and Rationality Testing," *FRB of St. Louis Working Paper No. 2007-047D*. Available at SSRN.

ACCEPTED MANUSCRIPT

Lima Neto, E., and de Carvalho, F. (2008), "Centre and Range Method for Fitting a Linear Regression Model to Symbolic Interval Data," *Computational Statistics and Data Analysis*. Vol. 52, pp. 1500-1515.

Lima Neto, E., and de Carvalho, F. (2010), "Constrained linear regression models for symbolic interval-valued variables," *Computational Statistics and Data Analysis*. Vol. 54, pp. 333-347.

McLeish, L. (1974), "Dependent Central Limit Theorems and Invariance Principles," *Annals of Probability*. Vol. 2, pp. 620-628.

Mittelhammer, R., Judge, G., and Miller, D. (2000). *Econometric Foundations*. Cambridge University Press, Cambridge.

Newey, W., and MacFadden, D. (1994), "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics, Vol. IV, R. Engle and D. McFadden, (eds.)*. North Holland, Amsterdam, pp. 2111-2245.

Newey, W., and West, K. (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*. Vol. 55, pp. 703-708.

Newey, W., and West, K. (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*. Vol. 61, pp. 631-653.

Nath, G. (1972), "Moments of a Linearly Truncated Bivariate Normal Distribution," *Australian Journal of Statistics*. Vol. 14, pp. 97-102.

Orme, C. (1989), "On the Uniqueness of the Maximum Likelihood Estimator in Truncated Regression Models," *Econometric Reviews*. Vol. 8, pp. 217-222.

Orme, C., and Ruud, P. (2002), "On the Uniqueness of the Maximum Likelihood Estimator," *Economics Letters*. Vol. 75, pp. 209-217.

Patton, A., Politis D., and White, H. (2009), "Correction to 'Automatic Block-Length Selection for the Dependent Bootstrap' by D. Politis and H. White," *Econometric Reviews*. Vol. 28, pp. 372-375.

Politis, D., and White, H. (2004). "Automatic Block-Length Selection for the Dependent Bootstrap," *Econometric Reviews*. Vol. **23**, pp. 53-70.

ACCEPTED MANUSCRIPT

Tobin, J. (1958), “Estimation of Relationships for Limited Dependent Variables,” *Econometrica*. Vol. 26, pp. 24-36.

Rodrigues, P. M. M., and Salish, N. (2011), “Modeling and Forecasting Interval Time Series with Threshold Models: An Application to S&P500 Index Returns,” *Working Paper*. Economics and Research Department, Banco de Portugal .

White, H. (2000). *Asymptotic Theory for Econometricians: Revised Edition*. Academic Press, New York.

Wooldridge, M., and White, H. (1988), “Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes,” *Econometric Theory*. Vol. 4, pp. 210-230.

ACCEPTED MANUSCRIPT

Table 1: Simulation Results for Case B-2 and Case B-4

(a) Simulation Results for Case B-2

	Small Sample Size (T = 250)						Large Sample Size (T = 2000)					
	OLS		Two-step		Modified Two-step		OLS		Two-step		Modified Two-step	
Parameters	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse
$\beta_{lc} = 0$	−1.0645	1.0813	−0.1282	1.2306	−0.0936	0.3890	−1.0054	1.0075	−0.0162	0.3819	−0.0115	0.1312
$\beta_{uc} = 0$	−0.2738	0.3026	−0.0338	0.8377	−0.0476	0.1658	−0.2448	0.2487	−0.0050	0.2442	−0.0076	0.0544
$\beta_{11} = 0.8$	0.4648	0.3531	0.7542	0.3869	0.7643	0.1540	0.4852	0.3173	0.7948	0.1247	0.7963	0.0539
$\beta_{12} = 0.1$	0.4188	0.3430	0.1263	0.4018	0.1171	0.1744	0.4125	0.3157	0.1022	0.1281	0.1006	0.0619
$\beta_{21} = 0.1$	0.0253	0.1000	0.1002	0.2587	0.0951	0.0718	0.0247	0.0789	0.0999	0.0787	0.0990	0.0256
$\beta_{22} = 0.8$	0.8602	0.0980	0.7842	0.2655	0.7899	0.0832	0.8731	0.0780	0.7976	0.0800	0.7987	0.0292
$C_l = -1.4564$			−1.3899	1.8628	−1.4373	0.1344			−1.4505	0.5331	−1.4563	0.0478
$C_u = -0.3479$			−0.3609	1.3450	−0.3354	0.0778			−0.3523	0.3541	−0.3475	0.0280
$\sigma_l^2 = 3$	2.1280	0.8944	2.9323	0.3818	2.9509	0.4000	2.1494	0.8536	3.0014	0.1448	3.0008	0.1431
$\sigma_u^2 = 1$	0.9424	0.1009	0.9832	0.0944	0.9858	0.0940	0.9509	0.0576	0.9991	0.0343	0.9993	0.0343
$\rho = 0.8$	0.8268	0.0336	0.8003	0.0291	0.7956	0.0301	0.8270	0.0279	0.8002	0.0102	0.7997	0.0103

(b) Simulation Results for Case B-4

	Small Sample Size (T = 250)						Large Sample Size (T = 2000)					
	OLS		Two-step		Modified Two-step		OLS		Two-step		Modified Two-step	
Parameters	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse
$\beta_{lc} = 0$	−1.1702	1.1777	−85.30	5807	−0.0269	0.5277	−1.1628	1.1638	239.3	5441	0.0035	0.1729
$\beta_{uc} = 0$	−0.2769	0.2948	−5.8502	3420	−0.0073	0.1732	−0.2782	0.2805	102.5	6127	−0.0010	0.0605
$\beta_{11} = 0.1$	0.0498	0.1378	−0.0606	8.3845	0.0659	0.2514	0.0572	0.0609	0.2017	8.2371	0.0957	0.0863
$\beta_{12} = 0.05$	0.0931	0.1799	0.3443	12.6523	0.0723	0.3470	0.0919	0.0720	0.0461	9.9632	0.0538	0.1158
$\beta_{21} = 0.05$	0.0399	0.0949	−0.1618	6.4568	0.0438	0.1092	0.0391	0.0341	0.0276	7.0154	0.0483	0.0368
$\beta_{22} = 0.1$	0.1051	0.1260	0.2913	10.9189	0.0999	0.1480	0.1098	0.0446	0.1642	7.1917	0.1007	0.0499
$C_l = -1.4564$			124.2	7062	−1.4326	0.2111			−318.0	7340	−1.4552	0.0746
$C_u = -0.3479$			15.77	4069	−0.3380	0.1019			−132.8	8371	−0.3459	0.0388
$\sigma_l^2 = 3$	1.6612	1.3476	2.1295	0.8977	2.9675	0.6387	1.6664	1.3349	2.1328	0.8711	3.0002	0.2209
$\sigma_u^2 = 1$	0.9213	0.1150	0.9438	0.1060	0.9955	0.1090	0.9224	0.0830	0.9484	0.0605	0.9981	0.0393
$\rho = 0.8$	0.8607	0.0631	0.8274	0.0367	0.7991	0.0367	0.8595	0.0598	0.8270	0.0284	0.7990	0.0137

Number of Simulation=1000

Table 2: Simulation Results for Case NB-2 and Case NB-4

Simulation Results for Case NB-2												
	Small Sample Size (T = 250)						Large Sample Size (T = 2000)					
	OLS		Two-step		Modified Two-step		OLS		Two-step		Modified Two-step	
Parameters	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse
$\beta_{lc} = -2$	−2.0208	0.9077	−2.0139	1.0513	−2.0204	0.9082	−2.0202	0.3932	−2.0132	0.4039	−2.0202	0.3932
$\beta_{uc} = 2$	2.0813	0.5217	2.1187	0.6194	2.0814	0.5217	2.0137	0.2271	2.0166	0.2338	2.0137	0.2271
$\beta_{11} = 0.8$	0.7869	0.0632	0.7873	0.0707	0.7869	0.0632	0.7972	0.0258	0.7976	0.0264	0.7972	0.0258
$\beta_{12} = 0.1$	0.0910	0.0824	0.0904	0.0952	0.0910	0.0825	0.1003	0.0343	0.0997	0.0353	0.1003	0.0343
$\beta_{21} = 0.1$	0.0985	0.0351	0.1008	0.0401	0.0985	0.0351	0.1000	0.0149	0.1002	0.0153	0.1000	0.0149
$\beta_{22} = 0.8$	0.7869	0.0486	0.7836	0.0572	0.7869	0.0486	0.7980	0.0200	0.7978	0.0206	0.7980	0.0200
$C_l = -1.4564$			$3.94E4$	$1.12E6$	−1.4285	0.0918			−471.9	9246	−1.4530	0.0299
$C_u = -0.3479$			$-2.70E4$	$6.08E5$	−0.3312	0.0623			−268.3	5036	−0.3459	0.0210
$\sigma_l^2 = 3$	2.9536	0.2745	2.9301	0.2777	2.9167	0.2790	2.9969	0.0893	2.9939	0.0894	2.9923	0.0894
$\sigma_u^2 = 1$	0.9871	0.0911	0.9790	0.0920	0.9818	0.0914	1.0009	0.0312	0.9999	0.0312	1.0003	0.0312
$\rho = 0.8$	0.7987	0.0227	0.7987	0.0227	0.7948	0.0234	0.8000	0.0079	0.8000	0.0079	0.7995	0.0079
Simulation Results for Case NB-4												
	Small Sample Size (T = 250)						Large Sample Size (T = 2000)					
	OLS		Two-step		Modified Two-step		OLS		Two-step		Modified Two-step	
Parameters	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse	mean	rmse
$\beta_{lc} = -2$	−2.0219	0.5431	−1.6596	14.0955	−2.0210	0.5437	−1.9995	0.2019	−1.4023	21.2427	−1.9987	0.2021
$\beta_{uc} = 2$	2.0008	0.3217	2.0827	7.5863	2.0011	0.3217	2.0006	0.1191	2.4545	16.7747	2.0008	0.1191
$\beta_{11} = 0.1$	0.0938	0.0996	0.0919	0.5700	0.0939	0.0997	0.0989	0.0363	0.0998	0.3399	0.0990	0.0363
$\beta_{12} = 0.05$	0.0542	0.1686	0.0474	0.9405	0.0541	0.1688	0.0491	0.0622	0.0414	0.5580	0.0491	0.0623
$\beta_{21} = 0.05$	0.0488	0.0582	0.0483	0.3187	0.0488	0.0582	0.0495	0.0212	0.0476	0.2029	0.0495	0.0212
$\beta_{22} = 0.1$	0.0978	0.0997	0.1015	0.5140	0.0978	0.0998	0.0995	0.0367	0.0922	0.3195	0.0995	0.0367
$C_l = -1.4564$			−98.20	64189	−1.4348	0.0907			−2593	78843	−1.4554	0.0312
$C_u = -0.3479$			160.6	34474	−0.3325	0.0633			−1757	58342	−0.3470	0.0214
$\sigma_l^2 = 3$	2.9703	0.2742	2.9494	0.2771	2.9356	0.2779	2.9993	0.0950	2.9989	0.0953	2.9971	0.0954
$\sigma_u^2 = 1$	0.9886	0.0924	0.9807	0.0932	0.9834	0.0926	0.9992	0.0315	0.9984	0.0315	0.9987	0.0315
$\rho = 0.8$	0.7982	0.0236	0.7981	0.0238	0.7943	0.0245	0.8001	0.0079	0.8000	0.0079	0.7996	0.0079
Number of Simulation=1000												

Table 3: Methodology Evaluation for DGP1 (HIGH persistence and BINDING observability restriction)

DGP1									
Multivariate Normal Distribution									
	RMSE		*CR & ER*			*MLF*		*MDE*	
	Lower	*Upper*	*CR*	*ER*	$\frac{CR+ER}{2}$	$p = 1$	$p = 2$	$q = 1$	$q = 2$
CCRM	1.5851	1.2286	0.7099	0.6086	0.6593	2.2377	4.0244	1.2510	1.4182
CRM	1.5201	1.2973	0.7066	0.6073	0.6569	2.2445	3.9958	1.2505	1.4131
TS	**1.2735**	**0.7689**	0.8244	0.7023	**0.7633**	**1.6280**	**2.2136**	**0.8954**	**1.0519**
MTS	**1.2738**	**0.7691**	0.8244	0.7022	**0.7633**	**1.6284**	**2.2148**	**0.8956**	**1.0522**
GARCH-N (99%)	2.9030	2.6388	0.9877	0.3604	0.6740	4.9289	15.3984	2.6293	2.7741
GARCH-N (99.5%)	3.1914	2.9510	0.9928	0.3343	0.6635	5.5520	18.9028	2.9342	3.0736
GARCH-T (99%)	2.2336	1.8782	0.9543	0.4440	0.6992	3.4777	8.5229	1.9096	2.0636
GARCH-T (99.5%)	2.5063	2.1911	0.9734	0.4064	0.6899	4.0569	11.0902	2.1988	2.3540
Multivariate Student's t Distribution ($\nu = 5$)									
	RMSE		*CR & ER*			*MLF*		*MDE*	
	Lower	*Upper*	*CR*	*ER*	$\frac{CR+ER}{2}$	$p = 1$	$p = 2$	$q = 1$	$q = 2$
CCRM	2.1135	1.6705	0.7064	0.5856	0.6460	2.8669	7.2746	1.5954	1.9050
CRM	2.0382	1.7506	0.7055	0.5864	0.6459	2.8739	7.2356	1.5951	1.8999
TS	**1.6877**	**1.0099**	0.8323	0.6894	**0.7609**	**1.9925**	**3.8747**	**1.0921**	**1.3908**
MTS	**1.6890**	**1.0107**	0.8327	0.6894	**0.7610**	**1.9938**	**3.8815**	**1.0928**	**1.3919**
GARCH-N (99%)	3.8119	3.4153	0.9819	0.3373	0.6596	6.4432	26.2352	3.4253	3.6191
GARCH-N (99.5%)	4.1732	3.8073	0.9874	0.3131	0.6503	7.2230	31.9582	3.8066	3.9945
GARCH-T (99%)	3.1679	2.6299	0.9598	0.4013	0.6806	4.9167	17.0211	2.6739	2.9115
GARCH-T (99.5%)	3.6105	3.1287	0.9761	0.3609	0.6685	5.8674	22.9462	3.1430	3.3783

ACCEPTED MANUSCRIPT

Table 4: Methodology Evaluation for DGP3 (HIGH persistence and NON-BINDING observability restriction)

DGP3										
Multivariate Normal Distribution										
	RMSE		*CR & ER*			*MLF*		*MDE*		
	Lower	*Upper*	*CR*	*ER*	$\frac{CR+ER}{2}$	$p=1$	$p=2$	$q=1$	$q=2$
CCRM	2.4973	2.0577	0.8655	0.8346	0.8500	3.6409	10.4764	2.0241	2.2882
CRM	2.1549	2.2905	0.8650	0.8336	0.8493	3.5523	9.8958	1.9724	2.2238
TS	**1.7255**	**0.9964**	0.9181	0.8945	**0.9063**	**2.1732**	**3.9711**	**1.1772**	**1.4090**
MTS	**1.7259**	**0.9965**	0.9181	0.8945	**0.9063**	**2.1737**	**3.9727**	**1.1774**	**1.4092**
GARCH-N (99%)	8.3435	8.0010	0.9999	0.4079	0.7039	15.6599	133.6454	8.0080	8.1741
GARCH-N (99.5%)	9.4680	9.1662	1.0000	0.3744	0.6872	18.0331	173.6818	9.1715	9.3184
GARCH-T (99%)	3.1196	2.0920	0.7845	0.8634	0.8239	4.1878	14.1224	2.3392	2.6564
GARCH-T (99.5%)	3.0973	2.0581	0.8430	0.8310	0.8370	4.0564	13.8528	2.2646	2.6298
Multivariate Student's t Distribution ($\nu = 5$)									
	RMSE		*CR & ER*			*MLF*		*MDE*	
	Lower	*Upper*	*CR*	*ER*	$\frac{CR+ER}{2}$	$p=1$	$p=2$	$q=1$	$q=2$
CCRM	3.0472	2.5013	0.8463	0.8044	0.8253	4.3590	15.5570	2.4207	2.7878
CRM	2.6563	2.7776	0.8447	0.8028	0.8237	4.2534	14.7856	2.3596	2.7178
TS	**2.1303**	**1.2356**	0.9110	0.8781	**0.8945**	**2.5234**	**6.0691**	**1.3680**	**1.7415**
MTS	**2.1308**	**1.2359**	0.9110	0.8780	**0.8945**	**2.5240**	**6.0717**	**1.3683**	**1.7419**
GARCH-N (99%)	8.9350	8.4656	0.9993	0.3992	0.6992	16.4829	151.5336	8.4814	8.7035
GARCH-N (99.5%)	10.0825	9.6664	0.9997	0.3666	0.6831	18.9300	195.1327	9.6775	9.8766
GARCH-T (99%)	3.7183	2.3954	0.8416	0.7932	0.8174	4.7388	19.6039	2.6477	3.1280
GARCH-T (99.5%)	3.8918	2.6454	0.8914	0.7533	0.8223	4.9802	22.2088	2.7703	3.3278

Table 5: Simulation Results of DGP1 and GDP3 with Multivariate Normal Errors

		DGP1 (*high persistence and binding O.R.*)						DGP3 (*high persistence and non-binding O.R.*)					
		b_{11}	b_{12}	b_{0L}	b_{21}	b_{22}	b_{0U}	b_{11}	b_{12}	b_{0L}	b_{21}	b_{22}	b_{0U}
	true	-0.8	0.1	0	-0.1	0.8	0	-0.8	0.1	-2	-0.1	0.8	2
Mean	CCRM	-0.0986	-0.0986	-0.1230	-0.0986	-0.0986	2.7143	-0.2168	-0.2168	1.5698	-0.2168	-0.2168	12.3923
	CRM	-0.1553	-0.0419	-0.2841	-0.0419	-0.1553	2.8754	-0.3703	-0.0634	-0.0909	-0.0634	-0.3703	14.0530
	TS	**-0.7930**	**0.1081**	**-0.0348**	**-0.1017**	**0.7977**	**0.0057**	**-0.8002**	**0.1014**	**-2.0135**	**-0.0999**	**0.7990**	**2.0098**
	MTS	**-0.7970**	**0.1046**	**-0.0112**	**-0.1017**	**0.7975**	**0.0067**	**-0.7998**	**0.1018**	**-2.0184**	**-0.1001**	**0.7988**	**2.0123**
Bias^2	CCRM	0.4920	0.0394	0.0151	**0.0000**	0.8075	7.3676	0.3401	0.1004	12.7436	0.0137	1.0340	107.9996
	CRM	0.4156	0.0201	0.0807	0.0034	0.9126	8.2677	0.1847	0.0267	3.6448	0.0013	1.3696	145.2742
	TS	**0.0000**	**0.0001**	**0.0012**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**2e-04**	**0.0000**	**0.0000**	**0.0001**
	MTS	**0.0000**	**0.0000**	**0.0001**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**3e-04**	**0.0000**	**0.0000**	**0.0002**
MSE	CCRM	0.4922	0.0397	**0.0163**	**0.0002**	0.8077	7.3736	0.3404	0.1007	12.7738	0.014	1.0340	108.048
	CRM	0.4164	0.0202	0.0813	0.0034	0.9134	8.2792	0.1861	0.0267	3.6501	0.0014	1.3710	145.4484
	TS	**0.0060**	**0.0061**	0.2076	0.0019	**0.0021**	**0.0686**	**0.0002**	**0.0005**	**0.0538**	**0.0001**	**0.0002**	**0.0186**
	MTS	**0.0013**	**0.0021**	**0.0295**	**0.0004**	**0.0007**	**0.0091**	**0.0002**	**0.0005**	**0.0510**	**0.0001**	**0.0002**	**0.0174**

Table 6: Simulation Results of DGP1 and DGP3 with Multivariate Student-*t* Errors

		DGP1 (*high persistence and binding O.R.*)						DGP3 (*high persistence and non-binding O.R.*)					
		b_{11}	b_{12}	b_{0L}	b_{21}	b_{22}	b_{0U}	b_{11}	b_{12}	b_{0L}	b_{21}	b_{22}	b_{0U}
	true	-0.8	0.1	0	-0.1	0.8	0	-0.8	0.1	-2	-0.1	0.8	2
Mean	CCRM	-0.0976	-0.0976	**-0.1508**	**-0.0976**	-0.0976	3.2462	-0.2080	-0.2080	1.492	-0.2080	-0.2080	12.502
	CRM	-0.1461	-0.0490	-0.3158	-0.0490	-0.1461	3.4111	-0.3537	-0.0624	-0.1120	-0.0624	-0.3537	14.1055
	TS	**-0.8889**	**0.0016**	1.1406	**-0.0522**	**0.8529**	**-0.6148**	**-0.7947**	**0.1072**	**-2.1132**	**-0.1030**	**0.7957**	**2.0657**
	MTS	**-0.9284**	-0.0329	1.5307	-0.0295	**0.8729**	-0.8406	**-0.7963**	**0.1059**	**-2.0963**	**-0.1021**	**0.7965**	**2.0559**
Bias^2	CCRM	0.4934	0.0390	**0.0228**	**0.0000**	0.8056	10.5376	0.3504	0.0949	12.192	0.0117	1.0161	110.29
	CRM	0.4276	0.0222	0.0997	0.0026	0.8951	11.6359	0.1992	0.0264	3.5645	0.0014	1.3310	146.54
	TS	**0.0079**	**0.0097**	1.3010	0.0023	**0.0028**	**0.3780**	**0.0000**	**0.0001**	**0.0128**	**0.0000**	**0.0000**	**0.0043**
	MTS	0.0165	0.0177	2.3431	0.0050	0.0053	0.7067	**0.0000**	**0.0000**	**0.0093**	**0.0000**	**0.0000**	**0.0031**
MSE	CCRM	0.4937	0.0393	**0.0244**	**0.0002**	0.8058	10.5479	0.3508	0.0952	12.225	0.0120	1.0160	110.35
	CRM	0.4285	0.0223	0.1008	0.0027	0.8960	11.6540	0.2007	0.0264	3.5710	0.0015	1.3325	146.74
	TS	**0.0134**	**0.0156**	1.7679	0.0042	**0.0047**	**0.5458**	**0.0003**	**0.0006**	**0.0827**	**0.0001**	**0.0002**	**0.0277**
	MTS	0.0200	0.0220	3.5410	0.0061	0.0066	1.1562	**0.0002**	**0.0006**	**0.0710**	**0.0001**	**0.0002**	**0.0234**

ACCEPTED MANUSCRIPT

Table 7: Methodology Evaluation for SP500 Daily Low/High Interval Returns

	RMSE		*CR & ER*			*MLF*		*MDE*	
Unstable Period (2007/1/1-2011/4/29); BINDING observability restriction									
	Lower	*Upper*	*CR*	*ER*	$\frac{CR+ER}{2}$	$p = 1$	$p = 2$	$q = 1$	$q = 2$
CCRM	1.1541	0.9990	0.6811	0.5939	0.6375	1.4506	2.3300	0.7827	1.0794
CRM	1.1541	0.9990	0.6811	0.5939	0.6375	1.4506	2.3300	0.7827	1.0794
TS	**1.1356**	**0.9828**	0.6810	0.5974	0.6392	**1.4337**	**2.2555**	**0.7711**	**1.0619**
MTS	**1.1379**	**0.9831**	0.6824	0.5952	0.6388	**1.4360**	**2.2611**	**0.7726**	**1.0633**
GARCH-N (99%)	1.8625	1.8329	0.9557	0.3630	**0.6594**	2.9936	6.8285	1.6239	1.8478
GARCH-N (99.5%)	2.0469	2.0246	0.9671	0.3399	**0.6535**	3.3268	8.2886	1.7960	2.0358
GARCH-T (99%)	2.2210	2.2064	0.9735	0.3272	0.6504	3.5868	9.8010	1.9308	2.2137
GARCH-T (99.5%)	2.6042	2.5989	0.9857	0.2921	0.6389	4.2692	13.5366	2.2743	2.6016
Stable Period (2004/1/1-2007/1/1); NON-BINDING observability restriction									
	RMSE		*CR & ER*			*MLF*		*MDE*	
	Lower	*Upper*	*CR*	*ER*	$\frac{CR+ER}{2}$	$p = 1$	$p = 2$	$q = 1$	$q = 2$
CCRM	0.4146	0.3958	0.7177	0.6427	0.6802	0.6396	0.3285	0.3471	0.4053
CRM	0.4146	0.3958	0.7177	0.6427	0.6802	0.6396	0.3285	0.3471	0.4053
TS	**0.4123**	**0.3914**	0.7184	0.6466	0.6825	**0.6337**	**0.3232**	**0.3435**	**0.4020**
MTS	**0.4129**	**0.3942**	0.7173	0.6443	0.6808	**0.6369**	**0.3258**	**0.3452**	**0.4036**
GARCH-N (99%)	0.6368	0.6240	0.9473	0.4410	**0.6942**	1.1157	0.7949	0.6142	0.6304
GARCH-N (99.5%)	0.7025	0.6908	0.9621	0.4143	**0.6882**	1.2435	0.9708	0.6818	0.6967
GARCH-T (99%)	0.7415	0.7302	0.9686	0.4002	0.6844	1.3194	1.0830	0.7211	0.7359
GARCH-T (99.5%)	0.8599	0.8499	0.9818	0.3621	0.6720	1.5562	1.4617	0.8405	0.8549

ACCEPTED MANUSCRIPT

List of Figures

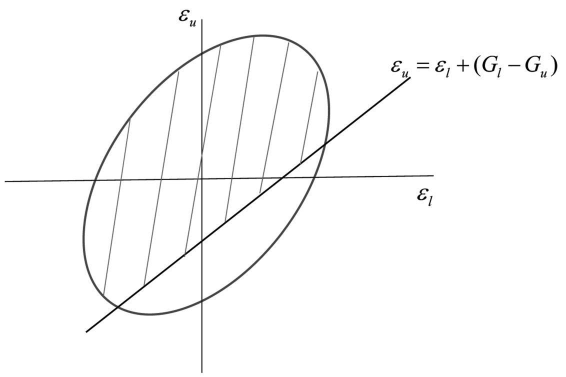


Figure 1: Truncated Distribution of the Error Term

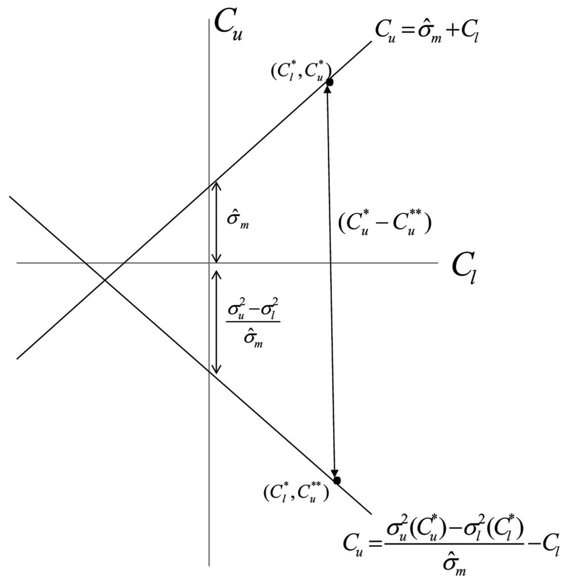
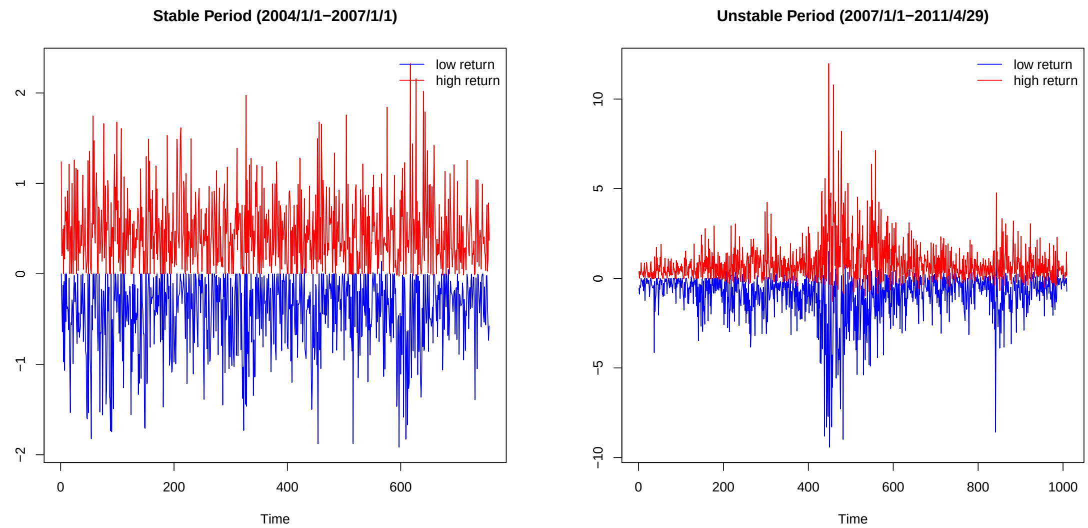


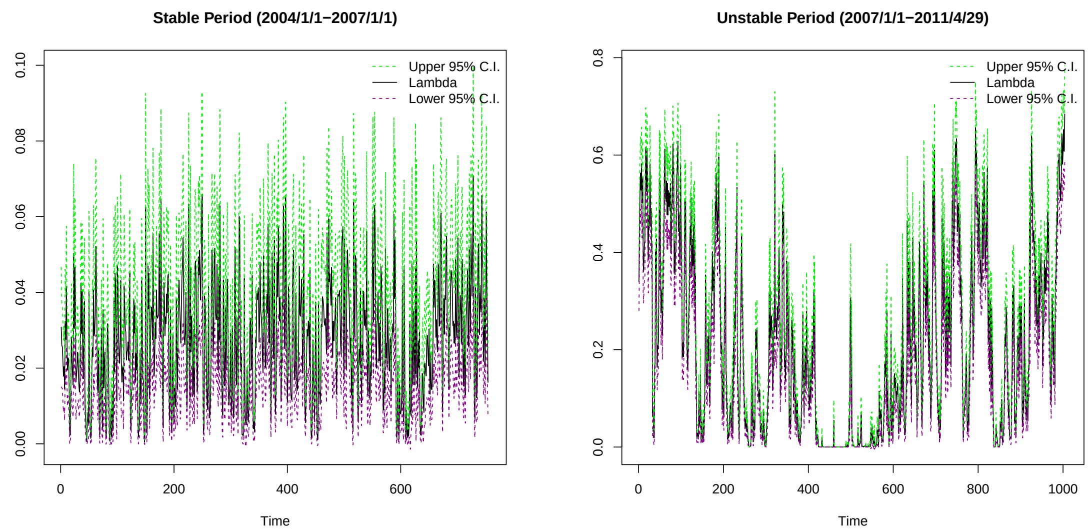
Figure 2: Minimum Distance Estimator

ACCEPTED MANUSCRIPT



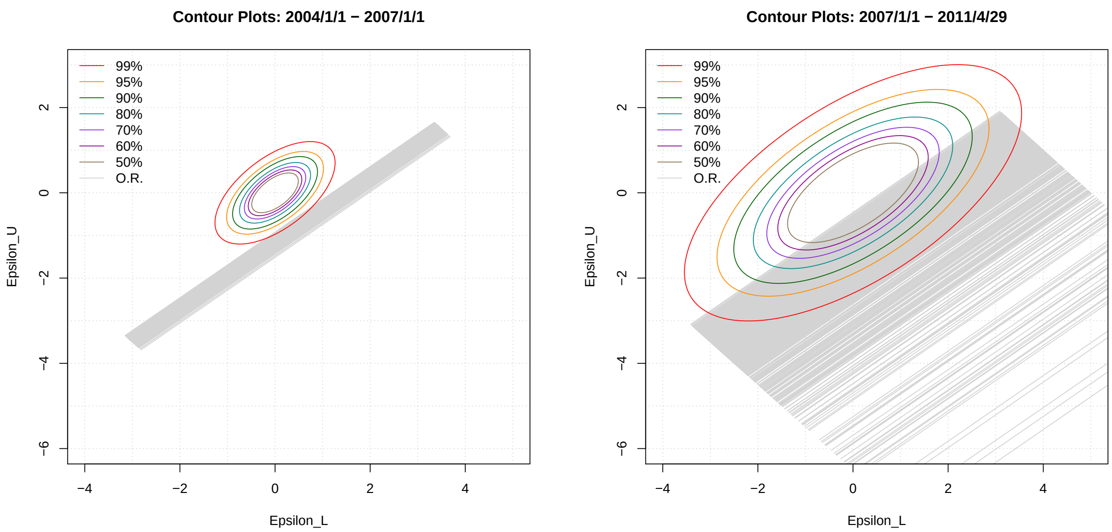
(a) Stable Period (2004/1/1-2007/1/1) (b) Unstable Period (2007/1/1-2011/4/29)

Figure 3: High/Low Returns of Daily SP500 Index for Stable and Unstable Periods



(a) Stable Period (2004/1/1-2007/1/1) (b) Unstable Period (2007/1/1-2011/4/29)

Figure 4: Estimated Inverse Mill's Ratio for Stable and Unstable Periods



(a) Stable Period (2004/1/1-2007/1/1) (b) Unstable Period (2007/1/1-2011/4/29)

Figure 5: Observability Restriction for Stable and Unstable Periods

ACCEPTED MANUSCRIPT