When A Factor Is Measured with Error: The Role of Conditional Heteroskedasticity in Identifying and Estimating Linear Factor Models

Prono, Todd

Commodity Futures Trading Commission

19 September 2011

When A Factor Is Measured with Error: The Role of Conditional Heteroskedasticity in Identifying and Estimating Linear Factor Models[1]

Todd Prono[2]

Commodity Futures Trading Commission

September 2011

Abstract

A new method is proposed for estimating linear triangular models, where identification results from the structural errors following a bivariate and diagonal GARCH(1,1) process. The associated estimator is a GMM estimator shown to have the usual $\sqrt{T}$-asymptotics. A Monte Carlo study of the estimator is provided as is an empirical application of estimating market betas from the CAPM. These market beta estimates are found to be statistically distinct from their OLS counterparts and to display expanded cross-sectional variation, the latter feature offering promise for their ability to provide improved pricing of cross-sectional expected returns.

JEL Codes: C3, C13, C32. Keywords: Measurement error, triangular models, factor models, heteroskedasticity, identification, many moments, GMM.

---

[1] I owe gratitude to Robin Lumsdaine, Arthur Lewbel, David Reiffen, and seminar participants at Binghamton University for helpful comments and discussions.

[2] Corresponding Author: Todd Prono, Commodity Futures Trading Commission, Office of the Chief Economist, 1155 21$^{\text{st}}$, N.W., Washington, DC 20581. (202) 418-5460, tprono@cftc.gov.

1. Introduction

This paper presents a new method for estimating linear triangular models where measurement error or endogeneity effects one of the regressors. Examples of these types of models include (i) asset return factor models where one of the factors is either measured inaccurately or an imperfect proxy for the true, latent, factor or (ii) restricted VAR models from the empirical macro literature. The traditional approach to identifying these models is through the use of exclusionary restrictions on parameters affecting the conditional mean or, equivalently, through the assumed existence of valid instruments. In contrast, this paper demonstrates how a certain parametric specification of the conditional heteroskedasticty (CH) affecting the structural errors to the triangular system allows for identification in the absence of traditional instruments. As such, this paper contributes to the literature on identification through various forms of heteroskedasticity. Based on this identification result, a continuous updating estimator (CUE) is proposed that is shown to be consistent and asymptotically normal. It is also robust to many moments bias. This estimator performs well in Monte Carlo experiments under moment existence criteria that allow for varying fat-tailed processes. The estimator is also applied to estimating market betas from the familiar CAPM, offering promising results for the ability of these estimates to price expected returns in the cross-section.

Consider the model

$$Y_{1,t} = X_t^{'}\beta_1 + Y_{2,t}\beta_2 + \epsilon_{1,t},$$

$$Y_{2,t} = X_t^{'}\delta + \epsilon_{2,t},$$

where $Y_t = \begin{bmatrix} Y_{1,t} & Y_{2,t} \end{bmatrix}^{'}$ is a vector of endogenous variables, X_t a vector of observable covariates than can include lags of Y_t, and $\epsilon_t = \begin{bmatrix} \epsilon_{1,t}, & \epsilon_{2,t} \end{bmatrix}^{'}$ a vector of structural errors. Of course, $E\left[X_t\epsilon_{i,t}\right] = 0$ for $i = 1, 2$ is insufficient for identifying the model. Rather than impose zero restrictions on certain elements in β_1, consider the following specification for ϵ_t:

$$E\left[\epsilon_t \mid F_{t-1}\right] = 0, \quad E\left[\epsilon_t\epsilon_t^{'} \mid F_{t-1}\right] = \left[h_{ij,t}\right], \tag{1}$$

$$h_{ij,t} = \omega_{ij} + a_{ij}\epsilon_{i,t-1}\epsilon_{j,t-1} + b_{ij}h_{ij,t-1}, \quad i, j = 1, 2, \tag{2}$$

where $\mathcal{F}_t = \sigma\left(Z_t, Z_{t-1}, \ldots\right)$ and $Z_t = \left(Y_t', X_t'\right)'$. This specification describes a bivariate, diagonal GARCH(1,1) model. The univariate version was introduced by Bollerslev (1986), the multivariate generalization by Bollerslev, Engle, and Wooldridge (1988). This model of CH can be shown to support identification of the triangular system in the same way as traditional zero restrictions imposed on β_1; namely, through an examination of the reduced form (see Prono 2010). Allowing for this result are the structural restrictions imposed by the parameterization in (2). This parameterization imposes a structure on $Cov\left[\epsilon_{i,t}\epsilon_{j,t}, \epsilon_{i,t-k}\epsilon_{j,t-k}\right]$ for $k \geq 2$, and functions of these covariances can be paired with the moment conditions $E\left[X_t\epsilon_{i,t}\right] = 0$ to grant identification. A benefit of this result is that identification of the triangular system is achieved without the need for considering all of the parameters in (2); rather, only a subset of these parameters needs to be considered. Before proceeding to the formal statement of identification and the properties of the associated estimator, it is instructive to further consider the source of identification in (1) and (2) as well as a factor model that would benefit from this result.

1.1 Identification Source

The identification problem confronting the triangular system can be recast in terms of a control function as in Klein and Vella (2010). Doing so provides a heuristic basis for understanding how (1) and (2) solve this problem. Consider the conditional regression

$$A\left(\mathcal{F}_{t-1}\right) \equiv \arg\min_A E\left[\epsilon_{1,t} - A\epsilon_{2,t} \mid \mathcal{F}_{t-1}\right]^2 = Cov\left[\epsilon_{1,t},\ \epsilon_{2,t} \mid \mathcal{F}_{t-1}\right]/Var\left[\epsilon_{2,t} \mid \mathcal{F}_{t-1}\right].$$

In this case, $U_t \equiv \epsilon_{1,t} - A\left(\mathcal{F}_{t-1}\right)\epsilon_{2,t}$ is uncorrelated with $\epsilon_{2,t}$ conditional on $\mathcal{F}_{t-1}$ and forms the basis for the controlled regression

$$Y_{1,t} = X_t'\beta_1 + Y_{2,t}\beta_2 + A\left(\mathcal{F}_{t-1}\right)\epsilon_{2,t} + U_t. \tag{3}$$

Let $V_t = \left[X_t', Y_{2,t}, \epsilon_{2,t}\right]$. Then, if ϵ_t is homoskedastic so that $A\left(\mathcal{F}_{t-1}\right)$ is constant, we have the usual identification problem, since (absent exclusionary restrictions for β_1) $E\left[V_t'V_t\right]$ is singular.[3] Now suppose, instead, that ϵ_t is CH, and let $W_t = \left[X_t', Y_{2,t}, A\left(\mathcal{F}_{t-1}\right)\epsilon_{2,t}\right]$. Then, $E\left[W_t'W_t\right]$ is nonsingular, and the identification problem is solved, provided that $A\left(\mathcal{F}_{t-1}\right)$ can be consistently estimated. This latter requirement necessitates (2) and illustrates why CH alone is not sufficient for identifying the triangular system.

One approach to make estimation of $A\left(\mathcal{F}_{t-1}\right)$ feasible is to assume a constant conditional covariance. Specifically, since $A\left(\mathcal{F}_{t-1}\right) = h_{12,t}/h_{22,t}$, if $h_{12,t} = \omega_{12}$, then $A\left(\mathcal{F}_{t-1}\right)$ can be consistently estimated because $h_{22,t}$ is parameterized as a univariate GARCH(1,1) model, and $\epsilon_{2,t}$ is identified provided that $E\left[X_tX_t'\right]$ is nonsingular. Sentana and Fiorentini (2001) employ this precise covariance restriction to identify a latent factor model, where univariate GARCH(1,1) processes characterize the conditional variances of the factors. Lewbel (2010) also relies upon a constant conditional covariance restriction for identifying triangular and simultaneous models. In a similar vein, Vella and Verbeek (1997) and Rummery et al. (1999), too, rely on a covariance restriction for identification by proposing rank order as an instrumental variable.

A contribution of this paper is to allow $h_{12,t}$ to be time-varying, parameterizing it as an ARMA(1,1) process, analogous to the specification of each conditional variance. Doing so complicates estimation of $A\left(\mathcal{F}_{t-1}\right)$ by requiring the control function to be treated simultaneously along with (3), since $h_{12,t}$ now depends on past values of $\epsilon_{1,t}$. The functional form in (2) allows for this simultaneous estimation. Rather than propose an estimator for the controlled regression, however, this paper demonstrates how the moment conditions

$$E\left[X_t\epsilon_{i,t}\right] = 0, \quad Cov\left[\epsilon_{i,t}\epsilon_{j,t}, \epsilon_{i,t-k}\epsilon_{j,t-k}\right] - \phi_{ij}Cov\left[\epsilon_{i,t}\epsilon_{j,t}, \epsilon_{i,t-(k-1)}\epsilon_{j,t-(k-1)}\right] = 0, \tag{4}$$

$\forall\ i, j = 1, 2$ excluding $i = j = 1$ where $\phi_{ij} = a_{ij} + b_{ij}$ identify the triangular system and how finite sample analogs to these moment conditions combine to form an estimator for that system. Since the parametric form of (2) implies this second set of moment conditions, the source of identification behind the controlled regression in (3) and the moment conditions in (4) is equivalent.

Klein and Vella (2010) is a work closely related to this one. They show identification

[3] Singularity follows from $\epsilon_{2,t}$ being a linear combination of $Y_{2,t}$ and X_t.

of the triangular model given heteroskedastic errors of a semi-parametric functional form. Their estimator is more complicated to implement than this one, owing to the generality of the heteroskedastic specification. In many applications of financial economics, however, the more restrictive CH specification of (6) and (7) proves warranted (see, for example, Hansen and Lunde 2005). Moreover, the Klein and Vella approach links the conditional covariance between errors directly to each conditional variance. In this paper, by contrast, $h_{12,t}$ is not a direct function of either $h_{11,t}$ or $h_{22,t}$.[4]

Other papers that exploit heteroskedasticity for identification include Rigobon (2003) and Rigobon and Sack (2003), where multiple unconditional variance regimes act as probabilistic instruments, and the correlation between structural errors is sourced to common, unobserved, shocks.

1.2 Measurement Error

Consider the CAPM of Sharpe (1964) and Lintner (1965), where $Y_{1,t}$ is a given excess security return, $Y_{2,t}^{*}$ is the excess return on the true market return, which is unobservable, and $Y_{2,t}$ is an observable proxy to the true excess market return. If the CAPM prices all security returns including the proxy return, then

$$Y_{1,t} = \beta_1 + Y_{2,t}^{*}\beta_2 + U_{1,t}, \tag{5}$$

$$Y_{2,t} = \gamma_1 + Y_{2,t}^{*}\gamma_2 + U_{2,t},$$

where $\beta_1 = \gamma_1 = 0$. If $E\left[Y_{2,t}\right] = E\left[Y_{2,t}^{*}\right]$, then

$$Y_{2,t} = Y_{2,t}^{*} + U_{2,t}, \tag{6}$$

which casts the relationship between the proxy return and the true market return as one of measurement error, although not, necessarily, in the classical sense, since the theory does not restrict $U_{1,t}$ to be orthogonal to $U_{2,t}$ or, equivalently, $U_{1,t}$ to be orthogonal to $Y_{1,t}$. Consider

[4] An example where $h_{12,t}$ is a direct function of $h_{11,t}$ and $h_{22,t}$ is the CCC model of Bollerslev (1990).

the projection equation $Y^*_{2,t} = \delta + V_{2,t}$. Substituting (6) into this projection equation and into (5) produces the triangular system

$$Y_{1,t} = \beta_1 + Y_{2,t}\beta_2 + \epsilon_{1,t}, \quad \epsilon_{1,t} = U_{1,t} - \beta_2 U_{2,t}, \tag{7}$$

$$Y_{2,t} = \delta + \epsilon_{2,t}, \quad \epsilon_{2,t} = U_{2,t} + V_{2,t}.$$

Given a means for consistently estimating (7), one has a response to the Roll critique (1977), since, in this case, $\widehat{\beta}_2$ is a measure of the security return's sensitivity to the true market return. Notice, however, that OLIVE from Meng, Hu, and Bai (2011) cannot, necessarily, provide this consistent estimate, since any additional excess security return, say, $Y_{3,t}$, which is priced according to $Y_{3,t} = \beta_3 + Y^*_{2,t}\beta_4 + U_{3,t}$, is only a valid instrument for $Y_{2,t}$ if all $U_{i,t}$ are orthogonal.

This paper explores estimation of the triangular system using (1) and (2), a system commonly employed on security returns. The associated estimator represents a response to the Roll critique insofar as one is willing to assume certain higher moment existence criteria for those security returns. A multi-factor generalization of the above example follows readily if the non-market factors are not also measured with error.

2. Identification

For the linear triangular model

$$Y_{1,t} = X_t'\beta_{1,0} + Y_{2,t}\beta_{2,0} + \epsilon_{1,t} \tag{8}$$

$$Y_{2,t} = X_t'\delta_0 + \epsilon_{2,t} \tag{9}$$

together with the following bivariate GARCH(1,1) specification for its structural errors $\epsilon_t = \left[\epsilon_{1,t}, \quad \epsilon_{2,t}\right]'$

$$E\left[\epsilon_t \mid F_{t-1}\right] = 0, \quad E\left[\epsilon_t\epsilon_t' \mid F_{t-1}\right] = \left[h_{ij,t}\right], \tag{10}$$

$$h_{ij,t} = \omega_{ij,0} + a_{ij,0}\epsilon_{i,t-1}\epsilon_{j,t-1} + b_{ij,0}h_{ij,t-1}, \quad i,j = 1,2, \tag{11}$$

$\beta_{1,0}$ is the true value of β_1 and similarly for the other parameters. Even if there are no zero restrictions for $\beta_{1,0}$, which is equivalent to saying that there are no instruments available for $Y_{2,t}$, this section shows that (8) may still be identified given the parametric form of CH in (11).

ASSUMPTION A1: *(i)* $E[X_t X_t']$ *and* $E[X_t Y_t]$ *are finite and identified from the data. (ii)* $E[X_t X_t']$ *is nonsingular. (iii)* $E\left[X_t \epsilon_{i,t}\right] = 0$.

Given A1, the secondary equation (9) is identified, as is the reduced form of the primary equation (8). Let the reduced form errors from (8) be $R_{1,t} = Y_{1,t} - X_t' E[X_t X_t']^{-1} E\left[X_t Y_{1,t}\right]$. The relationship between these reduced form errors and the structural errors is

$$R_{1,t} = \epsilon_{1,t} + \epsilon_{2,t}\beta_{2,0}. \tag{12}$$

ASSUMPTION A2: *(i) Let* $H_t = \left[h_{ij,t}\right]$. H_t *is positive definite almost surely. (ii)* $a_{ij,0} > 0$, $b_{ij,0} \geq 0$. *(iii) Let* $\phi_{ij,0} = a_{ij,0} + b_{ij,0}$. $\phi_{12,0} \neq \phi_{22,0}$.

In practice, positive definiteness under A2 can be satisfied using the BEKK parameterization of (11) proposed by Engle and Kroner (1995).[5] Allowing $b_{ij,0} = 0$ permits H_t to follow a diagonal ARCH(1) process. Let $Z_{ij,t} = \epsilon_{i,t}\epsilon_{j,t}$. Then

$$\begin{aligned} Z_{ij,t} &= h_{ij,t} + W_{ij,t} \\ &= \omega_{ij,0} + \phi_{ij,0} Z_{ij,t-1} - b_{ij,0} W_{ij,t-1} + W_{ij,t}, \end{aligned}$$

where, $E\left[W_t \mid \digamma_{t-1}\right] = 0$ and $E\left[W_t W_{t-s}\right] = 0 \ \forall \ s \geq 1$.

ASSUMPTION A3: $Z_{ij,t}$ *is covariance stationary* $\forall \ i, j = 1, 2$ *except* $i = j = 1$.

An implication of A3 is that $\phi_{ij,0} < 1$, in which case $E\left[Z_{ij,t}\right] = \sigma_{ij,0} = \frac{\omega_{ij,0}}{1-\phi_{ij,0}}$ (see Theorem 1 in Bollerslev 1986). Note that while $\epsilon_{2,t}$ is required to be fourth moment stationary, no such restriction is imposed on $\epsilon_{1,t}$. In fact, $\epsilon_{1,t}$ need not even have a finite variance (i.e.,

[5] See Proposition 2.6 of the aforementioned work.

$\phi_{11,0} = 1$ as in the IGARCH case is not ruled out). Given A3, if $z_{ij,0t} = Z_{ij,t} - \sigma_{ij,0}$, then

$$z_{ij,0t} = \phi_{ij,0} z_{ij,0t-1} - b_{ij,0} W_{ij,t-1} + W_{ij,t}. \tag{13}$$

Consider $z_{lm,0t}$, where $l, m = 1, 2$ excluding the case where $l = m = 1$. Multiplying both sides of (13) by $z_{lm,0t-k}$ for $k \geq 2$ and taking expectations produces

$$E\left[z_{ij,0t} z_{lm,0t-k}\right] = \phi_{ij,0} E\left[z_{ij,0t} z_{lm,0t-(k-1)}\right]. \tag{14}$$

This expression was derived in Bollerslev (1986, 1988) and He and Teräsvirta (1999) for the case where $i = j = l = m$.

ASSUMPTION A4: *Let* $U_{ij,t-2} = \left[Z_{ij,t-2}, \ldots, Z_{ij,t-K}\right]'$. *Given (12), the reduced from of* $U_{12,t-2}$ *is* $U^{(R)}_{12,t-2}$. *The matrix* $\Phi_R = \begin{bmatrix} Cov\left(R_{1,t}\epsilon_{2,t}, U^{(R)}_{12,t-1}\right) & Cov\left(\epsilon^2_{2,t}, U^{(R)}_{12,t-1}\right) \\ Cov\left(R_{1,t}\epsilon_{2,t}, U_{22,t-1}\right) & Cov\left(\epsilon^2_{2,t}, U_{22,t-1}\right) \end{bmatrix}$ *has full column rank.*

Since $R_{1,t}$ can be estimated by regressing $Y_{1,t}$ on X_t, and $\epsilon_{2,t}$ can be estimated by regressing $Y_{2,t}$ on X_t, the matrix rank test of Cragg and Donald (1996) can be applied to an estimate of Φ_R, rendering A4 testable. Alternatively, one can simply test if the determinant of $\Phi'_R \Phi_R$ is zero, since A4 requires this matrix to be nonsingular.

THEOREM 1. *Consider the model of (8)–(11). Let Assumptions A1–A4 hold. Then* $\beta_{1,0}$, $\beta_{2,0}$, $\phi_{12,0}$, *and* $\phi_{22,0}$ *are identified.*

Proofs are in the Appendix. The proof of Theorem 1 is based on the reduced form of (14). As a consequence, only the conditional covariance function and the conditional variance function for $\epsilon_{2,t}$ matter for identification (see section 1.1). Structural parameters can be retrieved from this reduced form because of the parametric specification in (11). This specification omits lags of $\epsilon^2_{1,t}$ and $\epsilon^2_{2,t}$ from the conditional covariance function and lags of $\epsilon^2_{1,t}$ and $\epsilon_{1,t}\epsilon_{2,t}$ from the conditional variance function of $\epsilon_{2,t}$. These restrictions are analogous to traditional zero restrictions that produce valid instruments. The diagonal GARCH(1,1) specification, therefore, is the key identifying assumption.

3. Estimation

Let $S_t = \{Y_t, X_t\}_{t=1}^{T}$. Define $\epsilon_{1,t} = Y_{1,t} - X_t'\beta_1 - Y_{2,t}\beta_2$, $\epsilon_{2,t} = Y_{2,t} - X_t'\delta$, $z_{ij,t} = \epsilon_{i,t}\epsilon_{j,t} - \sigma_{ij}$, and $u_{ij,t-2} = \begin{bmatrix} z_{ij,t-2}, & \ldots, & z_{ij,t-K} \end{bmatrix}'$ $\forall\ i,j = 1,2$ except $i = j = 1$, and $K \geq 2$.

ASSUMPTION A5: $\theta = \left\{\beta_1, \beta_2, \delta, \omega_{ij}, \phi_{ij}\right\}$. $\theta_0 \in \Theta \subseteq \mathbb{R}^7$, *located in the interior of* Θ, *a compact parameter space defined such that* ϕ_{12}/ϕ_{22} *excludes an open neighborhood of one.*

A5 is a standard regulatory condition. Its only nuance stems from the need to reconcile compactness with A2(iii). Consider the vector valued functions

$$g_1(S_t;\theta) = X_t\epsilon_{1,t}, \quad g_2(S_t;\theta) = X_t\epsilon_{2,t},$$

$$g_3(S_t;\theta) = z_{ij,t}$$

$$g_4(S_t;\theta) = z_{ij,t}\left(u_{lm,t-2} - \phi_{ij}u_{lm,t-1}\right), \quad \forall\ l,m = 1,2 \text{ excluding } l = m = 1,$$

where $g_3(S_t;\theta)$ and $g_4(S_t;\theta)$ stack the vector valued functions $\forall\ i,j,l,m$ into single column vectors. Let $g_t(\theta) = \begin{bmatrix} g_1(S_t;\theta), & g_2(S_t;\theta), & g_3(S_t;\theta)', & g_4(S_t;\theta)' \end{bmatrix}'$. In addition, let

$$\begin{aligned}
\widehat{g}(\theta) &= T^{-1}\sum_{t=K+1}^{T} g_t(\theta), \qquad \overline{g}(\theta) = E\left[g_t(\theta)\right], \\
\widehat{G}(\theta) &= \frac{\partial\widehat{g}(\theta)}{\partial\theta}, \qquad G(\theta) = E\left[\frac{\partial g_t(\theta)}{\partial\theta}\right], \\
g_{\theta t}(\theta) &= \frac{\partial g_t(\theta)}{\partial\theta}, \quad \widehat{\Sigma}(\theta) = T^{-1}\sum_{t=K+1}^{T} g_t(\theta)\, g_{\theta t}(\theta)', \quad \Sigma(\theta) = E\left[g_t(\theta)\, g_{\theta t}(\theta)'\right], \\
\Omega(\theta) &= \sum_{s=-(L-1)}^{s=(L-1)} E\left[g_{t-s}(\theta)\, g_t(\theta)'\right], \quad L \geq 1, \\
\widehat{\Omega}(\theta) &= \sum_{s=-(L-1)}^{s=(L-1)} T^{-1}\sum_{t=K+s+1}^{T} g_{t-s}(\theta)\, g_t(\theta)',
\end{aligned}$$

and consider the estimator

$$\widehat{\theta} = \arg\min_{\theta\in\Theta}\ \widehat{g}(\theta)'\Lambda_T\widehat{g}(\theta) \tag{15}$$

for some sequence of positive semi-definite Λ_T. For this estimator, the moment conditions defined from $g_1(S_t;\theta)$ and $g_2(S_t;\theta)$ are standard for linear models and are, of course, insufficient for identifying the associated triangular system. The moment conditions defined from $g_3(S_t;\theta)$ and $g_4(S_t;\theta)$ produce the finite sample version of (14) and are, therefore, instrumental in enabling identification given Theorem 1.

ASSUMPTION A6: *(i) $\{X_t\epsilon_{i,t}\}$ is an L^1 mixingale (see Andrews 1988 for a definition). (ii) $\exists$ an $r>1$ such that $E\left[\left|X_t\epsilon_{i,t}\right|^r\right]<M$. (iii) Let $v_{t,k}=z_{ij,0t}z_{lm,0t-k}-E\left[z_{ij,0t}z_{lm,0t-k}\right]$ for $k=1,\ldots,K$. $\{v_{t,k}\}$ is uniformly integrable.*

Mixingale properties for $\{g_t(\theta)\}$ factor prominently into establishing consistency of $\widehat{\theta}$ in Theorem 2 below. Also, notice that A6(ii) continues to allow $\epsilon_{1,t}$ to follow an IGARCH process.

THEOREM 2 (Consistency). *Consider the estimator in (15) for the model of (8)–(11). Assume that $\Lambda_T \xrightarrow{p} \Lambda_0$, a positive definite matrix. Let Assumptions A1–A6 hold. Then, $\widehat{\theta} \xrightarrow{p} \theta_0$.*

Consistency under Theorem 2 requires fourth moment existence for $\epsilon_{2,t}$ but no corresponding requirement for $\epsilon_{1,t}$. In fact, $\epsilon_{1,t}$ does not even need to be covariance stationary. Depending upon the specification of K, however, the estimator in (15) can involve many moment conditions. Works by Stock and Wright (2000), Newey and Smith (2004), Han and Phillips (2006), and Newey and Windmeijer (2009), for instance, highlight the bias caused by many moment conditions in GMM estimators. Newey and Windmeijer (2009) illustrate how the CUE of Hansen, Heaton, and Yaron (1996) is robust to this bias. Theorem 2 nests the CUE, provided that sufficient moment existence criteria are satisfied so that $\widehat{\Omega}\left(\widehat{\theta}\right)^{-1} \xrightarrow{p} \Omega(\theta_0)^{-1}$.[6] The bias-reducing feature of the CUE relative to GMM estimators motivates the following discussion of asymptotic normality to only consider the case where $\Lambda_T=\widehat{\Omega}(\theta)^{-1}$.

ASSUMPTION A7: *(i) $\{X_t\epsilon_{i,t}\}$ is an L^2 mixingale. (ii) $\exists$ a neighborhood N of θ_0 such that $E\left[\sup_{\theta\in N}\|g_t(\theta)\|^2\right]<\infty$. (iii) $\{g_{t-s}(\lambda_0,\sigma_0^2)\,g_t(\lambda_0,\sigma_0^2)'\}$ satisfies the UWLLN*

[6] These criteria include $\epsilon_{1,t}$ being, at least, covariance stationary, and $\epsilon_{2,t}$ having a finite eighth moment.

(see Wooldridge 1990, Definition A.1.). (iv) $G(\theta_0)' \Omega(\theta_0)^{-1} G(\theta_0)$ is nonsingular. (v) Assumption 1 of De Jong (1997) holds.

A7 is an extension of A6, since A7(i) implies A6(i), and A7(ii) implies both A6(ii) and A6(iii). The stronger mixingale properties of A7 permit a CLT to apply to $\widehat{g}(\theta_0)$ (see the proof of Theorem 3 in the Appendix).

THEOREM 3 (Asymptotic Normality). *Consider the estimator in (15) for the model of (8)–(11), where $\Lambda_T = \widehat{\Omega}(\theta)^{-1}$.* Assume that $\widehat{\Sigma}(\theta) \xrightarrow{p} \Sigma(\theta)$. *Let Assumptions A1–A7 hold. Then,*

$$\sqrt{T}\left(\widehat{\theta} - \theta_0\right) \xrightarrow{d} N\left(0, \left[G(\theta_0)' \Omega(\theta_0)^{-1} G(\theta_0)\right]^{-1}\right). \tag{16}$$

Admittedly, the assumptions supporting asymptotic normality are strong. For instance, (16) requires the eighth moment of $\epsilon_{2,t}$ to exist. This condition, however, is shared by the estimators in Kristensen and Linton (2006) and Baillie and Chung (2001) by nature of a shared reliance on the autocovariances of squared residuals for inference. The degree to which this condition limits the applicability of (15) is explored in the simulation experiments of the next section.

4. Monte Carlo

This section analyzes the finite sample performance of (15) with $\Lambda_T = \widehat{\Omega}(\theta)^{-1}$ benchmarked against the OLS estimator and the controlled regression (CR) estimator of (3), where $A\left(\mathcal{F}_{t-1}\right) = h_{22,t}^{-1}$, under the following simulation design:

$$\begin{aligned}
Y_{1,t} &= X_{1,t} + Y_{2,t} + \epsilon_{1,t}, \\
Y_{2,t} &= X_{1,t} + \epsilon_{2,t}, \\
X_{1,t} &\sim N(0,\ 1), \quad \epsilon_t = H_t^{1/2} \xi_t,
\end{aligned}$$

where each $\xi_{i,t}$ is distributed either as a $N(0,1)$ or standardized $\Gamma(2,1)$ random variable. In the specification of (11), $a_{11,0} = a_{12,0} = 0.10$, $a_{22,0} = 0.20$, $b_{11,0} = 0.80$, and $b_{12,0} = b_{22,0} = 0.70$. Conditional on these values, the constants $\omega_{ij,0}$ are set so that $Var\left[\epsilon_{1,t}\right] = Var\left[\epsilon_{2,t}\right] =$

1, and either $Cov\left[\epsilon_{1,t},\ \epsilon_{2,t}\right] = 0.20$ (the low correlation state) or $Cov\left[\epsilon_{1,t},\ \epsilon_{2,t}\right] = 0.40$ (the high correlation state). Given this specification, when $\xi_{i,t} \sim N(0,\ 1)$, the eighth moment of $\epsilon_{2,t}$ is finite (see Theorem 2 of Bollerslev 1986). On the other hand, when $\xi_{i,t}$ $\sim$standardized $\Gamma(2,1)$, the fourth moment of $\epsilon_{2,t}$ does not exist (see Corollary 6 of Carrasco and Chen 2002).

All simulations are conducted with 1,000 observations across 500 trials. When generating observations for each trial, the first 200 are dropped to avoid initialization effects. For each trial using (15), the starting values are the true parameter values. In addition, (15) is computed using either $K = 10$ or $K = 20$.[7] Summary statistics for all of the estimators include the mean and median bias, the standard deviation and decile range (defined as the difference between the 90th and the 10th percentiles), as well as the root mean squared error (RMSE) and median absolute error (MDAE), where both the RMSE and MDAE are measured with respect to the true parameter value. The median bias, decile range, and MDAE are robust measures of central tendency, dispersion, and efficiency, respectively, that are reported out of a concern over the existence of higher moments. For (15), the coverage rate for 95% confidence intervals as well as the rejection rate for the standard test for overidentification at a 5% level are also reported.

Table 1 summarizes the results for the OLS and CR estimators. As expected, the OLS estimates of β_{10} and β_{20} are biased. The absolute value of this bias increases when moving from the low correlation to the high correlation state and is generally higher when $\xi_{i,t}$ $\sim$standardized $\Gamma(2,1)$, the case reflecting a heavier-tailed process. In general, the magnitude of this bias is large. The CR estimator displays notably less bias than its OLS counterpart; however, the overall level of bias remains non-negligible, especially in the case of fat-tailed errors.

Table 2A summarizes the results for (15) when $K = 10$. In the case where $\xi_{i,t} \sim N(0,\ 1)$, the estimates of β_{10} and β_{20} are unbiased. The nuisance parameters are slightly biased, but this tendency does not effect the estimates from the conditional mean. The coverage rates tend to be too high and the rejection rates too low; however, the latter improves when moving to the high correlation state. Overall, the CUE in (15) offers a marked improvement

[7] When $K = 10$, (15) contains 40 moment conditions. When $K = 20$, the number of moment conditions is 80.

over both the OLS and CR estimators.

What is surprising in Table 2A is that (15) continues to produce unbiased estimates of β_{10} and β_{20} even when $\xi_{i,t}$ ~standardized $\Gamma(2,1)$. Irrespective of the correlation state, biases in the nuisance parameters increase significantly relative to the case where $\xi_{i,t} \sim N(0,\ 1)$, but this increase does not spill over onto the estimates from the conditional mean. Contrary to what the theory predicts, therefore, it seems as if (15) remains consistent even if the fourth moment of $\epsilon_{2,t}$ is not well defined. Also surprising is the finding that coverage rates for $\widehat{\beta}_1$ and $\widehat{\delta}$ correspond to the chosen confidence interval. The coverage rate for $\widehat{\beta}_2$, however, is too low. In addition, the overidentification test is significantly undersized.

Table 2B summarizes results for the CUE when $K = 20$. In general, these results (relative to those in Table 2A) confirm the CUE as being robust to many moments bias. For $\xi_{i,t} \sim N(0,\ 1)$ across both correlation states, moving from $K = 10$ to $K = 20$ results in diminished efficiency according to either the RMSE or MDAE. Coverage rates are generally improved, however, and the rejection rates are much closer to being appropriately sized. When $\xi_{i,t}$ ~standardized $\Gamma(2,1)$, the same results emerge as in the case where $K = 10$. Specifically, parameter estimates from the conditional mean remain unbiased even though the nuisance parameters display non-negligible bias, which, relative to the case where $K = 10$, is more severe. There is also a noticeable deterioration in coverage rates, counter-balanced against a marked improvement in rejection rates.

5. CAPM Betas

This section uses the CUE from section 3 to estimate CAPM betas for size, B/M, and momentum portfolios following the example in section 1.2. These portfolios are studied because they reflect the size, value, and momentum "premiums" that empirical applications of the CAPM struggle to explain. The returns are measured weekly (in percentage terms) from 10/6/67 through 9/28/07. Test results consider 20- and 10-year subperiods of this overall date range. The daily 25 size-B/M and 25 size-momentum return files (each 5×5 sorts with breakpoints determined by NYSE quintiles) formed from all securities traded on the NYSE, AMEX, and NASDAQ exchanges are used to construct the weekly return series.[8] The size portfolios considered are "Small," "Mid," and "Large." "Small" is the average of the five low market-cap portfolios, "Mid" the average of the five medium market-cap portfolios, and "Big" the average of the five large market-cap portfolios. The B/M portfolios considered are "Value," Neutral," and "Growth." Value" is the average of the five high B/M portfolios, "Neutral" the average of the five middle B/M portfolios, and "Growth" the average of the five low B/M portfolios.[9] Finally, the momentum portfolios considered are "Losers," "Draws," and "Winners." "Losers" is the average of the five low return-sorted portfolios, "Neutral" the average of the five middle return-sorted portfolios, and "Winners" the average of the five high return-sorted portfolios. The proxy return for the true market return is the CRSP value-weighted index return formed from all securities traded on the NYSE, AMEX, and NASDAQ exchanges. Excess returns are calculated using the one-month Treasury bill rate from Ibbotson Associates.

The most glaring take-away from Table 3, which summarizes estimation results for returns measured between 10/6/67 and 9/25/87, is that differences in beta estimates between OLS and the CUE are large (i.e., of economic significance) and statistically significant.[10] Moreover, this result is not impacted by the lag length chosen for the CUE. Since Theorem 1 nests the case of a zero covariance between structural errors–which, in the context of (7), means that there is no measurement error in the market return–this finding strongly suggests that the standard approach to estimating beta is biased. This finding is further supported by Table 4, which summarizes estimation results over the more-recent period 11/6/87 - 9/28/07, and by Tables 5–7, which consider ten-year subperiods of the two date ranges considered in Tables 3 and 4, respectively.[11]

Across the different portfolios, one can also observe an increase in the dispersion of the

[8]These return files are available on Kenneth French's website. Weekly returns are utilized because the CUE, which is based on higher moments, benefits from many observations in terms of finite sample performance. Weekly returns are selected over daily returns because the former reduces day-of-the-week and weekend effects as well as the effects of nonsynchronus trading and bid-ask bounce.

[9]Definitions for the "Small," "Large," "Value," and "Growth" portfolios are taken from Lewellen and Nagel (2006).

[10]Statistical significance is determined using 95% confidence intervals constructed from the standard errors of the CUE, which are consistent given general forms of heteroskedasticity and autocorrelation of the first order (i.e., $L = 2$).

[11]The subperiod 10/6/67 - 9/30/77 is not considered because the mean of the proxy return is negative.

beta estimates obtained using the CUE relative to OLS. Moreover, this increased dispersion does not seem to link to imprecision in the individual beta estimates, since the standard errors for the CUE are, at least, comparable in magnitude to their OLS counterparts. The implication, therefore, is that the beta estimates obtained under the CUE display elevated cross-sectional variation. In empirical asset pricing, betas obtained from time-series regressions are important for their assumed role in pricing expected returns in the cross-section. A well known empirical feature of cross-sectional expected returns is that (1) they tend to exhibit substantial variation, and (2) their associated betas vary correspondingly little (minor variations in betas cross-sectionally is evidenced in the first two panels of the Tables). This second feature explains the poor empirical performance of the CAPM, which uses individual asset sensitivities to the market return as its single pricing factor. Tables 3–7 suggest that this poor performance may be overstated; using consistent beta estimates may improve the ability of these estimates to explain variation in expected returns cross-sectionally.

Differences in alpha estimates between the CUE and OLS appear decidedly more muted. With minor exceptions, these estimates are statistically indistinguishable for the two 20-year time periods considered (see Tables 3 and 4). For the 10-year subperiods, however, statistically distinct alpha estimates do arise, and, when they do, increases in their magnitude (in absolute terms) under the CUE tend to explain the difference, as opposed to reductions in standard errors.

6. Conclusion

This paper presents a new method for estimating the linear triangular system, one which does not rely upon the existence of outside instruments for identification but, rather, a particular parametric form for the CH in the structural errors. This parametric form is common to empirical asset pricing specifications and tests. The estimator is shown to display the usual $\sqrt{T}$-asymptotics and is robust to many (potentially weak) moments bias. It also economizes on the number of nuisance parameters out of the CH process that need to be estimated.

The estimator is applied to estimating market betas in a CAPM setting. The resulting estimates differ significantly from the corresponding OLS estimates and appear to display increased cross-sectional variation. The two-pass method of Fama and MacBeth (1973) for testing asset pricing models relies upon time series beta estimates from the first pass. Works reliant upon this method have found the risk premium associated with these first-pass betas to be near zero or even negative (see, e.g., Jagannathan and Wang 1996 and Lettau and Ludvigson 2001). Inconsistent beta estimates from the first pass will affect the cross-sectional results from the second pass and may explain these counter-intuitive results. Increased cross-sectional variation in consistent beta estimates is a promising finding that supports this conjecture because of the empirical properties of cross-sectional expected returns.

Appendix

PROOF OF THEOREM 1: Given (14), first consider the case where $i = j = l = m = 2$. Then

$$Cov\left(\epsilon_{2,t}^2, U_{22,t-2}\right) = \phi_{22,0} Cov\left(\epsilon_{2,t}^2, U_{22,t-1}\right), \tag{17}$$

which identifies $\phi_{22,0}$ as

$$\phi_{22,0} = \left(Cov\left(\epsilon_{2,t}^2, U_{22,t-1}\right)' Cov\left(\epsilon_{2,t}^2, U_{22,t-1}\right)\right)^{-1} Cov\left(\epsilon_{2,t}^2, U_{22,t-1}\right)' Cov\left(\epsilon_{2,t}^2, U_{22,t-2}\right).$$

Next let $i = 1$, $j = 2$, and $l = m = 2$. In this case,

$$Cov\left(\epsilon_{1,t}\epsilon_{2,t}, U_{22,t-2}\right) = \phi_{12,0} Cov\left(\epsilon_{1,t}\epsilon_{2,t}, U_{22,t-1}\right),$$

the reduced form of which is

$$Cov\left(R_{1,t}\epsilon_{2,t}, U_{22,t-2}\right) = \phi_{12,0} Cov\left(R_{1,t}\epsilon_{2,t}, U_{22,t-1}\right) + \beta_{2,0}\left(\phi_{22,0} - \phi_{12,0}\right) Cov\left(\epsilon_{2,t}^2, U_{22,t-1}\right), \tag{18}$$

given (12) and (17). Finally, let $i = l = 1$ and $j = m = 2$. Then

$$Cov\left(\epsilon_{1,t}\epsilon_{2,t}, U_{12,t-2}\right) = \phi_{12,0} Cov\left(\epsilon_{1,t}\epsilon_{2,t}, U_{12,t-1}\right),$$

the reduced form of which simplifies to

$$Cov\left(R_{1,t}\epsilon_{2,t},U_{12,t-2}^{(R)}\right)=\phi_{12,0}Cov\left(R_{1,t}\epsilon_{2,t},U_{12,t-1}^{(R)}\right)+\beta_{2,0}\left(\phi_{22,0}-\phi_{12,0}\right)Cov\left(\epsilon_{2,t}^{2},U_{12,t-1}^{(R)}\right), \tag{19}$$

given (12), (17), (18), and the fact that

$$Cov\left(\epsilon_{2,t}^{2},U_{12,t-2}^{(R)}\right)=\phi_{22,0}Cov\left(\epsilon_{2,t}^{2},U_{12,t-1}^{(R)}\right).$$

Given (18) and (19),

$$\begin{bmatrix} Cov\left(R_{1,t}\epsilon_{2,t},U_{12,t-2}^{(R)}\right) \\ Cov\left(R_{1,t}\epsilon_{2,t},U_{22,t-2}\right) \end{bmatrix}=\Phi_{R}\lambda,$$

where $\lambda=\begin{bmatrix} \phi_{12,0} & \beta_{2,0}\left(\phi_{22,0}-\phi_{12,0}\right) \end{bmatrix}'$. Given A4, λ is identified as

$$\lambda=\left(\Phi_{R}^{'}\Phi_{R}\right)\Phi_{R}^{'}\begin{bmatrix} Cov\left(R_{1,t}\epsilon_{2,t},U_{12,t-2}^{(R)}\right) \\ Cov\left(R_{1,t}\epsilon_{2,t},U_{22,t-2}\right) \end{bmatrix},$$

from which $\phi_{12,0}$ is identified, and $\beta_{2,0}$ is identified conditional on the identification of both $\phi_{22,0}$ and $\phi_{12,0}$ and given A2. Finally, given A1, $\beta_{1,0}$ is identified as $\beta_{1,0}=E\left[X_{t}X_{t}'\right]^{-1}E\left[X_{t}\left(Y_{1,t}-Y_{2,t}\beta_{2,0}\right)\right]$.$\blacksquare$

PROOF OF THEOREM 2: Given A6(ii), $\left\{X_{t}\epsilon_{i,t}\right\}$ is uniformly integrable. A6(i) and A6(ii), therefore, allow of an application of Theorem 1 in Andrews (1988), which establishes Result R1: $T^{-1}\sum_{t}X_{t}\epsilon_{i,t}\xrightarrow{p}0$. Next, recursive substitution into (13) produces

$$z_{ij,0t}=\sum_{p=0}^{\infty}\psi_{p}W_{ij,t-p}, \tag{20}$$

where $\psi_{0}=1$ and $\psi_{p}=a_{ij,0}\phi_{ij,0}^{p-1}\ \forall\ p\geq1$. Since $\left\{\psi_{p}\right\}_{p=0}^{\infty}$ is absolutely summable given A2 and A3, and $E\left[\left|W_{ij,t}\right|\right]^{2}$ is finite given A3, $\{Z_{ij,t}\}$ is an L^{1} mixingale that is uniformly integrable. As a consequence, Theorem 1 of Andrews (1988) applies again to establish result R2: $T^{-1}\sum_{t}Z_{ij,t}\xrightarrow{p}\sigma_{ij,0}$. Next, (20) can also be used to show that

$\{v_{t,k}\}$ is a L^1 mixingale since $E\left[\left|W_{ij,t}\right|^2\right]$ is finite (see Hamilton 1994 p. 192-93 for a closely related proof). This result together with A6(iii) and R2 establishes Result R3: $T^{-1}\sum_t z_{ij,t}z_{lm,t-k} \xrightarrow{p} E\left[z_{ij,0t}z_{lm,0t-k}\right]$. Given results R1–R3, $\widehat{g}(\theta) \xrightarrow{p} \overline{g}(\theta)$. Since $\Lambda_T \xrightarrow{p} \Lambda_0$ by assumption, $\widehat{g}(\theta)' \Lambda_T \widehat{g}(\theta) \xrightarrow{p} \overline{g}(\theta)' \Lambda_0 \overline{g}(\theta)$ by continuity of multiplication. Finally, given Theorem 1, $\overline{g}(\theta)' \Lambda_0 \overline{g}(\theta)$ is uniquely minimized at $\theta = \theta_0$.■

PROOF OF THEOREM 3: Using well known results on derivatives of inverse matrices, the first order conditions for (15) with $\Lambda_T = \widehat{\Omega}(\theta)^{-1}$ are

$$\left[\widehat{G}\left(\widehat{\theta}\right)' \widehat{\Omega}(\theta)^{-1} - \widehat{g}\left(\widehat{\theta}\right)' \widehat{\Omega}\left(\widehat{\theta}\right)^{-1} \widehat{\Sigma}\left(\widehat{\theta}\right) \widehat{\Omega}\left(\widehat{\theta}\right)^{-1}\right] \widehat{g}\left(\widehat{\theta}\right) = 0.$$

Multiplying this expression by $\sqrt{T}$ and expanding $\widehat{g}\left(\widehat{\theta}\right)$ around θ_0 produces

$$\sqrt{T}\left(\widehat{\theta} - \theta_0\right) = \left[G(\theta_0)' \Omega(\theta_0)^{-1} G(\theta_0)\right]^{-1} G(\theta_0)' \Omega(\theta_0)^{-1} \sqrt{T}\widehat{g}(\theta_0),$$

given A7(iv) and the following Results: (R4) $\widehat{g}\left(\widehat{\theta}\right) \xrightarrow{p} \overline{g}(\theta_0) = 0$, $\widehat{\Sigma}\left(\widehat{\theta}\right) \xrightarrow{p} \Sigma(\theta_0)$, and $\widehat{G}\left(\widehat{\theta}\right) \xrightarrow{p} G(\theta_0)$ given Theorem 2 (specifically, $\widehat{G}\left(\widehat{\theta}\right) \xrightarrow{p} G(\theta_0)$ from the mixingale and uniform integrability properties of A6); (R5) $\widehat{\Omega}\left(\widehat{\theta}\right)^{-1} \xrightarrow{p} \Omega(\theta_0)^{-1}$ given Theorem 2, A7(ii), A7(iii), and Lemma 4.3 of Newey and McFadden (1994) applied to $a(z,\theta) = g_{t-s}(\theta)\, g_t(\theta)'$, where A7(iii) replaces the reliance on Khintchine's law of large numbers within the proof of this Lemma. Next, given absolute summability of $\{\psi_p\}_{p=0}^{\infty}$ (see the proof of Theorem 2), $\{z_{ij,0t}\}$ and $\{v_{t,k}\}$ are L^2 mixingales, since $E\left[\left|W_{ij,t}\right|^4\right]$ is finite under A7(ii). This result together with A7(i) establishes the L^2 mixingale property for $\{g_t(\theta_0)\}$, which satisfies the first element of Assumption 1 in De Jong (1997). Since the remaining elements hold under A7(v), $\sqrt{T}\widehat{g}(\theta_0) \xrightarrow{d} N(0, \Omega(\theta_0))$ by Theorem 1 of the aforementioned work, where $\Omega(\theta_0)$ is finite by A7(ii). The statement in (16) then follows by an application of the Slutzky theorem.■

References

[1] Andrews, D.W.K., 1988, Laws of large numbers for dependent non-identically distributed random variables, Econometric Theory, 4, 458-467.

[2] Baillie, R.T., and H. Chung, 2001, Estimation of GARCH models from the autocorrelations of the squares of a process, Journal of Time Series Analysis, 22, 631-650.

[3] Bollerslev, T., 1986, Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, 31, 307–327.

[4] Bollerslev, T., 1988, On the correlation structure for the generalized autoregressive conditional heteroskedastic process, Journal of Time Series Analysis, 9, 121-131.

[5] Bollerslev, T., 1990, Modelling the coherence in short run nominal exchange rates: a multivariate generalized ARCH model, Review of Economics and Statistics, 72, 498-505.

[6] Bollerslev, T., R.F Engle and J.M. Wooldridge, 1988, A capital asset pricing model with time-varying covariances, Journal of the Political Economy, 96, 116-131.

[7] Carrasco, M. and X. Chen, 2002, Mixing and moment properties of various GARCH and stochastic volatility models, Econometric Theory, 18, 17-39.

[8] Cragg, J. and S. Donald, 1997, Inferring the rank of a matrix, Journal of Econometrics, 76, 223-250.

[9] Engle, R.F and K.F. Kroner, 1995, Multivariate simultaneous generalized GARCH, Econometric Theory, 11, 121-150.

[10] Fama, E.F. and J. MacBeth, 1973, Risk, return and equilibrium: empirical tests, Journal of Political Economy, 81, 607-636.

[11] Hamilton, J.D., 1994, Time series analysis, Princeton University Press.

[12] Han, C. and P.C.B. Phillips, 2006, GMM with many moment conditions, Econometrica, 74, 147-192.

[13] Hansen, L.P., J. Heaton and A. Yaron, 1996, Finite-sample properties of some alternative GMM estimators, Journal of Business and Economic Statistics, 14, 262-280.

[14] Hansen, P.R. and A. Lunde, 2005, A forecast comparison of volatility models: does anything beat a GARCH(1,1)?, Journal of Applied Econometrics, 20, 873-889.

[15] He, C. and T. Teräsvirta, 1999, Properties of moments of a family of GARCH processes, Journal of Econometrics, 92, 173-192.

[16] Jagannathan, R. and Z. Wang, 1996, The conditional CAPM and the cross-section of expected returns, Journal of Finance, 53, 1285-1309.

[17] Kristensen, D. and O. Linton, 2006, A closed-form estimator for the GARCH(1,1)-model, Econometric Theory, 22, 323-327.

[18] Klein, R. and F. Vella, 2010, Estimating a class of triangular simultaneous equations models without exclusion restrictions, Journal of Econometrics, 154, 154-164.

[19] Lettau, M. and S. Ludvigson, 2001, Resurrecting the (C)CAPM: a cross-sectional test with risk premia that are time varying, Journal of Political Economy, 109, 1238-1287.

[20] Lewbel, A., 2010, Using heteroskedasticity to identify and estimate mismeasured and endogenous regressor models, Journal of Business and Economic Statistics, forthcoming.

[21] Lewellen, J. and S. Nagel, 2006, The conditional CAPM does not explain asset-pricing anomalies, Journal of Financial Economics, 82, 289-314.

[22] Lintner, J., 1965, The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets," Review of Economics and Statistics, 47, 13–37.

[23] Meng, J.G., G. Hu and J. Bai, 2011, OLIVE: a simple method for estimating betas when feactors are measured with error, Journal of Financial Research, 34, 27-60.

[24] Newey, W.K. and D. McFadden, 1994, Large sample estimation and hypothesis testing, in R.F. Engle and D. McFadded, eds, Handbook of Econometrics, Vol. 4, Amsterdam North Holland, chapter 36, 2111-2245.

[25] Newey, W.K. and R.J. Smith, 2004, Higher order properties of GMM and generalized empirical likelihood estimators, Econometrica, 72, 219-255.

[26] Newey, W.K and F. Windmeijer, 2009, Generalized method of moments with many weak moment conditions, Econometrica, 77, 687-719.

[27] Prono, T., 2010, GARCH-based identification and estimation of triangular systems, unpublished manuscript.

[28] Rigobon, R., 2003, Identification through heteroskedasticity, Review of Economics and Statistics, 85, 777-792.

[29] Rigobon, R. and B. Sack, 2003, Measuring the response of monetary policy to the stock market, Quarterly Journal of Economics, 118, 639-669.

[30] Roll, R., 1977, A critique of the asset pricing theory's tests: part I, Journal of Financial Economics, 4, 129–176.

[31] Rummery, S., F. Vella and M. Verbeek, 1999, Estimating the returns to education for Australian youth via rank-order instrumental variables, Labour Economics, 6, 491-507.

[32] Sentana, E. and G. Fiorentini, 2001, Identification, estimation and testing of conditionally heteroskedastic factor models, Journal of Econometrics, 102, 143-164.

[33] Sharpe, W.F., 1964, Capital asset prices: a theory of market equilibrium under conditions of risk, Journal of Finance, 19, 425–442.

[34] Stock, J. and J. Wright, 2000, GMM with weak identification, Econometrica, 68, 1055-1096.

[35] Vella, F. and M. Verbeek, 1997, Rank order as an instrumental variable, unpublished manuscript.

[36] Wooldridge, J.M., 1990, A unified approach to robust, regression-based specification tests, Econometric Theory, 6, 17-43.

TABLE 1

	S.E Dist.	Est.	Para.	Mean Bias	Med. Bias	SD	Dec. Rge.	RMSE	MDAE
$\rho = 0.20$	$N(0,1)$	OLS	β_1	-0.196	-0.197	0.057	0.146	0.205	0.197
			β_2	0.195	0.196	0.048	0.119	0.200	0.196
			δ	0.002	0.003	0.033	0.081	0.033	0.021
		CR	β_1	-0.066	-0.064	0.107	0.257	0.125	0.081
			β_2	0.064	0.064	0.100	0.248	0.119	0.077
			δ	0.001	0.002	0.026	0.067	0.026	0.018
	$\Gamma(2,1)$	OLS	β_1	-0.205	-0.204	0.070	0.182	0.217	0.204
			β_2	0.204	0.203	0.064	0.157	0.214	0.203
			δ	0.001	0.002	0.033	0.082	0.032	0.021
		CR	β_1	-0.057	-0.047	0.122	0.297	0.134	0.080
			β_2	0.056	0.044	0.119	0.283	0.131	0.071
			δ	0.000	0.001	0.027	0.071	0.027	0.018
$\rho = 0.40$	$N(0,1)$	OLS	β_1	-0.374	-0.374	0.061	0.159	0.379	0.374
			β_2	0.372	0.374	0.055	0.139	0.376	0.374
			δ	0.001	0.001	0.033	0.080	0.033	0.021
		CR	β_1	-0.125	-0.125	0.103	0.245	0.162	0.127
			β_2	0.124	0.123	0.097	0.228	0.158	0.124
			δ	0.001	0.000	0.027	0.066	0.027	0.019
	$\Gamma(2,1)$	OLS	β_1	-0.387	-0.387	0.088	0.223	0.397	0.387
			β_2	0.387	0.392	0.085	0.201	0.396	0.392
			δ	0.001	0.000	0.032	0.086	0.032	0.021
		CR	β_1	-0.121	-0.109	0.138	0.330	0.184	0.117
			β_2	0.121	0.104	0.136	0.322	0.182	0.113
			δ	0.000	0.000	0.027	0.070	0.027	0.019

Notes: Simulations are conducted using 1,000 observations across 500 trials. ρ is the correlation between structural errors. The true parameter vector is $\beta_{10} = \beta_{20} = \delta_0 = 1$. S.E. Dist. is the standardized error distribution, Para. the parameter estimate. CR is the controlled regression. Med. Bias is the median bias, SD the standard deviation, Dec. Rge. the decile range (measured as the difference between the 90th and 10th percentiles), RMSE the root mean squared error, and MDAE the median absolute error. RMSE and MDAE are measured with respect to the true parameter values.

TABLE 2A

	S.E Dist.	Para.	Mean Bias	Med. Bias	SD	Dec. Rge.	RMSE	MDAE	COV.	OVER
$\rho = 0.20$	$N(0,1)$	β_1	0.002	0.000	0.021	0.041	0.021	0.009	0.994	0.018
		β_2	-0.005	-0.004	0.019	0.041	0.020	0.010	0.990	
		δ	-0.001	0.000	0.023	0.052	0.023	0.011	0.960	
		ϕ_{12}	-0.010	-0.004	0.055	0.090	0.056	0.020		
		ϕ_{22}	-0.030	-0.014	0.051	0.107	0.059	0.016		
	$\Gamma(2,1)$	β_1	0.004	0.004	0.041	0.062	0.041	0.015	0.950	0.006
		β_2	-0.007	-0.007	0.042	0.063	0.042	0.015	0.846	
		δ	-0.002	-0.002	0.029	0.072	0.029	0.018	0.940	
		ϕ_{12}	-0.031	-0.012	0.119	0.295	0.123	0.059		
		ϕ_{22}	-0.082	-0.057	0.090	0.199	0.122	0.057		
$\rho = 0.40$	$N(0,1)$	β_1	0.001	0.000	0.014	0.035	0.014	0.008	1.000	0.030
		β_2	-0.004	-0.003	0.014	0.034	0.015	0.008	0.990	
		δ	-0.001	-0.001	0.021	0.046	0.021	0.009	0.974	
		ϕ_{12}	-0.014	-0.011	0.030	0.059	0.033	0.016		
		ϕ_{22}	-0.019	-0.011	0.034	0.049	0.039	0.013		
	$\Gamma(2,1)$	β_1	0.004	0.003	0.044	0.058	0.044	0.014	0.950	0.010
		β_2	-0.006	-0.006	0.044	0.058	0.045	0.015	0.866	
		δ	-0.002	0.000	0.027	0.067	0.027	0.016	0.950	
		ϕ_{12}	-0.044	-0.017	0.104	0.202	0.113	0.033		
		ϕ_{22}	-0.068	-0.036	0.081	0.179	0.106	0.036		

Notes: Simulations are conducted using 1,000 observations across 500 trials. For the CUE, $k = 10$, and $L = 1$. ρ is the correlation between structural errors, S.E. Dist. the standardized error distribution, Para. the parameter estimate. The true parameter vector is $\beta_{10} = \beta_{20} = \delta_0 = 1$. Med. Bias is the median bias, SD the standard deviation, Dec. Rge. the decile range (measured as the difference between the 90th and 10th percentiles), RMSE the root mean squared error, and MDAE the median absolute error. RMSE and MDAE are measured with respect to the true parameter values. COV. is the coverage rate for a 95% confidence interval, and OVER is the rejection rate for the standard test for overidentification restrictions.

TABLE 2B

	S.E Dist.	Para.	Mean Bias	Med. Bias	SD	Dec. Rge.	RMSE	MDAE	COV.	OVER
$\rho = 0.20$	$N(0,1)$	β_1	0.006	0.004	0.036	0.053	0.036	0.012	0.968	0.030
		β_2	-0.005	-0.005	0.034	0.049	0.034	0.012	0.930	
		δ	0.000	0.000	0.030	0.077	0.030	0.017	0.892	
		ϕ_{12}	-0.035	-0.019	0.102	0.224	0.108	0.037		
		ϕ_{22}	-0.067	-0.037	0.069	0.155	0.096	0.037		
	$\Gamma(2,1)$	β_1	0.003	0.006	0.045	0.061	0.045	0.016	0.900	0.028
		β_2	-0.002	-0.004	0.045	0.058	0.045	0.014	0.754	
		δ	-0.001	-0.002	0.032	0.078	0.032	0.022	0.858	
		ϕ_{12}	-0.040	-0.026	0.138	0.316	0.144	0.072		
		ϕ_{22}	-0.109	-0.097	0.087	0.206	0.140	0.097		
$\rho = 0.40$	$N(0,1)$	β_1	0.006	0.003	0.023	0.046	0.024	0.010	0.986	0.044
		β_2	-0.006	-0.005	0.022	0.041	0.023	0.010	0.952	
		δ	-0.001	-0.001	0.026	0.066	0.026	0.013	0.940	
		ϕ_{12}	-0.044	-0.032	0.073	0.118	0.086	0.035		
		ϕ_{22}	-0.049	-0.029	0.057	0.133	0.075	0.029		
	$\Gamma(2,1)$	β_1	0.004	0.005	0.041	0.063	0.041	0.015	0.936	0.042
		β_2	-0.006	-0.005	0.039	0.063	0.040	0.015	0.786	
		δ	-0.001	0.000	0.028	0.072	0.028	0.017	0.898	
		ϕ_{12}	-0.062	-0.045	0.119	0.273	0.134	0.063		
		ϕ_{22}	-0.095	-0.087	0.077	0.175	0.122	0.087		

Notes: Simulations are conducted using 1,000 observations across 500 trials. For the CUE, $k = 20$, and $L = 1$. ρ is the correlation between structural errors, S.E. Dist. the standardized error distribution, Para. the parameter estimate. The true parameter vector is $\beta_{10} = \beta_{20} = \delta_0 = 1$. Med. Bias is the median bias, SD the standard deviation, Dec. Rge. the decile range (measured as the difference between the 90th and 10th percentiles), RMSE the root mean squared error, and MDAE the median absolute error. RMSE and MDAE are measured with respect to the true parameter values. COV. is the coverage rate for a 95% confidence interval, and OVER is the rejection rate for the standard test for overidentification restrictions.

TABLE 3

	Size			B/M			Momentum		
	Small	Mid	Large	Value	Neutral	Growth	Losers	Draws	Winners
Panel A: Alpha (OLS)									
est.	0.045	0.060	0.016	0.113	0.053	-0.053	-0.106	0.045	0.132
std. error	0.045	0.026	0.013	0.033	0.023	0.028	0.043	0.022	0.033
Panel B: Beta (OLS)									
est.	0.925	0.953	0.950	0.890	0.882	1.196	1.152	0.898	1.061
std. error	0.030	0.019	0.008	0.022	0.017	0.020	0.027	0.013	0.032
Panel C: Alpha (CUE, $k = 12$)									
est.	0.069	0.075	0.021	0.130	0.069	-0.043	-0.186*	0.053	0.197
std. error	0.041	0.025	0.011	0.030	0.022	0.025	0.037	0.020	0.040
Panel D: Beta (CUE, $k = 12$)									
est.	0.761*	0.817*	0.902*	0.724*	0.755*	1.117*	1.376*	0.860*	0.628*
std. error	0.024	0.013	0.007	0.016	0.011	0.016	0.021	0.012	0.017
Panel E: Overidentification Test									
J-stat	33.44	37.02	41.48	33.13	33.72	30.49	32.87	36.00	33.94
p-value	0.793	0.648	0.450	0.804	0.783	0.885	0.813	0.692	0.775
Panel F: Alpha (CUE, $k = 24$)									
est.	0.086	0.084	0.018	0.138	0.074	-0.045	-0.195*	0.046	0.187
std. error	0.036	0.021	0.009	0.026	0.018	0.021	0.033	0.018	0.030
Panel G: Beta (CUE, $k = 24$)									
est.	0.715*	0.753*	0.908*	0.694*	0.751*	1.110*	1.368*	0.846*	0.615*
std. error	0.014	0.008	0.004	0.010	0.018	0.009	0.012	0.007	0.011
Panel H: Overidentification Test									
J-stat	72.43	80.74	83.76	76.76	73.61	74.59	62.62	68.54	86.82
p-value	0.899	0.722	0.637	0.819	0.880	0.863	0.985	0.947	0.546

Notes: The date range considered is 10/6/67 - 9/25/87. Estimates marked with a * have 95% confidence intervals that do not include the corresponding OLS estimate. All standard errors are consistent given general forms of heteroskedasticty and first-order autocorrelation (i.e., $L = 2$).

TABLE 4

	Size			B/M			Momentum		
	Small	Mid	Large	Value	Neutral	Growth	Losers	Draws	Winners
Panel A: Alpha (OLS)									
est.	0.049	0.040	0.027	0.094	0.063	-0.070	-0.129	0.059	0.148
std. error	0.048	0.028	0.023	0.034	0.026	0.036	0.058	0.026	0.039
Panel B: Beta (OLS)									
est.	0.810	0.917	0.879	0.779	0.811	1.160	1.139	0.761	1.121
std. error	0.039	0.023	0.029	0.032	0.020	0.027	0.050	0.028	0.038
Panel C: Alpha (CUE, $k = 12$)									
est.	0.074	0.055	0.057	0.118	0.079	-0.048	-0.215	0.082	0.201
std. error	0.040	0.026	0.018	0.030	0.025	0.031	0.050	0.023	0.032
Panel D: Beta (CUE, $k = 12$)									
est.	0.760*	0.829*	0.735*	0.615*	0.652*	1.206*	1.358*	0.631*	0.943*
std. error	0.023	0.018	0.012	0.018	0.015	0.015	0.035	0.013	0.017
Panel E: Overidentification Test									
J-stat	42.51	42.18	31.45	46.82	48.48	38.47	34.88	37.69	45.48
p-value	0.406	0.420	0.859	0.246	0.205	0.583	0.738	0.618	0.291
Panel F: Alpha (CUE, $k = 24$)									
est.	0.074	0.049	0.058*	0.122	0.082	-0.046	-0.202	0.072	0.189
std. error	0.034	0.023	0.014	0.025	0.021	0.027	0.038	0.019	0.028
Panel G: Beta (CUE, $k = 24$)									
est.	0.742*	0.818*	0.781*	0.609*	0.623*	1.220*	1.356*	0.606*	0.918*
std. error	0.014	0.010	0.006	0.012	0.009	0.009	0.018	0.008	0.011
Panel H: Overidentification Test									
J-stat	72.92	83.35	83.89	78.43	79.13	79.54	70.60	82.41	86.49
p-value	0.892	0.649	0.633	0.781	0.764	0.754	0.925	0.676	0.556

Notes: The date range considered is 11/6/87 - 9/28/07. Estimates marked with a * have 95% confidence intervals that do not include the corresponding OLS estimate. All standard errors are consistent given general forms of heteroskedasticty and first-order autocorrelation (i.e., $L = 2$).

TABLE 5

	Size			B/M			Momentum		
	Small	Mid	Large	Value	Neutral	Growth	Losers	Draws	Winners
Panel A: Alpha (OLS)									
est.	0.091	0.081	-0.002	0.130	0.068	-0.045	-0.081	0.048	0.141
std. error	0.053	0.031	0.016	0.039	0.025	0.036	0.052	0.026	0.046
Panel B: Beta (OLS)									
est.	0.826	0.922	0.972	0.803	0.873	1.177	1.001	0.859	1.110
std. error	0.040	0.024	0.010	0.033	0.018	0.025	0.035	0.017	0.041
Panel C: Alpha (CUE, $k = 12$)									
est.	0.096	0.069	0.023	0.207*	0.062	-0.086	-0.175*	0.037	0.178
std. error	0.040	0.026	0.016	0.033	0.020	0.032	0.046	0.021	0.037
Panel D: Beta (CUE, $k = 12$)									
est.	0.947*	0.974*	0.816*	0.575*	0.886	1.421*	1.287*	0.874	1.107
std. error	0.027	0.019	0.008	0.020	0.015	0.019	0.021	0.014	0.026
Panel E: Overidentification Test									
J-stat	31.87	31.49	37.34	39.55	34.84	36.387	44.13	41.27	29.14
p-value	0.846	0.858	0.634	0.535	0.740	0.676	0.341	0.459	0.917
Panel F: Alpha (CUE, $k = 24$)									
est.	0.061	0.069	0.027	0.184*	0.075	-0.131*	-0.188*	0.041	0.185
std. error	0.033	0.019	0.014	0.027	0.016	0.026	0.036	0.016	0.031
Panel G: Beta (CUE, $k = 24$)									
est.	0.959*	0.983*	0.794*	0.579*	0.890*	1.440*	1.205*	0.875*	1.137
std. error	0.017	0.010	0.004	0.012	0.008	0.012	0.012	0.007	0.016
Panel H: Overidentification Test									
J-stat	73.55	72.20	84.14	81.37	77.25	79.88	79.81	77.89	74.79
p-value	0.881	0.903	0.626	0.705	0.808	0.745	0.747	0.794	0.859

Notes: The date range considered is 10/7/77 - 9/25/87. Estimates marked with a * have 95% confidence intervals that do not include the corresponding OLS estimate. All standard errors are consistent given general forms of heteroskedasticty and first-order autocorrelation (i.e., $L = 2$).

TABLE 6

	Size			B/M			Momentum		
	Small	Mid	Large	Value	Neutral	Growth	Losers	Draws	Winners
Panel A: Alpha (OLS)									
est.	0.029	0.039	0.008	0.079	0.047	-0.074	-0.134	0.044	0.112
std. error	0.054	0.031	0.017	0.035	0.025	0.041	0.058	0.024	0.038
Panel B: Beta (OLS)									
est.	0.693	0.873	1.031	0.789	0.821	1.075	1.007	0.786	1.085
std. error	0.040	0.026	0.013	0.026	0.019	0.031	0.045	0.020	0.028
Panel C: Alpha (CUE, $k = 12$)									
est.	0.060	0.051	-0.012	0.040	0.028	0.004	-0.226*	0.025	0.170
std. error	0.043	0.028	0.015	0.033	0.020	0.038	0.047	0.019	0.036
Panel D: Beta (CUE, $k = 12$)									
est.	0.624	0.845	1.092*	0.776	0.885*	0.938*	1.349*	0.828	0.866*
std. error	0.058	0.030	0.017	0.042	0.028	0.037	0.058	0.031	0.039
Panel E: Overidentification Test									
J-stat	41.92	33.50	29.08	42.88	42.04	38.72	49.45	41.86	38.76
p-value	0.431	0.791	0.919	0.391	0.426	0.572	0.171	0.433	0.571
Panel F: Alpha (CUE, $k = 24$)									
est.	0.123*	0.074	-0.015	0.120	0.054	0.009	-0.240*	0.047	0.240
std. error	0.033	0.020	0.011	0.021	0.016	0.030	0.036	0.015	0.027
Panel G: Beta (CUE, $k = 24$)									
est.	0.650*	0.854	1.107*	0.748*	0.881*	1.114*	1.358*	0.830*	0.864*
std. error	0.018	0.013	0.007	0.018	0.010	0.014	0.018	0.008	0.019
Panel H: Overidentification Test									
J-stat	84.65	82.24	72.25	82.70	80.20	72.61	84.03	86.34	82.69
p-value	0.611	0.681	0.902	0.668	0.736	0.90	0.629	0.560	0.668

Notes: The date range considered is 11/6/87 - 9/26/97. Estimates marked with a * have 95% confidence intervals that do not include the corresponding OLS estimate. All standard errors are consistent given general forms of heteroskedasticty and first-order autocorrelation (i.e., $L = 2$).

TABLE 7

	Size			B/M			Momentum		
	Small	Mid	Large	Value	Neutral	Growth	Losers	Draws	Winners
Panel A: Alpha (OLS)									
est.	0.087	0.048	0.021	0.107	0.077	-0.052	-0.102	0.070	0.190
std. error	0.078	0.047	0.039	0.056	0.044	0.058	0.099	0.044	0.068
Panel B: Beta (OLS)									
est.	0.863	0.937	0.810	0.775	0.807	1.199	1.199	0.750	1.138
std. error	0.051	0.032	0.036	0.045	0.028	0.034	0.072	0.039	0.053
Panel C: Alpha (CUE, $k = 12$)									
est.	0.221	0.110	0.044	0.207*	0.126	0.000	-0.161	0.155*	0.301
std. error	0.065	0.044	0.028	0.050	0.051	0.044	0.079	0.040	0.051
Panel D: Beta (CUE, $k = 12$)									
est.	0.561*	0.718*	0.696*	0.461*	0.398*	1.239*	1.337*	0.469*	0.783*
std. error	0.021	0.016	0.011	0.019	0.015	0.013	0.032	0.013	0.017
Panel E: Overidentification Test									
J-stat	44.00	35.78	34.55	40.36	44.77	33.89	38.62	39.77	41.66
p-value	0.346	0.701	0.751	0.499	0.317	0.776	0.577	0.525	0.442
Panel F: Alpha (CUE, $k = 24$)									
est.	0.125	0.078	0.038	0.148	0.120	-0.057	-0.116*	0.104	0.229
std. error	0.043	0.035	0.018	0.034	0.033	0.030	0.050	0.030	0.037
Panel G: Beta (CUE, $k = 24$)									
est.	0.615*	0.738*	0.697*	0.503*	0.509*	1.231*	1.294*	0.514*	0.877*
std. error	0.013	0.009	0.005	0.011	0.009	0.009	0.016	0.007	0.011
Panel H: Overidentification Test									
J-stat	81.25	77.66	72.05	92.11	86.37	64.19	68.15	79.47	101.73
p-value	0.708	0.799	0.905	0.390	0.559	0.978	0.951	0.755	0.168

Notes: The date range considered is 10/3/97 - 9/28/07. Estimates marked with a * have 95% confidence intervals that do not include the corresponding OLS estimate. All standard errors are consistent given general forms of heteroskedasticty and first-order autocorrelation (i.e., $L = 2$).