DEAKIN UNIVERSITY AUSTRALIA

Faculty of Business and Law

School of Accounting, Economics and Finance

School Working Paper

Financial Econometrics Series

SWP 2014/01

Testing for Predictability in Conditionally Heteroskedastic Stock Returns

J. Westerlund and P.K. Narayan

TESTING FOR PREDICTABILITY IN CONDITIONALLY HETEROSKEDASTIC STOCK RETURNS*

Joakim Westerlund†
Deakin University
Australia

Paresh Narayan
Deakin University
Australia

September 9, 2013

Abstract

The difficulty of predicting stock returns has recently motivated researchers to start looking for more powerful tests, and the current paper takes a step in this direction. Unlike existing tests, the test proposed here exploits the information contained in the heteroskedasticity of returns, which is expected to lead to higher power, a result that is confirmed by our results. In order to also maintain good size accuracy, subsample critical values are used.

JEL Classification: C12; C22; G1.

Keywords: Predictability; FQGLS; Conditional heteroskedasticity; Subsampling; Stock returns.

1 Introduction

Consider the vector $(y_t, x_t)'$, observable for $t = 1, ..., T$ time periods. The data generating process (DGP) of this vector is given by

$$y_t = \theta + \beta x_{t-1} + \epsilon_{y,t}, \tag{1}$$

$$x_t = \mu(1-\rho) + \rho x_{t-1} + \epsilon_{x,t}, \tag{2}$$

$$\epsilon_{y,t} = \gamma \epsilon_{x,t} + \epsilon_{y.x,t}, \tag{3}$$

*The first author would like to thank the Jan Wallander and Tom Hedelius Foundation for financial support under research grant number P2005–0117:1.

†Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance, Melbourne Burwood Campus, 221 Burwood Highway, VIC 3125, Australia. Telephone: +61 3 924 46973. Fax: +61 3 924 46283. E-mail address: j.westerlund@deakin.edu.au.

where $x_0 = 0$, and $\epsilon_{x,t}$ and $\epsilon_{y.x,t}$ are mean zero disturbance terms that are independent of each other. Equation (1) is the prototypical predictive regression model that has been widely used in the finance literature to test whether stock returns (y_t) can be predicted using some other macroeconomic or financial variable (x_t). The standard econometric approach is to fit (1) by least squares (LS), and to test whether $\beta = 0$ using a conventional t-test. The main finding of such tests is that the observed value of the t-statistic is typically greater than two. Most earlier studies, which tended to rely on normal critical values, were therefore able to reject the null hypothesis. However, it has since then become clear that the standard distribution theory for stationary processes can be quite misleading when testing the predictability of returns, and that some of the rejections might actually be due to size distortions. The literature has identified two major problems, which both stems from the observed behavior of the predictors.

The first problem is if $\gamma \neq 0$, such that $\epsilon_{x,t}$ and $\epsilon_{y,t}$ are correlated, which raises the issue of endogeneity bias. For example, if x_t is dividend yield, then an increase in the stock price will lower dividends and raise returns.

The second problem is that, although according to theory many of the predictors used should be stationary, empirically most predictors are only slowly mean-reverting, and the evidence that they do not contain unit roots is weak. Standard asymptotic theory, which presumes that $|\rho| < 1$, is therefore likely to be inappropriate. In fact, even if $\{x_t\}_{t=1}^{T}$ is known to be stationary, if $\rho \approx 1$, the standard asymptotic theory is likely to provide a poor approximation in small samples (Elliott and Stock, 1994).

These econometric problems have recently caused some researchers to consider an alternative framework based on local asymptotic theory (see, for example, Cavanagh et al., 1995; Elliott and Stock, 1994; Lanne, 2002; Torous et al., 2004). This theory has not only made it possible to study in a coherent framework the effects of both endogeneity ($\gamma \neq 0$) and near unit roots ($\rho \approx 1$), but has also led to the development of several new, improved, tests, such as the uniformly most powerful tests of Lewellen (2004), and Campbell and Yogo (2006) that make better use of the predictor information.

The point here is that these developments all stem from the observed behavior of x_t. In view of this, it is quite surprising to find that there has been no attempts to account for the part of the variation in returns that cannot be attributed to x_t. In fact, most studies tend to view this variation as pure noise. The following quotation, taken from Campbell and Yogo

(2006, page 29), serves as an illustration:

> A regression of stock returns onto a lagged financial variable has low power because stock returns are extremely noisy. If we can eliminate some of this noise, we can increase the power of the test. When the innovations to returns and the predictor variable are correlated, we can subtract off the part of the innovation to the predictor variable that is correlated with returns to obtain a less noisy dependent variable for our regression.

Thus, stock returns are noisy, and by conditioning on $\epsilon_{x,t}$, we can reduce that noise, leading to a more powerful test. But what about the part of the variation in the returns that cannot be conditioned out in this way? In particular, what about volatility clustering, one of most well-documented features of stock returns? Intuitively, if the endogeneity is important for power, then there should be some merit in accounting also for heteroskedasticity. This is a potentially very important issue, because even if returns are truly predictable, the noisiness of the regression error ($\epsilon_{y,t}$) may well overwhelm the signal coming from x_t, causing deceptive inference (Ferson et al., 2003).[1]

Motivated by the above observation, the current paper develops a test based on generalized LS (GLS). In particular, a feasible quasi-GLS (FQGLS) t-test is proposed that is robust to misspecification of the conditional variance, and that is asymptotically equivalent to GLS if the model is correct. We show, both analytically and using simulations, that the information contained in the heteroskedasticity is useful, and that the power of the FQGLS test can be increased above that achievable by existing LS-based tests that do not make use of this information. Unfortunately, the asymptotic test depends on nuisance parameters reflecting the persistency, endogeneity and heteroskedasticity of the data, which means that while the test has good power properties, it can be size distorted. In order to account for this lack of robustness, a subsample FQGLS test is proposed that has correct asymptotic size even in the presence nuisance parameters, which stands in sharp contrast to many of the existing Bonferroni-type tests that are known to be conservative (see, for example, Campbell and Yogo, 2006; Cavanagh et al., 1995; Lewellen, 2004; Torous et al., 2004).

The rest of the paper is organized as follows. Section 2 outlines the FQGLS test and Section 3 discusses the assumptions under which which we will be working. Section 4 derives

[1]The "spurious regression problem" in finance refers to the case when expected return is highly persistent but this persistency is dwarfed by the noise coming from the regression error (Ferson et al., 2003).

the asymptotic distribution of the FQGLS t-statistic, which is compared to the asymptotic distributions of the LS and infeasible quasi-GLS (QGLS) t-statistics. Monte Carlo and empirical results are reported in Sections 6 and 7, respectively. Section 8 concludes. All proofs are provided in Appendix.

2 Test procedure

The DGP in (1) and (2) has become very popular in recent years, and the main reason for this is its ability to capture both the endogeneity and persistency of most commonly used predictors. However, there is at least one empirical regularity that is not captured here, namely, conditional heteroskedasticity. Let us therefore define

$$E(\epsilon_{y,t}^2|F_{t-1}) = \sigma_{y,t}^2 = \gamma^2 E(\epsilon_{x,t}^2|F_{t-1}) + E(\epsilon_{y.x,t}^2|F_{t-1}) = \gamma^2\sigma_{x,t}^2 + \sigma_{y.x,t}^2, \tag{4}$$

where F_t is the sigma-field generated by $\{(x_n, y_n)\}_{n=1}^t$.

One of the most well-known results from classical regression theory is that LS is inefficient in the presence of heteroskedasticity. Our approach is therefore based on GLS. However, since we do not assume knowledge of $\sigma_{y,t}^2$, the conditional variance of the error term in (1), the specific estimator considered here is more appropriately referred to as FQGLS, and the proposed predictability test is simply the FQGLS t-statistic.

The FQGLS estimator relies on having an estimator $\hat{s}_{y,t}^2$ of $\sigma_{y,t}^2$. In this paper $\hat{s}_{y,t}^2$ is the fitted value from a parametric specification of the conditional heteroskedasticity, such as an ARCH model, that is estimated using the LS residuals from (1) and (2). We do not assume that the conditional heteroskedasticity estimator is consistent. Therefore, the chosen parametric model need not be correct, as when specifying an ARCH model when the true model is GARCH. The reason is that $\{\hat{s}_{y,t}^2\}_{t=2}^T$ is such that they approximate a stationary F_{t-1}-adapted sequence $\{s_{y,t}^2\}_{t=2}^T$ in the sense that certain normalized sums have the same asymptotic distribution whether $\hat{s}_{y,t}^2$ or $s_{y,t}^2$ appears in the sum.

As an example, assume for simplicity that $\theta = \gamma = 0$ is known, such that $y_t = \beta x_{t-1} + \epsilon_{y,t}$ and $\sigma_{y,t}^2 = \sigma_{y.x,t}^2$. Let $\hat{\epsilon}_{y,t} = \hat{\epsilon}_{y,t}(\hat{\beta}) = y_t - \hat{\beta}x_{t-1}$, where $t = 2, ..., T$ and $\hat{\beta}$ is the LS estimator of β. The results allow for the case when $\hat{s}_{y,t}^2 = s_{y,t}^2(\hat{\lambda}_0, \hat{\lambda}_1, \hat{\beta}) = \hat{\lambda}_0 + \hat{\lambda}_1\hat{\epsilon}_{y,t-1}^2(\hat{\beta})$ ($t = 3, ..., T$) is the fitted value from a regression of $\hat{\epsilon}_{y,t}$ onto a constant and $\hat{\epsilon}_{y,t-1}$. While $\hat{\beta}$ is consistent for β, the probability limits of $\hat{\lambda}_0$ and $\hat{\lambda}_1$ are given by the "pseudo-true" values λ_0 and λ_1.

Hence, in this case $s^2_{y,t} = s^2_{y,t}(\lambda_0, \lambda_1, \beta) = \lambda_0 + \lambda_1 \hat{\epsilon}^2_{y,t-1}(\beta)$, where $\hat{\epsilon}_{y,t}(\beta) = \epsilon_{y,t}$; that is, $s^2_{y,t}$ is just $\hat{s}^2_{y,t}$ with the estimation error eliminated.

To define $\hat{s}^2_{y,t}$ in the general case, it is convenient to introduce $y^d_t = y_t - \overline{y}$ and $\overline{y} = \sum_{t=1}^T y_t/T$ with similar definitions of x^d_t and x^d_{t-1}. In this notation, $\hat{\epsilon}_{y,t} = \hat{\epsilon}_{y,t}(\hat{\beta}) = \hat{y}^d_t - \hat{\beta} x^d_{t-1}$ and $\hat{\epsilon}_{x,t} = \hat{\epsilon}_{x,t}(\hat{\rho}) = x^d_t - \hat{\rho} x^d_{t-1}$, where $\hat{\beta}$ ($\hat{\rho}$) is the LS slope in a regression of y^d_t (x^d_t) onto x^d_{t-1}. Let $\hat{\epsilon}_{y.x,t} = \hat{\epsilon}_{y.x,t}(\hat{\beta}, \hat{\rho}) = \hat{\epsilon}_{y,t}(\hat{\beta}) - \hat{\gamma} \hat{\epsilon}_{x,t}(\hat{\rho})$, where $\hat{\gamma} = \hat{\gamma}(\hat{\beta}, \hat{\rho})$ is the LS slope in a regression of $\hat{\epsilon}_{y,t}$ onto $\hat{\epsilon}_{x,t}$. Let $m \in \{x, y.x\}$ and denote by $\hat{s}^2_{m,t}$ the fitted value from a LS regression with $\hat{\epsilon}^2_{m,t}$ as the dependent variable. In principle, this fit can come from any estimable regression model. The list of potential regressors is very long and includes, for example, dummy variables, known nonlinear functions of the lags of $\hat{\epsilon}^2_{x,t}$ and $\hat{\epsilon}^2_{y.x,t}$, and other stationary predetermined variables. However, autoregressive (AR) models are most common, and in this paper we therefore focus on the case when $\hat{s}^2_{y.x,t} = s^2_{y.x,t}(\hat{\beta}, \hat{\rho}, \hat{\lambda}_{y.x})$ and $\hat{s}^2_{x,t} = s^2_{x,t}(\hat{\rho}, \hat{\lambda}_x)$, where

$$
\begin{aligned}
s^2_{x,t} &= s^2_{x,t}(\rho, \lambda_x) = \lambda'_x \hat{e}_{x,t}(\rho), &(5)\\
s^2_{y.x,t} &= s^2_{y.x,t}(\beta, \rho, \lambda_{y.x}) = \lambda'_{y.x} \hat{e}_{y.x,t}(\beta, \rho), &(6)
\end{aligned}
$$

with $t = q_m + 2, ..., T$, $\lambda_m = (\lambda_{0,m}, ..., \lambda_{j,q_m})'$, $\hat{e}_{m,t} = (1, \hat{\epsilon}_{m,t-1}, ..., \hat{\epsilon}_{m,t-q_m})'$ and $\hat{\lambda}_m$ is obtained by regressing $\hat{\epsilon}^2_{m,t}$ onto $\hat{e}_m$. Hence, since $\hat{\gamma}$ is consistent for γ, we have that $\hat{s}^2_{y,t}$ can be written $\hat{s}^2_{y,t} = s^2_{y,t}(\hat{\beta}, \hat{\rho}, \hat{\gamma}, \hat{\lambda}_{y.x}, \hat{\lambda}_x)$, where

$$
s^2_{y,t} = s^2_{y,t}(\beta, \rho, \gamma, \lambda_{y.x}, \lambda_x) = \gamma^2 s^2_{x,t}(\rho, \lambda_x) + s^2_{y.x,t}(\beta, \rho, \lambda_{y.x}). \qquad (7)
$$

The above discussion suggests the following stepwise procedure to obtain $\hat{s}^2_{y,t}$:

1. Obtain $\hat{\epsilon}_{y,t}$ and $\hat{\epsilon}_{x,t}$ by fitting (1) and (2) by LS.

2. Fit (3) by LS after replacing $\epsilon_{y,t}$ onto $\epsilon_{x,t}$ by $\hat{\epsilon}_{y,t}$ and $\hat{\epsilon}_{x,t}$, respectively. Obtain $\hat{\epsilon}_{y.x,t} = \hat{\epsilon}_{y,t} - \hat{\gamma} \hat{\epsilon}_{x,t}$, where $\hat{\gamma}$ is the estimated LS slope.

3. Retrieve $\hat{s}^2_{m,t}$ as the fitted value from a regression of $\hat{\epsilon}^2_{m,t}$ onto $\hat{e}_{m,t}$.

4. Compute $\hat{s}^2_{y,t} = \hat{\gamma}^2 \hat{s}^2_{x,t} + \hat{s}^2_{y.x,t}$.

Given $\hat{s}^2_{y,t}$ the computation of the FQGLS t-statistic is very simple;

$$
t\text{–FQGLS} = \frac{\sum_{t=q_m+2}^T \hat{s}^{-2}_{y,t} x^d_{t-1} y^d_t}{\sqrt{\sum_{t=q_m+2}^T \hat{s}^{-2}_{y,t} (x^d_{t-1})^2}},
$$

where $q = \max\{q_x, q_{y.x}\}$.

3 Assumptions

The basic DGP is given by (1)–(3), where $\epsilon_{x,t}$, $\epsilon_{x.y,t}$, β and ρ are assumed to satisfy Assumptions EPS, BET and RHO.

Assumption EPS.

(i) $\epsilon_{x,t}$ and $\epsilon_{y.x,t}$ are mutually independent with $E(\epsilon_{m,t}|F_{t-1}) = 0$, $E(\epsilon_{m,t}^2) = \sigma_m^2 \in (0, \infty)$, and $E(\epsilon_{m,t}^4) < \infty$ for $m \in \{x, y.x\}$;

(ii) $x_0 = 0$.

The requirement that $\epsilon_{x,t}$ and $\epsilon_{y.x,t}$ are serially uncorrelated is restrictive, but is made here in order to facilitate discussion and to focus on the essence of the problem, namely, the treatment of the conditionditional heteroskedasticity. In Section 3 we consider a subsample FQGLS-based tests that do not require serially uncorrelated innovations. Similarly, as long as $x_0 = O_p(1)$, the initialization does not affect the results. Hence, we can just as well set $x_0 = 0$. While the no serial correlation and zero initial value requirements can be relaxed, the assumption of constant unconditional variances is crucial. If the unconditional homoskedasticity assumption is violated the properties of the FQGLS test procedure become suspect, and in that case it is better to consider approaches that are robust to such unconditional heteroskedasticity (see Smeekes and Taylor, 2012, and the references provided therein). However, unconditional homoskedasticity does not rule out models of stochastic volatility (SV). In Section 4 we elaborate on this.

Assumption BET.

$$\beta = \frac{b}{T},$$

where $|b| < \infty$ is a constant not depending on T.

Assumption RHO.

$$\rho = 1 + \frac{c}{T},$$

where $c \in (-\infty, 0]$ is a constant not depending on T.

The local-to-zero specification in Assumption BET is very convenient, because it nests both the null and local alternative hypotheses, as determined by the drift parameter b. If $b = 0$ ($\beta = 0$), then we are under the no predictability null, whereas if $b \neq 0$, then we are under the local alternative with $\beta \to 0$ as $T \to \infty$. The main reason for considering a local alternative of this type is that it allows us to evaluate power analytically. This is in contrast to the conventional formulation where $\beta \neq 0$ is assumed to be fixed, in which case we only learn if the test is consistent and, if so, at what rate. Similarly, since the predictor can be quite persistent, we follow, for example, Cavanagh et al. (1995), Elliott and Stock (1994), and Lanne (2002) and assume that ρ is local-to-unity with c measuring the degree of persistency in $\{x_t\}_{t=1}^T$. If $c = 0$, then $\{x_t\}_{t=1}^T$ has an exact unit root, whereas if $c < 0$, then $\{x_t\}_{t=1}^T$ is "locally stationary" in the sense that ρ approaches one from below.

Assumption HET.

(i) $\{s_{y,t}^2\}_{t=2}^T$ is stationary and strong mixing with mixing coefficients $\alpha_n = O(n^{-3r/(r-3)})$ for $r > 3$, $n \geq 1$, $E(s_{y,t}^{-2}) = \phi_y \in (0, \infty)$ and $E(s_{y,t}^{-4}) \in (0, \infty)$;

(ii) $\lambda_{0,m} > 0$ and $\lambda_{1,m}, ..., \lambda_{q_m,m} \geq 0$ for $m \in \{x, y.x\}$;

(iii) $q_m < \infty$.

Note how Assumption HET does not put any restrictions on $\sigma_{y,t}^2$, the true conditional variance. The FQGLS t-test is therefore widely applicable. The specification of $\sigma_{y,t}^2$ is not irrelevant, however, as only when $s_{y,t}^2 = \sigma_{y,t}^2$ will FQGLS be efficient (see Section 3). In this sense (5) and (6) put restrictions on the types of conditional volatility models in which FQGLS can be expected to be efficient. Some examples of such models are given subsequently.

Example 1. An often-used multivariate conditional heteroskedasticity model is the constant-correlation ARCH of Bollerslev (1990). In this case,

$$E(\epsilon_t \epsilon_t' | F_{t-1}) = \Sigma_{\epsilon,t} = \begin{bmatrix} \sigma_{x,t}^2 & \sigma_{xy,t} \\ \sigma_{xy,t} & \sigma_{y,t}^2 \end{bmatrix} = D\Theta_t D' = \begin{bmatrix} \sigma_{x,t}^2 & \gamma\sigma_{x,t}^2 \\ \gamma\sigma_{x,t}^2 & \sigma_{y.x,t}^2 + \gamma^2\sigma_{x,t}^2 \end{bmatrix},$$

where $\epsilon_t = (\epsilon_{x,t}, \epsilon_{y.x,t})$,

$$D = \begin{bmatrix} 1 & 0 \\ \gamma & 1 \end{bmatrix},$$

$\Theta_t = \text{diag}(\sigma^2_{x,t}, \sigma^2_{y.x,t})$, $\sigma^2_{m,t} = \alpha_{0,m} + \sum_{j=1}^{p_m} \alpha_{j,m}\varepsilon^2_{m,t-j}$, $\varepsilon_t = (\varepsilon_{x,t}, \varepsilon_{y.x,t})'^{-1}\epsilon_t$, $\alpha_{0,m} > 0$, and $\alpha_{1,m}, ..., \alpha_{p_m,m} \geq 0$. Since $\varepsilon_{x,t} = \epsilon_{x,t}$ and $\varepsilon_{y.x,t} = \epsilon_{y,t} - \gamma\epsilon_{x,t} = \epsilon_{y.x,t}$, this is identically the structure in (5)–(7) with $q_m \geq p_m$ lags.

Example 2. The fact that the conditional variance equation is estimated using LS precludes generalized ARCH (GARCH) terms. However, since the purpose here is just to obtain the fitted value (and not to estimate the coefficients of the GARCH equation), the assumption that (5) and (6) are of finite order is not a restriction, because any GARCH model can be approximated arbitrarily well by taking q_m sufficiently large (see Andrews and Guggenberger, 2009, Section 7), a result that is verified in small samples in Section 4.[2]

Example 3. As mentioned in Section 2, while we focus on the case when (5) and (6) have an AR structure, this is not necessary. In fact, (5) and (6) can be easily generalized to accommodate, dummy variables, known (nonlinear) functions of the lags of $\hat{\epsilon}^2_{x,t}$ and $\hat{\epsilon}^2_{y.x,t}$, as well as other predetermined regressors. "X–ARCH" models of the type $\sigma^2_{y.x,t} = \alpha_{0,y.x} + \alpha_{1,y.x}\epsilon^2_{y.x,t-1} + \alpha_{2,y.x}\epsilon^2_{x,t-1}$ (see Brenner et al., 1996) can therefore be accommodated, as can asymmetric threshold ARCH models of the type $\sigma^2_{y.x,t} = \alpha_{0,y.x} + \alpha_{1,y.x}\epsilon^2_{y.x,t-1}1(\epsilon_{y.x,t-1} > 0) + \alpha_{2,y.x}\epsilon^2_{y.x,t-1}1(\epsilon_{y.x,t-1} \leq 0)$, where $1(A)$ is the indicator function for the event A (see Glosten et al., 1993).

4 Asymptotic results

In this section, we provide results for the (infeasible) quasi-GLS (QGLS) t-statistic based on $s^2_{y,t}$ rather than $\hat{s}^2_{y,t}$. We then show that this statistic is asymptotically equivalent to t–FQGLS. However, to be able to analyze the effect of the accounting for conditional heteroskedasticity, we start by considering the LS estimator.

Most, if not all, studies concerned with the testing of the predictability hypothesis are based on the LS estimator $\hat{\beta}$ of β and its t-statistic (see, for example, Stambaugh, 1999; Elliott and Stock, 1994; Lanne, 2002; Torous et al., 2004), as given by

$$t\text{–LS} = \frac{\sum_{t=2}^{T} x^d_{t-1} y^d_t}{\hat{\sigma}_y \sqrt{\sum_{t=2}^{T} (x^d_{t-1})^2}},$$

[2]Of course, the proofs with $q_m \to \infty$ will be more complex, although the current ones can probably be extended along the lines of Said and Dickey (1986).

where $\hat{\sigma}_y^2 = T^{-1}\sum_{t=2}^{T}\hat{\epsilon}_{y,t}^2$. The asymptotic distribution of this test statistic under Assumptions EPS, BET and RHO is given in Proposition 1. Anticipating this, it is useful to define

$$U(\rho_{xy}, b/\gamma, c) = \rho_{xy} DF(b/\gamma, c) + \sqrt{1-\rho_{xy}^2}Z, \tag{8}$$

where

$$DF(b/\gamma, c) = \frac{b}{\gamma}\sqrt{\int_0^1 J_{x,c}^d(s)^2 ds} + \frac{\int_0^1 J_{x,c}^d(s)dW_x(s)}{\sqrt{\int_0^1 J_{x,c}^d(s)^2 ds}}, \tag{9}$$

with $\rho_{xy} = \text{corr}(\epsilon_{y,t}, \epsilon_{x,t}) = \gamma\sigma_x/\sigma_y \in [-1, 1]$, $J_{x,c}^d(s) = J_{x,c}(s) - \int_0^1 J_{x,c}(r)dr$, $J_{x,c}(s) = \int_0^s \exp((s-r)c)dW_x(r)$ is a standard Ornstein–Uhlenbeck process, and $W_x(s)$ is a standard Brownian motion that is independent of $Z \sim N(0, 1)$.

Proposition 1. *Under Assumptions EPS, BET and RHO, as $T \to \infty$,*

$$t\text{–LS} \to_w U(\rho_{xy}, b/\gamma, c),$$

where $\to_w$ signifies weak convergence.

Remark 1. It is instructive to compare the asymptotic distribution of t–LS with that of the Dickey–Fuller (DF) t-test of the hypothesis of $\rho = 1$ in (2), which is given by

$$U(1, c, c) = DF(c, c) = c\sqrt{\int_0^1 J_{x,c}^d(s)^2 ds} + \frac{\int_0^1 J_{x,c}^d(s)dW_x(s)}{\sqrt{\int_0^1 J_{x,c}^d(s)^2 ds}}.$$

Hence, the asymptotic distribution of t–LS can be seen as a weighted sum of $N(0, 1)$ and a version of the DF local distribution with b/γ taking the place of c. The weight, ρ_{xy}^2, measures the extent of the endogeneity. At the one extreme, if $\rho_{xy}^2 = 1$, so that $\epsilon_{y,t}$ and $\epsilon_{x,t}$ are perfectly correlated, then t–LS essentially becomes a unit root test of y_t, and therefore $U(\rho_{xy}, b/\gamma, c) = DF(b/\gamma, c)$. On the other hand, if $\rho_{xy}^2 = 0$, so that the endogeneity is absent, then $U(\rho_{xy}, b/\gamma, c) = Z$.

Note that while dependent on ρ_{xy}, b, γ and c, the asymptotic distribution of t–LS does not depend on $\sigma_{y,t}^2$. Unattended conditional heteroskedasticity will therefore not interfere with inference, at least not asymptotically. However, there might still be important power gains to be made by accounting for the information contained in the heteroskedasticity. The results reported in Proposition 2 for the QGLS t-statistic enables analysis of this issue. Here $\rho_{xu} = \text{corr}(\epsilon_{x,t}, u_t)$, where $u_t = s_{y,t}^{-2}\epsilon_{y,t}$.

Proposition 2. *Under Assumptions EPS, BET, RHO and HET, as $T \to \infty$,*

$$t\text{–QGLS} \to_w U(\rho_{xu}, b/\gamma, c).$$

From $\rho_{xy} = \gamma\sigma_x/\sigma_y$ we obtain

$$\frac{\partial U(\rho_{xy}, b/\gamma, c)}{\partial b} = \frac{\rho_{xy}}{\gamma}\sqrt{\int_0^1 J_{x,c}^d(s)^2 ds} = \frac{\sigma_x}{\sigma_y}\sqrt{\int_0^1 J_{x,c}^d(s)^2 ds} > 0,$$

suggesting that the power of t–LS is increasing in b, which is also true for t–QGLS. In fact, it is not difficult to show that

$$\frac{\partial U(\rho_{xu}, b/\gamma, c)/\partial b}{\partial U(\rho_{xy}, b, c, \gamma)/\partial b} = \frac{\rho_{xu}}{\rho_{xy}} = \sigma_y\sqrt{\phi_y}, \tag{10}$$

which can be both larger and smaller than one. The relative power of t–LS and t–QGLS is therefore generally indetermine. One exception is when $s_{y,t}^2 = \sigma_{y,t}^2$, such that QGLS is also (true) GLS. In this case, we have that by the Jensen inequality, $\phi_y = E(s_{y,t}^{-2}) = E(\sigma_{y,t}^{-2}) \geq \sigma_y^{-2}$ with equality only if there is no heteroskedasticity. It follows that $\sigma_y\sqrt{\phi_y} \geq 1$, suggesting that t–QGLS should be more powerful than t–LS, which is in agreement with the results of Seo (1999), who considers a maximum likelihood-based unit root test with GARCH. We also see that the power gain is in creasing in the extent of the heteroskedasticity, as measured by ϕ_y/σ_y^2, and that power is equalized when $\sigma_{y,t}^2 = \sigma_y^2$ ($\rho_{xu} = \rho_{xy}$).

Remark 2. As Campbell and Yogo (2006) show, the bias-corrected t-test of Lewellen (2004), t–LEW henceforth, is asymptotically uniformly most powerful when $c = 0$ (and the innovations are conditionally homoskedastic). In view of this, the question naturally arises if the power of this test, which is based on LS, can be increased even further by accounting for the conditional heteroskedasticity? Interestingly, there is an asymptotically equivalent version of t–LEW that arises naturally in our framework if one takes as the objective to eliminate ρ_{xy} from the asymptotic distribution of t–LS. The probably most obvious way to accomplish this is to make (1) conditional on $\epsilon_{x,t}$, leading to the following augmented test regression:

$$y_t = \theta - \gamma\mu(1-\rho) + \beta^* x_{t-1} + \gamma\Delta x_t + \epsilon_{y.x,t},$$

where $\beta^* = \beta - \gamma(\rho - 1)$. t–LEW is asymptotically equivalent to the LS t-statistic for testing $\beta^* = 0$ in the above regression.[3] As Campbell and Yogo (2006, Appendix B) show, the

[3]Specifically, t–LEW is the t-statistic based on $\hat{\beta}^* = \hat{\beta} - \hat{\gamma}(\rho_0 - 1)$, where $\rho_0 = 0.9999$.

asymptotic distribution of this test statistic is given by

$$U^*(\rho_{xy}, b/\gamma, c) = \frac{\rho_{xy}}{\sqrt{1-\rho_{xy}^2}} \frac{(b-\gamma c)}{\gamma} \sqrt{\int_0^1 J_{x,c}^d(s)^2 ds} + Z.$$

The corresponding augmented QGLS t-statistic uses $\sigma_{y.x,t}^{-2}$ as weight and has the following asymptotic distribution: $U^*(\rho_{xy}, b/\gamma, c\rho_{y.xv})$, where $\rho_{y.xv} = \mathrm{corr}(\epsilon_{y.x,t}, v_t) = 1/\sqrt{\phi_{y.x}}\sigma_{y.x}$, $v_t = \sigma_{y.x,t}^{-2}\epsilon_{y.x,t}$ and $\phi_{y.x} = E(\sigma_{y.x,t}^{-2})$ (a formal proof is available upon request). It follows that

$$\frac{\partial U^*(\rho_{xy}, b/\gamma, c\rho_{y.xv})/\partial b}{\partial U^*(\rho_{xy}, b/\gamma, c)/\partial b} = \frac{1}{\rho_{y.xv}} = \sqrt{\phi_{y.x}}\sigma_{y.x} \geq 1,$$

with equality only if $\sigma_{y.x,t}^2 = \sigma_{y.x}^2$. Hence, again QGLS leads to more powerful tests than LS. It is also not difficult to show that this QGLS test should be more powerful than t–QGLS (without augmentation). Unfortunately, while augmentation leads to more power, it also leads to tests that are relatively sensitive to variations in c and γ. Indeed, since c and γ enter $U^*(\rho_{xy}, b/\gamma, c)$ and $U^*(\rho_{xy}, b/\gamma, c\rho_{y.xu})$ in very much the same ways as b, in practice what could be taken for power could just as well be a reflection of size distortion. Because this in the present paper we do not consider augmentation.

Remark 3. The results reported in Remark 2 above are important not only for the tests developed in this study, but also for what they imply for t–LEW and other testst that are similar in construction, such as those of Campbell and Yogo (2006). In particular, not only is the size of these tests expected to be very sensitive to variations in c and γ, but power is also expected to be highly dependent on these parameters. This is easily appreciated by looking at $U^*(\rho_{xy}, b/\gamma, c)$; if γ is "large" and/or $b \approx \gamma c$, then power is expected to be low.

In Proposition 3, we show that the FQGLS t-statistic not only has the same asymptotic distribution as its infeasible QGLS counterpart, but that they are in fact asymptotically equivalent.

Proposition 3. *Under the conditions of Proposition 2,*

$$|t\text{–FQGLS} - t\text{–QGLS}| = o_p(1).$$

Remark 4. In Proof of Proposition 3 we show that $|\hat{s}_{y,t}^2 - s_{y,t}^2| = o_p(1)$. This means that in absence of heteroskedadticity (such that $\sigma_{y,t}^2 = \sigma_y^2$) t–LS and t–FQGLS have the same asymptotic distribution. The asymptotic "price" of accounting for ARCH is therefore zero.

Remark 5. The simulation of critical values for use with t–FQGLS is complicated by the dependence of $U(\rho_{xu}, b/\gamma, c)$ on c and ρ_{xu}^2 (γ is not present under the no predictability null). However, if we assume that $c = 0$, then it is possible to simulate critical values for a few equally spaced values of ρ_{xu}^2. This approach is used by Hansen (1995a), who tabulates critical values for his unit root t-test with regressors for ρ_{xu}^2 from 0.1 to one in steps of 0.1. Interestingly, since the distribution of this test is the same as the one given in Proposition 1, the appropriate 5% critical values for use with t_{FGLS} when testing $\beta = 0$ versus the one-sided alternative that $\beta < 0$ can be taken directly from Hansen (1995a, Table 1).[4] In order to determine the specific critical value to use we need an estimator of ρ_{xu}. In this paper we use $\hat{\rho}_{xu} = \hat{\sigma}_{xu}/\hat{\sigma}_x\hat{\sigma}_u$, where $\hat{\sigma}_x^2 = T^{-1}\sum_{t=2}^T \hat{\epsilon}_{x,t}^2$, $\hat{\sigma}_u^2 = T^{-1}\sum_{t=q+2}^T \hat{u}_t^2$ and $\hat{u}_t = \hat{s}_{y,t}^{-2}\hat{\epsilon}_{y,t}$.

5 Subsampling

The main problem with t–FQGLS is that the asymptotic critical values are invalid if $c \neq 0$. The single most popular approach to accommodate violations of this type is to use a two-step procedure in which the predictor is pretested for a unit root, and where the predictability test is implemented conditional on the outcome of the pretest. Unfortunately, this means loosing control of the overall significance level of the joint test. Therefore, in order to at least put an upper limit on the joint significance level, Campbell and Yogo (2006), Cavanagh et al. (1995), Lewellen (2004), and Torous et al. (2004), among others, have made use of the Bonferroni principle, which states that the significance level for the joint hypothesis that at least one of the tests end up in a rejection is less than or equal to the sum of the individual significance levels. Hence, if the individual unit root and predictability tests are performed at the 5% level, then their joint significance level cannot be larger than 10%.

The advantage of the Bonferroni principle is that it is applicable even if the test statistics are correlated. The main drawback is that the upper limit it provides is a very crude one, and that the test is bound to be conservative. As a response to this, in the present section we consider the subsampling approach, which uses subsamples of length $n < T$ to obtain the empirical distribution function of t–FQGLS. Andrews and Guggenberger (2009a) develop a general subsample theory that applies to a broad class of non-regular models in which most

[4]The corresponding critical values for a two-sided equal-tailed test can be obtained upon request. These are obtained via simulation. In particular, given the asymptotic results reported in Section 2.1, critical values for the case when $c = 0$ can be obtained using simulated random walks of length $T = 1,000$ in place of $W_x(r)$.

other methods fail. As an example, they consider an AR model that is of the same form as (2) where $\{\epsilon_{x,t}\}_{t=1}^T$ is conditionally heteroskedastic and ρ is local-to-unity (see Andrews and Guggenberger, 2009a, Section 7; 2009b, Section S10). The test statistic that they consider is the FQGLS t-statistic of ρ that employs $\hat{s}_{x,t}^2$ as an estimator of $\sigma_{x,t}^2$. The form of the test statistic is therefore the same as that of t–FQGLS.[5] The authors prove that the resulting symmetric two-sided subsampling and equal-tailed two-sided hybrid-subsampling confidence intervals have correct asymptotic size. This is true even if $c \neq 0$, which makes the subsampling approach very appealing also in the present context.

To formally describe the subsampling approach, it is convenient to denote by t_s–FQGLS the FGLS statistic when applied to subsample $s = 1, ..., T-n+1$ covering the n observation pairs $\{(y_s, x_s), ..., (y_{s+n-1}, x_{s+n-1})\}$. For a nominal α-level two-sided symmetric test, the subsample critical value is given by $q_{T,n}(1-\alpha)$, the $(1-\alpha)$-quantile of the empirical distribution of $\{|t_s\text{–FQGLS}|\}_{s=1}^{T-n+1}$. The corresponding upper and lower subsample critical values for a two-sided equal-tailed test are given by $q_{T,n}^*(1-\alpha/2)$ and $q_{T,n}^*(\alpha/2)$, respectively, where $q_{T,n}^*(\alpha)$ is the α-quantile of the empirical distribution of $\{t_s\text{–FQGLS}\}_{s=1}^{T-n+1}$. Define the hybrid upper and lower subsample critical values as $q_{T,n}^{**}(\alpha/2) = \min\{q_{T,n}^*(\alpha/2), q^*(\alpha/2)\}$ and $q_{T,n}^{**}(1-\alpha/2) = \max\{q_{T,n}^*(1-\alpha/2), q^*(1-\alpha/2)\}$, respectively, where $q^*(\alpha)$ is the α-quantile of the asymptotic distribution of t–FQGLS under the null hypothesis, $U(\rho_{xu}, 0, c)$.

Proposition 4. *Under the conditions of Proposition 2, as n, $T \to \infty$ with $n/T \to 0$,*

$$P(|t\text{–FQGLS}| \leq q_{T,n}(1-\alpha)|H_0) \to 1-\alpha,$$
$$P(q_{T,n}^{**}(\alpha/2) \leq t\text{–FQGLS} \leq q_{T,n}^{**}(1-\alpha/2)|H_0) \to 1-\alpha.$$

Remark 6. As far as we, the two-sided symmetric and hybrid equal-tailed FQGLS t-tests are the only predictability tests that have correct asymptotic size when $c \neq 0$. In fact, as Andrews and Guggenberger (2009a) show, subsampling is valid even if $\{(\epsilon_{x.t}, \epsilon_{y.x,t})\}_{t=1}^T$ is stationary and strong mixing (but not necessarily serially uncorrelated). Hence, while the proof makes use of Assumption EPS, this can probably be relaxed to allow for serially correlated disturbances.

[5]The main difference is that while we employ the conventional FQGLS standard error estimator, they employ a sandwich-type heteroskedasticity-consistent (HC) standard error estimator. However, the HC standard errors are only needed in the case when $c = \infty$, which is ruled out by Assumption RHO.

6 Monte Carlo simulations

6.1 DGP

Two experiments, denoted E1 and E2, are considered. In E1 we investigate the performance of the FQGLS tests when the conditional variance is generated as in Example 1 of Section 2, which means that the fitted conditional variance model nests the true model, suggesting that FQGLS should be asymptotically efficient. The purpose of E2 is to evaluate the performance when the true model is not nested in this way.

The DGP used in E1 is given by a restricted version of (1) and (2) where the conditional variance generated as described in Example 1 of Section 2. The restrictions are $\theta = 1$, $\mu = 0$, $q_x = q_{y.x} = 1$, $\alpha_{0,y.x} = \alpha_{0,x} = \alpha_0$, $\alpha_{1,y.x} = \alpha_{1,x} = \alpha_1$, $\alpha_0 = 1 - \alpha_1$ (to ensure that $\sigma^2_{y.x} = \sigma^2_x = 1$) and $\varepsilon_t \sim N(0, \Theta_t)$. By setting α_1 and γ, we can control the values taken by ρ_{xy} and ρ_{xu}. Note first that by the definition of ρ_{xy}, since $\sigma^2_{y.x} = \sigma^2_x = 1$, we have $\sigma^2_y = 1 + \gamma^2$ and therefore $\rho_{xy} = \gamma/\sqrt{\gamma^2 + 1}$. There is no such explicit formula for ρ_{xu}. However, since $\sigma_x = 1$, we know that $\rho_{xu} = \gamma\sqrt{\phi_y}$, suggesting an inverse relationship with α_1. Also, by setting c we can control the persistency of $\{x_t\}_{t=1}^T$.

The DGP used in E2 is again given by (1) and (2) where $\theta = 1$ and $\mu = 0$, but now we fix $c = -20$ and $\gamma = -0.5$, which allows us to focus on $\sigma^2_{y.x,t}$ and σ^2_{xt}. In particular, while the latter variance is as in E1, for the former four specifications are considered.

- GARCH: $\sigma^2_{y.x,t} = 1 + 0.4\varepsilon^2_{y.x,t-1} + 0.4\sigma^2_{y.x,t-1}$.
- X–ARCH: $\sigma^2_{y.x,t} = 1 + 0.4\varepsilon^2_{y.x,t-1} + 0.4\varepsilon^2_{x,t-1}$.
- Asymmetric threshold ARCH: $\sigma^2_{y.x,t} = 1 + 0.2\varepsilon^2_{y.x,t-1}1(\varepsilon_{y.x,t-1} > 0)$ $+ 0.6\varepsilon^2_{y.x,t-1}1(\varepsilon_{y.x,t-1} \leq 0)$.
- AR–SV: $\ln(\sigma^2_{y.x,t}) = 1 + 0.4\ln(\sigma^2_{y.x,t-1}) + \eta_t$, where $\eta_t \sim N(0, 1)$.

All results are based on $3,000$ replications of samples of size $T \in \{50, 100, 200\}$. The reason for considering only relatively small values of T is to demonstrate how the subsample FGLS test does not require T very large to work properly. In fact, according to the results, $T = 50$ is enough. However, we also considered some relatively large sample sizes, the results of which are described in Section 6.4.

6.2 Implementation issues

In interest of comparison, in addition to symmetric and hybrid subsample FQGLS, t–LS and t–LEW are also simulated.[6] Thus, all-in-all we have four tests.

As for FGLS, even if GARCH, X–ARCH and asymmetric threshold ARCH models are actually estimable in our approach (see Examples 2 and 3 in Section 3), the implementation is done under the ARCH specification in (5) and (6), which is correct in E1 but not in E2. Hence, in E2 we study the performance of FQGLS when the conditional variance is misspecified. In both experiments, we need to determine the appropriate ARCH orders, q_x and $q_{y.x}$. We experimented with several rules, but opted for the Schwarz Bayesian information criterion (BIC); see Section 6.4 for a discussion of the performance under some of the other rules. The maximum number of lags is set to $\lfloor 4(T/100)^{2/9} \rfloor$, where $\lfloor x \rfloor$ denoted the integer part of x (Ng and Perron, 1995). The subsample critical values are based on the full sample lag estimates, although $\sigma^2_{y.x,t}$ and σ^2_{xt} are reestimated for each subsample.[7]

As for the subsample length, n, from a theoretical point of view any choice satisfying $n \to \infty$ with $n/T \to 0$ as $T \to \infty$ will do. Applying the procedure in practice, however, one has to be more careful. Choosing n "too small" may lead to a bad approximation of the actual distribution, because if the sample is small enough t–FQGLS may not be well-behaved. On the other hand, choosing n too close to T will lead to very little variation in $\{t_s\text{–FQGLS}\}_{s=1}^{T-n+1}$, and therefore an underestimation of the true dispersion of t–FQGLS. Because of these concerns, we follow Romano and Wolf (2001, Algorithm 5.1) and pick n by minimizing the variance of the critical values. The algorithm proceeds as follows:

1. For each $n \in [n_{min}, n_{max}]$ compute a the subsample critical values. Romano and Wolf (2001) recommend setting $n_{min} = r_{min}\sqrt{T}$ and $n_{max} = r_{max}\sqrt{T}$ with $r_{min} \in [0.5, 1]$ and $r_{max} \in [2, 3]$. We set $r_{min} = 1$ and $r_{max} = 2$.

2. For each n, compute $S(n)$ as the standard deviation of the subsample critical values obtained in $[n-k, n+k]$, where $k > 0$. Following Romano and Wolf (2001), we set

[6] Results were obtained also for the asymptotic and Bonferroni t-tests, Bonferroni Q-test and sup-bound Q-test of Campbell and Yogo (2006). However, in terms of size accuracy these were dominated by the other tests. Therefore, in order to keep our tables uncluttered, we do not report the results of the Campbell and Yogo (2006) tests, which can be made available upon request. Also available upon request are the results for asymptotic FQGLS.

[7] The full-sample estimators of $\sigma^2_{y.x,t}$ and σ^2_{xt} can also be used. The justification for this is that FQGLS is asymptotically equivalent in the full sample and in subsamples. However, unreported results suggest that the test based on the subsample estimator works best.

$k = 2$.

3. Set $\hat{n} = \arg\min_{n \in [n_{min}, n_{max}]} S(n)$.

It is important in practice that the estimated variance is bounded away from zero. One way to ensure this is to follow Hansen (1995b) and to use a trimmed estimator such as $\max\{\kappa\hat{\sigma}_y^2, \hat{\sigma}_{y,t}^2\}$, where $\kappa \in [0,1]$. He sets $\kappa = 0.1$, which is an arbitrary choice. A better approach that is asymptotically justifiable involves setting $\kappa \sim T^{-\eta}$ with $\eta > 0$, which, together with the consistency of $\hat{\sigma}_{y,t}^2$ (see Appendix), implies that $P(\hat{\sigma}_{y,t}^2 < \kappa\sigma_y^2) \to 0$ as $T \to \infty$. Hence, under these conditions, truncation is asymptotically irrelevant. In the simulations, we set $\kappa = T^{-1/2}$, although in the robustness trials we ran the results for $\kappa = 0$, 0.1 were almost identical (see Section 6.4).

All tests are double-sided. Therefore, since critical values reported by Hansen (1995a) are for a one-sided test, the asymptotic critical values for use with t–LS and hybrid subsample FQGLS had to be obtained via simulation (see Remark 5). t-LEW has a $N(0,1)$ distribution. The appropriate 5% level critical value is therefore given by 1.96.

6.3 Results

We begin by considering the results from E1, which are reported in Table 1. Because of space constraints, we focus on the size and size-adjusted power of a nominal 5% level test.[8] The information content of this table may be summarized as follows:

- With $b = c = \gamma = 0$, all four tests perform quite similarly with size being close to the nominal 5% level, which is just as expected. However, we also see that whenever $c, \gamma \neq 0$, t–LS and t–LEW have difficulties. The latter test is particularly distorted, which is partly expected given Remark 3. By contrast, the subsample FQGLS tests seem to perform well with only minor size distortions.

- The power of the tests is rather stable across the three values of T considered, which is to be expected given the local specification of β. Thus, as alluded to in Section 3, while a fixed value $\beta \neq 0$ would cause power to increase with T, since now $\beta \to 0$ as $T \to \infty$, power should be flat in T.

[8]Some results on the estimated lag lengths and root mean squared error of the different estimators are available upon request.

- If, on the one hand, $c = \gamma = 0$, then all tests tend to perform very similarly in terms of power. The subsample tests generally perform best; however, the difference in power is very small. The results for $c, \gamma \neq 0$ are, on the other hand, quite different with the power of the subsample FQGLS tests being well above that of the other tests. In fact, the subsample FQGLS tests are uniformly more powerful than the other tests.

- The fact that power is decreasing in both $|\gamma|$ and $|c|$ is in accordance with our theoretical results (see Section 4), and so is the relative sensitivity of t–LEW (see Remark 3). In fact, based on the results reported here, given its apparent inability to discriminate between the null and alternative hypotheses, t–LEW should not be used unless c and γ are known to be zero.

- The overall best power is obtained by using the symmetric subsample FQGLS test, which is in agreement with the Monte Carlo results reported by Andrews and Guggenberger (2009a, Table II) for their subsample FQGLS unit root tests.

Consider next the results reported in Table 2 for E2. These can be summarized as follows.

- Given their insensitivity to a misspecification of the conditional heteroskedasticity, the fact that the size of the subsample FQGLS tests is not affected by the variance model is well expected.

- While power vary quite substantially across variance models, it does so for all tests. The relative ranking of the tests is therefore the same as in Table 1 with the subsample FQGLS tests leading to highest power. The poorest performance is again obtained by using t–LEW, which has no power beyond size.

In sum, the results reported in this section suggests that the subsample FQGLS tests have good size accuracy and high power relative to other tests. This is true regardless of whether the conditional variance is correctly specified or not. The new tests should therefore be a valuable addition to the already existing menu of predictability tests.

6.4 Robustness

The robustness of the above results was assessed along several dimensions.

- As usual, the choice of lag orders in (5) and (6) should be done with care, as test results can be quite sensitive to different lag lengths. Specifically, while underspecification can lead to size distortions, overspecification will lead to a loss of power. Although this is of course a general concern, it is particularly relevant in the current context, as there is not just one choice that has to be made but two (q_x and $q_{y.x}$). Interestingly, the results were very robust across the various rules considered. For example, the results based on the Akaike information criterion (AIC) were almost identical to the results based on the BIC. The same was true when considering a general-to-specific test rule based on the t-statistic of the last ordered lag. Hence, when choosing among the "usual suspects", the choice of lag length selection rule does not seem to be that important after all.

- The subsample FQGLS tests are centered by the true parameter value under the null hypothesis, which is typically referred to as "uncentered" subsampling. There is also "centered" subsampling, in which case the test statistic is centered by the full-sample estimate of that parameter. Unreported results show that centering (by the full-sample estimate) causes the subsample FQGLS tests to become oversized, which is partly expected given the results of McMurry et al. (2013, Theorem 3.5), showing that centered rather than uncentered subsampling leads to relatively more rejections.

- The above results are based on setting κ (in the trimming of $\hat{s}^2_{y,t}$) to $T^{-1/2}$. We carried out the same simulations for $\kappa = 0$, 0.1, and obtained very similar results. This is because the trimming almost never took effect. However, we still recommend trimming, as a single variance estimate close to zero is enough cause misleading inference. We also ran some simulations based on trimming $\hat{\lambda}_x$ and $\hat{\lambda}_{y.x}$ rather than $\hat{s}^2_{y,t}$, an idea that was motivated by the work of Francq and Zakoïan (2008). However, trimming $\hat{s}^2_{y,t}$ almost always led to better performance.

- In the empirical part $T > 1,500$. Therefore, in order to test the robustness of our findings with respect to sample size, we did some trial runs with $T = 1,500$. The results from these runs were very similar to the ones for $T = 200$, which is partly expected given how little variation there is between $T = 50$ and $T = 200$. Note especially how the subsample FGLS tests seem to have good size accuracy already when $T = 50$. Hence, T does not have to be very large for the asymptotic approximation to provide a reasonable description of actual test behavior.

7 Empirical illustration

The return predictability literature can be divided into two strands depending on the type of predictor used; while the first uses various financial ratios (see, for example, Westerlund and Narayan, 2012; Rapach et al., 2010; Welch and Goyal, 2008; Campbell and Thompson, 2008; Campbell and Yogo, 2006; Lewellen, 2004), the second uses macroeconomic variables (see, for example, Rapach et al., 2005; Narayan and Sharma, 2011; Diesprong et al., 2008).

While obviously similar in aim, the two strands have met with varying interest. The financial ratios have been the most popular by far and therefore this strand has accumulated a rich volume of studies, giving it both depth and scope, even from an econometric point of view. In particular, the importance of accounting for the persistency, endogeneity, and to some extent also the heteroskedasticity of the data is well-known (see, for example, Campbell and Yogo, 2006; Lewellen, 2004; Westerlund and Narayan, 2012b). Unfortunately, the same cannot be said about studies that use macroeconomic predictors. In fact, unlike the financial ratios, which are generally believed to be able to predict returns, there is still lack of consensus as to the predictive content of macroeconomic variables. As a way of "compensating" for this lack of consensus, in the present paper we investigate the predictive ability of the oil price and the exchange rate, two of the most common macroeconomic predictors. However, we do not ignore financial ratios. One of the implications of the results of Section 4–6 is that the existing evidence that financial ratios predict returns based on the tests of Lewellen (2004), and Campbell and Yogo (2006) may well be due to size distortions. Hence, there is a need to reconsider the empirical evidence based on financial ratios.

7.1 Data

We have two data sets. The first data set is at a weekly frequency and contains returns of the S&P500 index, the US dollar-British pound exchange rate (ER), and the oil price (OIL), as proxied by the US crude oil measured in dollars per barrel. The span of the sample was dictated by the availability of data for the predictors. The exchange rate data series is longest and spans the 01/01/1971–5/25/2012 period, culminating into no less than 2,161 weekly observations. The oil price data cover the 5/13/1983–5/25/2012 period, and comprise 1,516 observations. All data are obtained from BLOOMBERG.

Our second data set is obtained from Amit Goyal's web page and is the same as in Goyal and Welch (2008). The published data set ends in 2006, and has since then been updated to include two more years. A complete description of the data is provided in Goyal and Welch (2008), and in an appendix available from Goyal's web page. The data set consists of monthly returns on the S&P500 index and 12 predictor variables including book-to-market (BM) ratio, long-term bond yield (LTY), earnings-price ratio (EP), long-term bond return (LTR), inflation (INFL), stock variance (SVAR), net equity expansion (NTIS), term spread (TMS), corporate bond rate (CORPR), default yield spread (DFY), default return spread (DFR), and dividend-price ratio (DP). The sample period for the variable varies depending on data availability, and is given in Table 3.

7.2 Descriptive statistics

We begin by looking at Table 3, which reports some commonly used descriptive statistics. We pay particular attention to the extent of persistency and heteroskedasticity. However, before we do this we discuss briefly the reported sample moments. We begin by considering the macroeconomic predictors. OIL is at least five times as volatile as ER. Both predictors are skewed; OIL is right-skewed, and ER left-skewed. Moreover, while ER is roughly mesokurtic, OIL is leptokurtic with relatively thin tails. The macroeconomic predictors therefore behave quite differently, which is suggestive of differing predictive abilities. As for the financial ratios, we see that DP is most volatile, followed by LTR and CORPR. We also see that most financial ratios tend to be skewed and the kurtosis measure is larger than three for all predictors, which is suggestive of fat tails.

We now turn to the persistency and heteroskedasticity of the data, which have more direct implications for the FQGLS tests. We begin with heteroskedasticity. As a first step we filtered each of the variables using an AR model (of order six). A Lagrange multiplier test for the null hypothesis of no ARCH is then applied to the resulting filtered series. The results reported in Table 3 suggest that the no ARCH null can be rejected at the 1% significance level for all variables.

The finding that all variables seem to exhibit ARCH is in itself of course nothing new, but is actually a rather expected result. However, it highlights an often ignored source of information in studies of return predictability. There are studies that recognize the presence of heteroskedasticity, although typically little or nothing is done about it. To give one example, in their study of the predictability of the equity premium Welch and Goyal (2008, page 1474) state that "There are other diagnostics (heteroskedasticity, residual autocorrelation, etc.) that stable models should pass but we do not explore them in our article." Other studies, such as Kothari and Shanken (1997), and Ferson et al. (2003), use robust standard errors, which means that although the heteroskedasticity is not ignored, the information it contains is. The return predictability literature generally suffers from these same issues.

Next, we attempt to ascertain the order of integration of the variables. The ADF test results reported in Table 3 suggest that the null hypothesis of a unit root has to be rejected at the 10% level for all variables but returns (in both samples), BM and LTY. We therefore conclude that there is evidence of stationarity not only for returns but also for a majority of the predictors. However, in many cases the estimated AR coefficient is still very close to one, suggesting that while statistically different from one, the actual difference is not very large. Most of the predictors therefore exhibit unit root-like behavior, which leads us to conclude that Assumption RHO is appropriate.[9] Another implication of this result is that both t–LS and t-LEW are likely to be size distorted, and that the subsample FQGLS tests are more reliable. In the next section we therefore only consider these tests.

7.3 Predictability test results

Having considered briefly the properties of each variable in our two samples, we now turn to the FQGLS test results. But before we come to the actual test results, in Table 4 we take a look at some regression diagnostics that are obtained as a by-product when implementing the FQGLS tests. In particular, we report the results from a Wald test of the null of no ARCH effect in (5) and (6). Corroborating the previously reported evidence of ARCH, we see that the null of no ARCH is rejected in both models for all variables. The only exception is for (5) in case of OIL.

In addition to the ARCH test results, Table 4 also reports the results of the LS estimator $\hat{\gamma}$ of γ in (3). If the predictors are exogenous, $\gamma = 0$, whereas is they are endogenous, $\gamma \neq 0$. The results show that $\hat{\gamma}$ is significantly different from zero for all but three predictors, INFL, NTIS and TMS. The fact that INFL turns up insignificant is in agreement with the results of Valkanov (2003). BM, EP and DP are among the predictors where a relatively high value of $\hat{\gamma}$ is observed. Coincidentally, these are also the predictors that have received most attention

[9] This finding regarding financial ratio predictors seems largely in agreement with studies such as Pontiff and Schall (1998), Lewellen (2004), Campbell and Yogo (2006), and Stambaugh (1999), in which the predictors are found to be highly persistent.

in the literature.[10]

The results from the predictability tests are presented in Table 5. In particular, we focus on a 95% confidence interval for β using the subsample FQGLS tests. We begin by looking at the macroeconomic predictors. The confidence intervals for OIL include the value zero, which means that there is no evidence against the no predictability null. By contrast, for ER the confidence intervals lie to the left of zero, suggesting that the no predictability null should be rejected. The results for the financial ratios are mixed too. Over the full sample period, we count five rejections, for EP, SVAR, NTIS, DFY and DP. One of the most popular predictors of returns, BM, shows limited evidence of predictability. The post-war data are slightly more supportive of predictability. In this case, we find evidence of predictability for seven predictors, including BM.

How do the above results compare to those reported in the extant literature? The cross-sectional return predictability literature celebrates BM as a very successful predictor. We fail to find strong support that BM predicts returns. In addition, from the work of Lewellen (2004) we learn that EP is a weak predictor of returns. This is in sharp contrast to our results; we find that EP strongly predicts returns. Similarly, while Hjalmarsson (2010) do not find any evidence that EP and DP predict returns, according to our findings both variables are in fact able to predict returns. Hence, different tests lead to different conclusions, suggesting that the choice of test is key when trying to assess the reliability of the evidence on the predictability hypothesis.

As already explained, the choice of which test to use is dictated primarily by the extent of predictor persistency and endogeneity. We show that our samples, comprising not only a large number of financial ratios but also two of the most popular macroeconomic predictors, are characterized by both features. This means that existing tests, such as t-LS and t–LEW are likely to be misleading, and that subsample FQGLS is more reliable. We also show that accounting for heteroskedasticity is important for power, and this feature is present in both samples. The main implication of these findings, one that can simply be drawn from the descriptive statistics, is that subsample FQGLS should be most suitable.

[10]The preference for these predictors is mainly due to their importance for financial models of risk and return (see Campbell and Shiller, 1988; Cochrane, 1991).

8 Concluding remarks

The literature on return predictability has evolved considerably over the last 20 years. While initial tests produced strong evidence of predictability, subsequent research has shown that much of this evidence is in fact due to bias caused by inappropriate assumptions regarding the predictors. For example, while initially thought to be stationary, many popular predictors have since then been shown to be highly persistent and often one cannot reject the notion that they are in fact unit root non-stationary. It is therefore natural to model these predictors as having a root near unity. Similarly, while the predictors were initially modeled as being exogenous, often this is not what one finds in practice.

Recent studies have therefore moved away from the assumption of an exogenous stationary predictor and towards more realistic models of endogeneity and near unit roots. However, while certainly more general when it comes to the allowable predictors, these studies ignore the fact that returns are typically characterized by conditional heteroskedasticity. In the current study we show, both analytically and numerically, that this practice can be rather costly, and that the inclusion of the extraneous information contained in the heteroskedasticity can lead to substantial gains in performance, not only in terms of power but also in terms of size accuracy when combined with subsample critical values. This is verified in small samples using both simulated and real data.

References

Andrews, D. W. K., and P. Guggenberger (2009a). Hybrid and Size-Corrected Subsampling Methods. *Econometrica* **77**, 721–762.

Andrews, D. W. K., and P. Guggenberger (2009b). Supplement to "Hybrid and Size-Corrected Subsampling Methods". (*Econometrica* **77**, 721–762). Available on the Econometric Society website.

Andrews, D. W. K., and P. Guggenberger (2012). Asymptotics for LS, GLS, and Feasible GLS Statistics in an AR(1) Model with Conditional Heteroskedasticity. *Journal of Econometrics* **169**, 196–210.

Bollerslev, T. (1990). Modeling Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach. *Review of Economics and Statistics* **72**, 498–505.

Brenner, R. J., R. H. Harjes and K. F. Kroner (1996). Another Look at Models of the Short-Term Interest Rate. *Journal of Financial and Quantitative Analysis* **31**, 85–107.

Cavanagh, C., G. Elliott and J. Stock (1995). Inference in Models with Nearly Integrated Regressors. *Econometric Theory* **11**, 1131–1147.

Campbell, J. Y., and S. B. Thompson (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average. *Review of Financial Studies* **21**, 1509–1531.

Campbell, J. Y., and M. Yogo (2006). Efficient Tests of Stock Return Predictability. *Journal of Financial Economics* **81**, 27–60.

Diesprong, G., B. Jacobsen and B. Maat (2008). Striking Oil: Another Puzzle? *Journal of Financial Economics* **89**, 307–327

Elliott, G., and J. H. Stock (1994). Inference in Time Series Regression When the Order of Integration of a Regressor is Unknown. *Econometric Theory* **10**, 672–700.

Ferson, W., S. Sarkissian and T. Simin (2003). Spurious Regression in Financial Economics? *Journal of Finance* **58**, 1393–1413.

Francq, C., and J.-M. Zakoïan (2008). Estimating ARCH Models when the Coefficients are Allowed to be Equal to Zero. *Austrian Journal of Statistics* **37**, 31–40.

Glosten, L. R., R. Jagannathan and D. Runkle (1993). On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance* **48**, 1779–1801.

Hansen, B. E. (1995a). Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power. *Econometric Theory* **11**, 1148–1171.

Hansen, B. E. (1995b). Regression with Nonstationary Volatility. *Econometrica* **63**, 1113–1132.

Kothari, S. P., and J. Shanken (1997). Book-to-market, dividend yield, and expected market returns: A Time Series Analysis. *Journal of Financial Economics* **44**, 169–203.

Lanne, M. (2002). Testing the Predictability of Stock Returns. *The Review of Economics and Statistics* **84**, 407–415.

Lewellen, J. (2004). Predicting Returns with Financial Ratios. *Journal of Financial Economics* **74**, 209–235.

McMurry, T., D. N. Politis and J. P. Romano (2013). Subsampling Inference with K Populations and a Nonstandard Behrens-Fisher Problem. Forthcoming in *International Statistical Review*.

Narayan, P. K., and S. Sharma (2011). New Evidence on Oil Price and Firm Returns. *Journal of Banking and Finance* **35**, 3253–3262.

Ng, S., and P. Perron (1995). Unit Root Test in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag. *Journal of the American Statistical Association* **90**, 268–281.

Rapach, D. E., J. K. Strauss and G. Zhou (2010). Out-of-sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *Review of Financial Studies* **23**, 821–862.

Rapach, D. E., J. K. Strauss and J. Rangvid (2005). Macro Variables and International Stock Return Predictability. *International Journal of Forecasting* **21**, 137–166.

Romano, J. P., and M. Wolf (2001). Subsampling Intervals in Autoregressive Models with Linear Time Trend. *Econometrica* **69**, 1283–1314.

Said, E. D., and D. A. Dickey (1984). Testing for Unit Roots in Autoregressive Moving Average Models of Unknown Order. *Biometrika* **71**, 599–607.

Seo, B. (1999). Distribution Theory for Unit Root Tests with Conditional Heteroskedasticity. *Journal of Econometrics* **91**, 113–144.

Smeekes, S., and R. A. M. Taylor (2012). Bootstrap Union Tests For Unit Roots In The Presence Of Nonstationary Volatility. *Econometric Theory* **28**, 422–456.

Stambaugh, R. F. (1999). Predictive Regressions. *Journal of Financial Economics* **54**, 375–421.

Torous, W., R. Valkanov and S. Yan (2004). On Predicting Stock Returns with Nearly Integrated Explanatory Variables. *Journal of Business* **77**, 937–966.

Valkanov, R. (2003). Long-horizon Regressions: Theoretical Results and Applications. *Journal of Financial Economics* **68**, 201–232.

Welch, I., and A. Goyal (2008). A Comprehensive Look at the Empirical Performance of Equity Premium Prediction. *Review of Financial Studies* **21**, 1455–1508.

Westerlund, J., and P. Narayan (2012). Does the Choice of Estimator Matter when Forecasting Returns? *Journal of Banking and Finance* **36**, 2632–2640.

Appendix: Proofs

Proof of Proposition 1.

The proof of Proposition 1 is a direct consequence of the results of Elliott and Stock (1994, Section 2.2). Details are therefore omitted. ■

Lemma A.1. *Under Assumption EPS and HET, as $T \to \infty$,*

$$\frac{1}{\sqrt{T}} \sum_{n=1}^{\lfloor sT \rfloor} \begin{bmatrix} \epsilon_{x,n} \\ \epsilon_{y,n} \\ u_n \end{bmatrix} \to_w \begin{bmatrix} B_x(s) \\ B_y(s) \\ B_u(s) \end{bmatrix},$$

where

$$\begin{aligned} B_x(s) &= \sigma_x W_x(s), \\ B_y(s) &= \sigma_y \rho_{xy} W_x(s) + \sigma_y (1-\rho_{xy}^2)^{1/2} W_y(s), \\ B_u(s) &= \sqrt{\phi_y} \left[\frac{1-\rho_{xu}^2}{1-\rho_{xy}^2} \left(1 - \frac{\rho_{xy}^2}{\rho_{xu}^2} \right) \right]^{1/2} W_u(s) + \sqrt{\phi_y} \rho_{xu} W_x(s) \\ &+ \sqrt{\phi_y} \frac{\rho_{xy}(1-\rho_{xu}^2)}{\rho_{xu}(1-\rho_{xy}^2)^{1/2}} W_y(s), \end{aligned}$$

with $u_t = s_{y,t}^{-2} \epsilon_{y,t}$, and $W_u(s)$, $W_y(s)$ and $W_x(s)$ being independent Brownian motions.

Proof of Lemma A.1.

By Lemma 5 (e) in Andrews and Guggenberger (2012),

$$\frac{1}{\sqrt{T}} \sum_{n=1}^{\lfloor sT \rfloor} \begin{bmatrix} \epsilon_{x,n} \\ \epsilon_{y,n} \\ u_n \end{bmatrix} \to_w \begin{bmatrix} B_x(s) \\ B_y(s) \\ B_u(s) \end{bmatrix} = \Sigma^{1/2} \begin{bmatrix} W_x(s) \\ W_y(s) \\ W_u(s) \end{bmatrix} = \Sigma^{1/2} \begin{bmatrix} W_\epsilon(s) \\ W_u(s) \end{bmatrix}, \tag{A1}$$

as $T \to \infty$, where $W_u(s)$, $W_y(s)$ and $W_x(s)$ are three Brownian motions that are independent of each other, $W_\epsilon(s) = [W_x(s), W_y(s)]'$ and

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xu} \\ \sigma_{xy} & \sigma_y^2 & \sigma_{yu} \\ \sigma_{xu} & \sigma_{yu} & \sigma_u^2 \end{bmatrix} = \begin{bmatrix} \Sigma_\epsilon & \Sigma_{\epsilon u} \\ \Sigma_{\epsilon u}' & \sigma_u^2 \end{bmatrix},$$

with an obvious definition of Σ_ϵ and $\Sigma_{\epsilon u}$. As for the elements of Σ, by the law of iterated expectations,

$$\begin{aligned}
\sigma_{xy} &= E(\epsilon_{x,t}\epsilon_{y,t}) = E[\epsilon_{x,t}(\gamma\epsilon_{x,t} + \epsilon_{y.x,t})] = \gamma E(\epsilon_{x,t}^2) = \gamma\sigma_x^2, && \text{(A2)}\\
\sigma_u^2 &= E(u_t^2) = E(s_{y,t}^{-4}\epsilon_{y,t}^2) = E[s_{y,t}^{-4}E(\epsilon_{y,t}^2|F_{t-1})] = E(s_{y,t}^{-2}) = \phi_y, && \text{(A3)}\\
\sigma_{yu} &= E(\epsilon_{y,t}u_t) = E(s_{y,t}^{-2}\epsilon_{y,t}^2) = E[s_{y,t}^{-2}E(\epsilon_{y,t}^2|F_{t-1})] = 1, && \text{(A4)}\\
\sigma_{xu} &= E(\epsilon_{x,t}u_t) = E(s_{y,t}^{-2}\epsilon_{y,t}\epsilon_{x,t}) = E[s_{y,t}^{-2}E(\epsilon_{y,t}\epsilon_{x,t}|F_{t-1})]\\
&= E[s_{y,t}^{-2}E((\gamma\epsilon_{x,t} + \epsilon_{y.x,t})\epsilon_{x,t}|F_{t-1})] = \gamma E[s_{y,t}^{-2}E(\epsilon_{x,t}^2|F_{t-1})] = \gamma\sigma_x^2\phi_y. && \text{(A5)}
\end{aligned}$$

Write $\Sigma = \Sigma^{1/2}(\Sigma^{1/2})'$. The Cholesky decomposition of $\Sigma^{1/2}$ is given by

$$\Sigma^{1/2} = \begin{bmatrix} \Sigma_\epsilon^{1/2} & 0 \\ \Sigma_{\epsilon u}'(\Sigma_\epsilon^{-1/2})' & \sigma_{u.\epsilon} \end{bmatrix}, \tag{A6}$$

where $\sigma_{u.\epsilon}^2 = \sigma_u^2 - \Sigma_{\epsilon u}'\Sigma_\epsilon^{-1}\Sigma_{\epsilon u}$, suggesting

$$\begin{aligned}
\begin{bmatrix} B_x(s) \\ B_y(s) \\ B_u(s) \end{bmatrix} &= \begin{bmatrix} \Sigma_\epsilon^{1/2} & 0 \\ \Sigma_{\epsilon u}'(\Sigma_\epsilon^{-1/2})' & \sigma_{u.\epsilon} \end{bmatrix} \begin{bmatrix} W_\epsilon(s) \\ W_u(s) \end{bmatrix}\\
&= \begin{bmatrix} \Sigma_\epsilon^{1/2}W_\epsilon(s) \\ \sigma_{u.\epsilon}W_u(s) + \Sigma_{\epsilon u}'(\Sigma_\epsilon^{-1/2})'W_\epsilon(s) \end{bmatrix}. && \text{(A7)}
\end{aligned}$$

By using $\Sigma_{\epsilon u} = (\sigma_{xu}, 1)' = (\gamma\phi_y\sigma_x^2, 1)'$ and

$$\Sigma_\epsilon^{-1} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}^{-1} = \begin{bmatrix} \sigma_x^2 & \gamma\sigma_x^2 \\ \gamma\sigma_x^2 & \sigma_y^2 \end{bmatrix}^{-1} = \frac{1}{\sigma_x^2(\sigma_y^2 - \gamma^2\sigma_x^2)} \begin{bmatrix} \sigma_y^2 & -\gamma\sigma_x^2 \\ -\gamma\sigma_x^2 & \sigma_x^2 \end{bmatrix}$$

we obtain

$$\begin{aligned}
\sigma_{u.\epsilon}^2 &= \sigma_u^2 - \Sigma_{\epsilon u}'\Sigma_\epsilon^{-1}\Sigma_{\epsilon u}\\
&= \phi_y - \frac{1}{\sigma_y^2 - \gamma^2\sigma_x^2}\left(\frac{\rho_{xu}^2}{\rho_{xy}^2} - 1\right)(\rho_{xu}^2 - 1)\\
&= \phi_y\left[1 - \frac{\rho_{xy}^2}{\rho_{xu}^2(1-\rho_{xy}^2)}\left(\frac{\rho_{xu}^2}{\rho_{xy}^2} - 1\right)(\rho_{xu}^2 - 1)\right] = \phi_y\frac{1-\rho_{xu}^2}{1-\rho_{xy}^2}\left(1 - \frac{\rho_{xy}^2}{\rho_{xu}^2}\right), && \text{(A8)}
\end{aligned}$$

where $\rho_{xu} = \sigma_{xu}/\sigma_u\sigma_x = \gamma\sqrt{\phi_y}\sigma_x$ and $\rho_{xy} = \sigma_{xy}/\sigma_x\sigma_y = \gamma\sigma_x/\sigma_y$.

Let us now consider $\Sigma_\epsilon^{1/2}$. By another use of the Cholesky decomposition,

$$\Sigma_\epsilon^{1/2} = \begin{bmatrix} \sigma_x & 0 \\ \gamma\sigma_x & \sigma_{y.x} \end{bmatrix} = \begin{bmatrix} \sigma_x & 0 \\ \sigma_y\rho_{xy} & \sigma_y(1-\rho_{xy}^2)^{1/2} \end{bmatrix}, \tag{A9}$$

where $\sigma^2_{y.x} = \sigma^2_y - \gamma^2\sigma^2_x = \sigma^2_y(1-\rho^2_{xy})$, suggesting

$$\Sigma_\epsilon^{1/2}W_\epsilon(s) = \begin{bmatrix} \sigma_x W_x(s) \\ \sigma_y\rho_{xy}W_x(s) + \sigma_y(1-\rho^2_{xy})^{1/2}W_y(s) \end{bmatrix}, \tag{A10}$$

and

$$\begin{aligned}\Sigma'_{\epsilon u}(\Sigma_\epsilon^{-1/2})'W_\epsilon(s) &= \gamma\phi_y\sigma_x W_x(s) + \frac{(1-\rho^2_{xu})}{\sigma_{y.x}}W_y(s) \\ &= \sqrt{\phi_y}\rho_{xu}W_x(s) + \sqrt{\phi_y}\frac{\rho_{xy}(1-\rho^2_{xu})}{\rho_{xu}(1-\rho^2_{xy})^{1/2}}W_y(s).\end{aligned} \tag{A11}$$

Direct substitution now yields

$$\begin{aligned}B_x(s) &= \sigma_x W_x(s), \\ B_y(s) &= \sigma_y\rho_{xy}W_x(s) + \sigma_y(1-\rho^2_{xy})^{1/2}W_y(s), \\ B_u(s) &= \sqrt{\phi_y}\left[\frac{1-\rho^2_{xu}}{1-\rho^2_{xy}}\left(1-\frac{\rho^2_{xy}}{\rho^2_{xu}}\right)\right]^{1/2} W_u(s) + \sqrt{\phi_y}\rho_{xu}W_x(s) \\ &+ \sqrt{\phi_y}\frac{\rho_{xy}(1-\rho^2_{xu})}{\rho_{xu}(1-\rho^2_{xy})^{1/2}}W_y(s),\end{aligned}$$

as was to be shown. ■

Proof of Proposition 1.

t–QGLS can be expanded as

$$\begin{aligned}t\text{–QGLS} &= \frac{\sum_{t=2}^T s_{y,t}^{-2}x^d_{t-1}y_t}{\sqrt{\sum_{t=2}^T s_{y,t}^{-2}(x^d_{t-1})^2}} \\ &= T\beta\left(T^{-2}\sum_{t=2}^T s_{y,t}^{-2}(x^d_{t-1})^2\right)^{1/2} + \frac{T^{-1}\sum_{t=2}^T s_{y,t}^{-2}x^d_{t-1}\epsilon^d_{y,t}}{\sqrt{T^{-2}\sum_{t=2}^T s_{y,t}^{-2}(x^d_{t-1})^2}}.\end{aligned} \tag{A12}$$

By Lemma 4 (c) Andrews and Guggenberger (2012), $T^{-3/2}\sum_{t=2}^T(s_{y,t}^{-2} - \phi_y)x^d_{t-1} = O_p(1)$. By using this and then taking deviations from means,

$$\begin{aligned}\frac{1}{T^{3/2}}\sum_{t=2}^T s_{y,t}^{-2}x^d_{t-1} &= \phi_y\frac{1}{T^{3/2}}\sum_{t=2}^T x^d_{t-1} + \frac{1}{T^{3/2}}\sum_{t=2}^T(s_{y,t}^{-2}-\phi_y)x^d_{t-1} \\ &= \phi_y\frac{1}{T^{3/2}}\sum_{t=2}^T x^d_{t-1} + O_p(T^{-1/2}) = O_p(T^{-1/2}).\end{aligned} \tag{A13}$$

By use of $\rho = \exp(T^{-1}c/T) = 1 + T^{-1}c + o(1)$, we obtain

$$T^{-1/2}x_t = \frac{1}{\sqrt{T}}\sum_{s=1}^{t}\rho^{t-s}\epsilon_{x,s} = \frac{1}{\sqrt{T}}\sum_{s=1}^{t}\exp[T^{-1}(t-s)c]\epsilon_{x,s} + o_p(1),$$

which in turn implies

$$T^{-1/2}x_{\lfloor sT \rfloor} \to_w \sigma_x J_{x,c}(s)$$

as $T \to \infty$, where $J_{x,c}(s) = \int_0^s \exp[(s-r)c]dW_x(r)$. Therefore, via Lemma A.1,

$$\frac{1}{T}\sum_{t=2}^{T} s_{y,t}^{-2}x_{t-1}^{d}\epsilon_{y,t} = \frac{1}{T}\sum_{t=2}^{T} x_{t-1}^{d}u_t \to_w \sigma_x \int_0^1 J_{x,c}^{d}(s)dB_u(s),$$

where $J_{x,c}^{d}(r) = J_{x,c}(r) - \int_0^1 J_{x,c}(s)ds$ (see also Andrews and Guggenberger, 2012, Lemma 5 (g)). This implies

$$\begin{aligned}
\frac{1}{T}\sum_{t=2}^{T} s_{y,t}^{-2}x_{t-1}^{d}\epsilon_{y,t}^{d} &= \frac{1}{T}\sum_{t=2}^{T} s_{y,t}^{-2}x_{t-1}^{d}\epsilon_{y,t} - \frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=2}^{T} s_{y,t}^{-2}x_{t-1}^{d}\epsilon_{y,s} \\
&= \frac{1}{T}\sum_{t=2}^{T} x_{t-1}^{d}u_t + O_p(T^{-1/2}) \\
&\to_w \sigma_x \int_0^1 J_{x,c}^{d}(s)dB_u(s). \qquad \text{(A14)}
\end{aligned}$$

By Lemma 4 (c) of Andrews and Guggenberger (2012), $T^{-3/2}\sum_{t=2}^{T}(s_{y,t}^{-2} - \phi_y)(x_{t-1}^{d})^2 = O_p(1)$, from which we obtain

$$\begin{aligned}
\frac{1}{T^2}\sum_{t=2}^{T} s_{y,t}^{-2}(x_{t-1}^{d})^2 &= \phi_y \frac{1}{T^2}\sum_{t=2}^{T}(x_{t-1}^{d})^2 + \frac{1}{T^2}\sum_{t=2}^{T}(s_{y,t}^{-2} - \phi_y)(x_{t-1}^{d})^2 \\
&= \phi_y \frac{1}{T^2}\sum_{t=2}^{T}(x_{t-1}^{d})^2 + O_p(T^{-1/2}) \\
&\to_w \phi_y \sigma_x^2 \int_0^1 J_{x,c}^{d}(s)^2 ds. \qquad \text{(A15)}
\end{aligned}$$

From Lemma A.1, and the fact that $J_x(s)$ is independent of $W_u(s)$ and $W_y(s)$, we get

$$\begin{aligned}
&\frac{\frac{1}{T}\sum_{t=2}^{T} s_{y,t}^{-2} x_{t-1}^{d} \epsilon_{y,t}^{d}}{\sqrt{\frac{1}{T^2}\sum_{t=2}^{T} s_{y,t}^{-2} (x_{t-1}^{d})^2}} \\
\rightarrow_w \quad & \frac{\int_0^1 J_{x,c}^d(s)[\phi_y^{-1/2} dB_u(s)]}{\sqrt{\int_0^1 J_{x,c}^d(s)^2 ds}} \\
= \quad & \left[\frac{1-\rho_{xu}^2}{1-\rho_{xy}^2}\left(1-\frac{\rho_{xy}^2}{\rho_{xu}^2}\right)\right]^{1/2} \frac{\int_0^1 J_{x,c}^d(s) dW_u(s)}{\sqrt{\int_0^1 J_{x,c}^d(s)^2 ds}} + \rho_{xu}\frac{\int_0^1 J_{x,c}^d(s) dW_x(s)}{\sqrt{\int_0^1 J_{x,c}^d(s)^2 ds}} \\
+ \quad & \frac{\rho_{xy}(1-\rho_{xu}^2)}{\rho_{xu}\sqrt{1-\rho_{xy}^2}} \frac{\int_{s=0}^1 J_{x,c}^d(s) dW_y(s)}{\sqrt{\int_0^1 J_{x,c}^d(s)^2 ds}} \\
=_d \quad & \rho_{xu}\frac{\int_0^1 J_{x,c}^d(s) dW_x(s)}{\sqrt{\int_0^1 J_{x,c}^d(s)^2 ds}} + \left[\frac{1-\rho_{xu}^2}{1-\rho_{xy}^2}\left(1-\frac{\rho_{xy}^2}{\rho_{xu}^2}\right) + \frac{\rho_{xy}^2(1-\rho_{xu}^2)^2}{\rho_{xu}^2(1-\rho_{xy}^2)}\right]^{1/2} Z \\
= \quad & \rho_{xu}\frac{\int_0^1 J_{x,c}^d(s) dW_x(s)}{\sqrt{\int_0^1 J_{x,c}^d(s)^2 ds}} + \sqrt{1-\rho_{xu}^2} Z \qquad \text{(A16)}
\end{aligned}$$

as $T \to \infty$, where $=_d$ signifies equality in distribution. By combining this with $\beta = b/T$ and

$$T\beta\left(\frac{1}{T^2}\sum_{t=2}^{T} s_{y,t}^{-2}(x_{t-1}^d)^2\right)^{1/2} = b\left(\frac{1}{T^2}\sum_{t=2}^{T} s_{y,t}^{-2}(x_{t-1}^d)^2\right)^{1/2} \rightarrow_w \frac{b\rho_{xu}}{\gamma}\left(\int_0^1 J_{x,c}^d(s)^2 ds\right)^{1/2}, \quad \text{(A17)}$$

we obtain

$$t\text{–QGLS} \rightarrow_w \frac{b\rho_{xu}}{\gamma}\sqrt{\int_0^1 J_{x,c}^d(s)^2 ds} + \rho_{xu}\frac{\int_0^1 J_{x,c}^d(s) dW_x(s)}{\sqrt{\int_0^1 J_{x,c}^d(s)^2 ds}} + \sqrt{1-\rho_{xu}^2} Z, \qquad \text{(A18)}$$

as required for the proof. ■

Proof of Proposition 2.

Let us denote by $\hat{\beta}$ and $\hat{\rho}$ the LS estimators of β and ρ, respectively. Since $T(\hat{\beta}-\beta)$ and $T^{-1/2}x_{t-1}^d$ are $O_p(1)$,

$$\hat{\epsilon}_{y,t} \quad = \quad y_t^d - \hat{\beta}x_{t-1}^d = \epsilon_{y,t} - (\hat{\beta}-\beta)x_{t-1}^d = \epsilon_{y,t} + O_p(T^{-1/2}),$$

and, by further use of $T(\hat{\rho}-\rho) = O_p(1)$, we can similarly show that $\hat{\epsilon}_{x,t} = \epsilon_{x,t} + O_p(T^{-1/2})$.

But $T^{-1}\sum_{t=2}^{T} x_{t-1}^{d}\epsilon_{x,t}$, $T^{-1}\sum_{t=2}^{T} x_{t-1}^{d}\epsilon_{y,t}$ and $T^{-2}\sum_{t=2}^{T}(x_{t-1}^{d})^2$ are also $O_p(1)$, and therefore

$$\begin{aligned}
\frac{1}{T}\sum_{t=2}^{T}\hat{\epsilon}_{y,t}\hat{\epsilon}_{x,t} &= \frac{1}{T}\sum_{t=2}^{T}[\epsilon_{y,t}-(\hat{\beta}-\beta)x_{t-1}^{d}][\epsilon_{x,t}-(\hat{\rho}-\rho)x_{t-1}^{d}] \\
&= \frac{1}{T}\sum_{t=2}^{T}\epsilon_{y,t}\epsilon_{x,t}-(\hat{\beta}-\beta)\frac{1}{T}\sum_{t=2}^{T}x_{t-1}^{d}\epsilon_{x,t}-(\hat{\rho}-\rho)\frac{1}{T}\sum_{t=2}^{T}x_{t-1}^{d}\epsilon_{y,t} \\
&+ (\hat{\rho}-\rho)(\hat{\beta}-\beta)\frac{1}{T}\sum_{t=2}^{T}(x_{t-1}^{d})^2 \\
&= \frac{1}{T}\sum_{t=2}^{T}\epsilon_{y,t}\epsilon_{x,t}+O_p(T^{-1/2}),
\end{aligned} \tag{A19}$$

where

$$\frac{1}{T}\sum_{t=2}^{T}\epsilon_{y,t}\epsilon_{x,t}=\frac{1}{T}\sum_{t=2}^{T}(\gamma\epsilon_{x,t}+\epsilon_{y.x,t})\epsilon_{x,t}=\gamma\frac{1}{T}\sum_{t=2}^{T}\epsilon_{x,t}^2+O_p(T^{-1/2}). \tag{A20}$$

By the same arguments

$$\frac{1}{T}\sum_{t=2}^{T}\hat{\epsilon}_{x,t}^2=\frac{1}{T}\sum_{t=2}^{T}\epsilon_{x,t}^2+O_p(T^{-1/2}), \tag{A21}$$

and so we obtain

$$\hat{\gamma}=\gamma+O_p(T^{-1/2}). \tag{A22}$$

Consider $\hat{s}_{y,t}^2$. We now show that $(\hat{\sigma}_{y,t}^2-\sigma_{y,t}^2)=o_p(1)$. The proof begins by noting that by the consistency of $\hat{e}_{m,t}$ for $m\in\{x,y.x\}$, $(\hat{\epsilon}_{m,t}^2-\epsilon_{m,t}^2)$, and hence also $(\hat{e}_{m,t}-e_{m,t})$, are $O_p(T^{-1/2})$. Now, while these results are just point-wise in t, it can be shown (by direct insertion and computation) that

$$\frac{1}{T}\sum_{t=q_m+2}^{T}\hat{e}_{m,t}\hat{e}_{m,t}' = \frac{1}{T}\sum_{t=q_m+2}^{T}e_{m,t}e_{m,t}'+O_p(T^{-1/2}), \tag{A23}$$

$$\frac{1}{\sqrt{T}}\sum_{t=q_m+2}^{T}\hat{e}_{m,t}\hat{\epsilon}_{m,t}^2 = \frac{1}{\sqrt{T}}\sum_{t=q_m+2}^{T}e_{m,t}\epsilon_{m,t}^2+O_p(T^{-1/2}). \tag{A24}$$

Thus, since $\epsilon_{m,t}^2=\lambda_m'e_{m,t}+(\epsilon_{m,t}^2-s_{m,t}^2)$,

$$\hat{\lambda}_m-\lambda_m=\left(\frac{1}{T}\sum_{t=q_m+2}^{T}e_{m,t}e_{m,t}'\right)^{-1}\frac{1}{T}\sum_{t=q_m+2}^{T}e_{m,t}(\epsilon_{m,t}^2-s_{m,t}^2)+O_p(T^{-1/2}), \tag{A25}$$

whose order depends on the assumed number of finite moments of $\epsilon_{m,t}$. With four moments, by the ergodic theorem, $T^{-1}\sum_{t=q_m+2}^{T}e_{m,t}(\epsilon_{m,t}^2-s_{m,t}^2)=o_p(1)$. It follows that

$$\hat{\lambda}_m-\lambda_m=o_p(1), \tag{A26}$$

and therefore

$$\hat{s}_{m,t}^2 = \hat{\lambda}_m' \hat{e}_{m,t} = \lambda_m' e_{m,t} + (\hat{\lambda}_m - \lambda_m)' \hat{e}_{m,t} + \lambda_m' (\hat{e}_{m,t} - e_{m,t}) = s_{m,t}^2 + o_p(1). \tag{A27}$$

By using this, the consistency of $\hat{\gamma}$, and Taylor expansion,

$$\begin{aligned}
\hat{s}_{y,t}^2 &= \hat{\gamma}^2 \hat{s}_{x,t}^2 + \hat{s}_{y.x,t}^2 \\
&= \gamma^2 s_{x,t}^2 + s_{y.x,t}^2 + (\hat{\gamma}^2 - \gamma^2)\hat{s}_{x,t}^2 + \gamma^2(\hat{s}_{x,t}^2 - s_{x,t}^2) + (\hat{s}_{y.x,t}^2 - s_{y.x,t}^2) \\
&= s_{y,t}^2 + o_p(1).
\end{aligned} \tag{A28}$$

Note that $(\hat{s}_{y,t}^{-2} - s_{y,t}^{-2})^2 \leq 2(\hat{s}_{y,t}^{-4} - s_{y,t}^{-4})$. This, $(\hat{s}_{y,t}^2 - s_{y,t}^2) = o_p(1)$ and Taylor expansion give

$$\frac{1}{T}\sum_{t=q+2}^{T} (\hat{s}_{y,t}^{-2} - s_{y,t}^{-2})^2 = o_p(1),$$

where $q = \max\{q_x, q_{y.x}\}$. Application of the Cauchy–Schwarz inequality now yields

$$\frac{1}{T^2}\sum_{t=q+2}^{T} (\hat{s}_{y,t}^{-2} - s_{y,t}^{-2})(x_{t-1}^d)^2 \leq \left(\frac{1}{T}\sum_{t=q+2}^{T} (\hat{s}_{y,t}^{-2} - s_{y,t}^{-2})^2\right)^{1/2} \left(\frac{1}{T^3}\sum_{t=q+2}^{T} (x_{t-1}^d)^4\right)^{1/2} = o_p(1).$$

Hence,

$$\begin{aligned}
\frac{1}{T^2}\sum_{t=q+2}^{T} \hat{s}_{y,t}^{-2}(x_{t-1}^d)^2 &= \frac{1}{T^2}\sum_{t=q+2}^{T} s_{y,t}^{-2}(x_{t-1}^d)^2 + \frac{1}{T^2}\sum_{t=q+2}^{T} (\hat{s}_{y,t}^{-2} - s_{y,t}^{-2})(x_{t-1}^d)^2 \\
&= \frac{1}{T^2}\sum_{t=q+2}^{T} s_{y,t}^{-2}(x_{t-1}^d)^2 + o_p(1),
\end{aligned} \tag{A29}$$

and, by using the same arguments as in Proof of Theorem 2 of Andrews and Guggenberger (2012), we can further show that

$$\begin{aligned}
\frac{1}{T}\sum_{t=q+2}^{T} \hat{s}_{y,t}^{-2} x_{t-1}^d \epsilon_{y,t} &= \frac{1}{T}\sum_{t=q+2}^{T} s_{y,t}^{-2} x_{t-1}^d \epsilon_{y,t} + \frac{1}{T}\sum_{t=q+2}^{T} (\hat{s}_{y,t}^{-2} - s_{y,t}^{-2}) x_{t-1}^d \epsilon_{y,t} \\
&= \frac{1}{T}\sum_{t=q+2}^{T} s_{y,t}^{-2} x_{t-1}^d \epsilon_{y,t} + o_p(1).
\end{aligned} \tag{A30}$$

These results, together with Taylor expansion, imply

$$\begin{aligned}
t\text{–FQGLS} &= \frac{\sum_{t=q+2}^{T} \hat{s}_{y,t}^{-2} x_{t-1}^d y_t}{\sqrt{\sum_{t=q+2}^{T} \hat{s}_{y,t}^{-2}(x_{t-1}^d)^2}} \\
&= T\beta\left(T^{-2}\sum_{t=q+2}^{T} \hat{s}_{y,t}^{-2}(x_{t-1}^d)^2\right)^{1/2} + \frac{T^{-1}\sum_{t=q+2}^{T} \hat{s}_{y,t}^{-2} x_{t-1}^d \epsilon_{y,t}^d}{\sqrt{T^{-2}\sum_{t=q+2}^{T} \hat{s}_{y,t}^{-2}(x_{t-1}^d)^2}} \\
&= T\beta\left(T^{-2}\sum_{t=q+2}^{T} s_{y,t}^{-2}(x_{t-1}^d)^2\right)^{1/2} + \frac{T^{-1}\sum_{t=q+2}^{T} s_{y,t}^{-2} x_{t-1}^d \epsilon_{y,t}^d}{\sqrt{T^{-2}\sum_{t=q+2}^{T} s_{y,t}^{-2}(x_{t-1}^d)^2}} + o_p(1) \\
&= t\text{–QGLS} + o_p(1),
\end{aligned} \tag{A31}$$

and so the proof is complete. ■

Proof of Proposition 3.

As we show in Proposition 2 the asymptotic distribution of t–FQGLS under the null hypothesis is exactly the same as the asymptotic distribution of the unit root t-statistic of Andrews and Guggenberger (2009a; 2009b). Given this, by the arguments in Andrews and Guggenberger (2009b), the limit of the coverage probabilities of the t–FQGLS confidence intervals equal that of the corresponding intervals in Andrews and Guggenberger (2009a; 2009b), which is greater than or equal to the nominal size $1 - \alpha$, where α is the significance level. Hence, the t–FQGLS confidence intervals have correct asymptotic size. ■

Table 1: 5% size and size-adjusted power in E1.

	$b=0$				$b=5$				$b=10$				$b=20$			
T	SYM	EQ	LS	LEW	SYM	EQ	LS	LEW	SYM	EQ	LS	LEW	SYM	EQ	LS	LEW
								$c=\gamma=0$								
50	3.7	1.8	6.7	7.4	45.0	45.1	43.5	42.8	81.1	81.3	79.3	77.8	98.3	98.3	97.5	96.8
100	5.0	3.1	6.2	6.0	47.6	49.6	46.9	44.6	83.5	84.7	81.5	79.8	99.2	99.4	98.7	98.3
200	3.8	2.9	5.5	5.9	50.9	50.9	44.9	43.6	87.5	87.5	82.8	82.3	99.5	99.5	98.6	98.5
								$c=-10$, $\gamma=-0.5$								
50	2.4	1.7	11.0	25.1	15.6	9.8	10.0	0.9	47.5	35.6	33.4	5.5	91.5	86.8	84.0	49.8
100	2.1	2.9	9.9	20.4	17.6	13.6	13.0	0.9	52.4	45.1	41.1	5.3	94.3	92.7	88.5	51.9
200	2.1	3.0	9.4	18.3	18.6	12.9	11.8	0.8	55.8	46.0	38.0	4.9	96.7	95.3	90.7	54.5
								$c=-10$, $\gamma=-1$								
50	2.2	2.4	16.0	55.9	10.4	7.5	6.7	0.3	27.1	18.3	15.9	0.0	83.2	74.4	68.5	6.0
100	1.8	4.0	15.5	50.5	13.2	9.9	7.6	0.6	37.2	27.6	20.1	0.1	89.8	85.3	74.5	4.7
200	1.3	4.1	14.6	48.8	13.5	10.1	6.6	0.3	38.1	28.8	17.5	0.1	92.3	88.7	73.7	5.3
								$c=-20$, $\gamma=-0.5$								
50	3.0	2.1	14.4	45.0	10.7	9.3	8.2	0.8	28.1	23.4	21.3	0.2	79.3	74.8	68.3	5.0
100	2.5	2.7	11.8	38.5	12.7	10.4	10.4	1.0	31.7	26.6	25.3	0.1	83.2	79.7	73.8	5.3
200	2.8	3.5	10.8	34.4	11.1	9.6	8.0	0.6	30.9	27.4	21.2	0.4	84.8	81.8	71.0	4.9
								$c=-20$, $\gamma=-1$								
50	3.2	3.0	21.6	83.5	9.6	6.3	6.3	1.1	17.0	12.9	12.3	0.1	55.3	44.4	41.8	0.0
100	2.0	3.8	18.6	79.3	9.9	6.9	7.3	1.3	20.9	15.5	14.8	0.3	65.5	56.0	47.3	0.0
200	1.8	4.0	17.9	79.8	8.0	6.8	5.8	1.0	18.7	14.8	10.6	0.1	66.0	58.8	37.3	0.0

Notes: γ refer to the covariance between the errors in the predictive and regressor equations, and c and b refer to the drifts in the regressor AR parameter and predictive slope, respectively. "SYM", "EQ", "LS" and "LEW" refer to symmetric subsample FQGLS, equal-tailed hybrid subsample FQGLS, t–LS, and the Lewellen (2004) t-test, respectively.

Table 2: 5% size and size-adjusted power in E2.

	$b = 0$				$b = 5$				$b = 10$				$b = 20$			
T	SYM	EQ	LS	LEW	SYM	EQ	LS	LEW	SYM	EQ	LS	LEW	SYM	EQ	LS	LEW
								GARCH								
50	3.1	1.8	8.1	19.5	7.7	6.5	5.8	2.5	13.4	11.4	10.9	1.9	35.5	30.1	30.9	5.8
100	3.7	2.1	8.1	15.4	6.5	5.8	5.8	2.9	10.9	9.6	9.7	2.0	30.5	28.8	27.5	5.3
200	2.7	2.0	6.5	11.8	6.5	5.8	6.0	2.4	12.2	10.1	9.1	1.6	31.6	28.6	24.3	5.2
								X–ARCH								
50	4.0	2.0	13.9	27.9	7.8	6.4	6.4	3.0	14.3	10.8	10.7	2.0	41.5	34.0	32.0	5.5
100	4.8	3.1	14.4	26.8	7.2	6.1	5.7	2.9	13.4	11.1	9.8	2.5	40.8	36.1	28.1	5.7
200	2.5	2.2	10.9	21.9	7.8	6.8	5.0	2.6	16.1	13.9	9.6	2.0	47.6	43.6	28.2	5.5
								Asymmetric threshold ARCH								
50	3.4	2.0	10.6	31.0	9.3	6.4	7.9	1.4	20.0	15.0	18.0	0.5	60.3	52.1	55.3	5.8
100	3.6	3.2	10.6	26.4	8.7	7.5	7.6	1.3	18.9	16.4	15.9	0.7	59.7	56.9	52.0	5.3
200	2.6	2.1	9.1	23.2	9.6	8.0	7.0	1.1	21.9	18.4	14.8	0.7	64.2	59.5	50.7	5.8
								AR–SV								
50	2.7	1.4	7.4	15.2	5.8	5.2	5.0	4.0	8.6	7.2	7.4	3.7	19.9	17.1	17.2	5.5
100	4.0	2.1	7.2	11.0	5.7	5.8	5.7	3.5	7.7	7.9	8.4	3.1	17.0	17.5	16.9	5.1
200	3.5	1.9	5.9	8.2	6.2	6.1	5.6	3.6	8.2	7.9	7.3	2.9	18.2	16.9	16.1	5.3

Notes: See the text for a more detailed description of the various conditional variance models. See Table 1 for an explaneation of the test.

Table 3: Descriptive statistics.

Variable	Description	Sample	Mean	CV	Skewness	Kurtosis	ADF	p-value	ARCH	p-value
		Sample: Financial ratios								
Returns	S&P500 returns	1921M03–2008M12	0.005	11.72	−0.54	11.23	−26.25	0.00	20.6	0.00
BM	Book-to-market	1921M03–2008M12	250.100	1.59	1.89	5.21	−0.27	1.00	508.3	0.00
LTY	Long-term bond yield	1919M01–2008M12	0.053	0.52	1.07	3.68	−1.17	0.69	389.6	0.00
EP	Earning price	1871M01–2008M12	0.075	0.36	0.98	4.09	−3.56	0.01	1674.0	0.00
LTR	Long-term bond return	1926M01–2008M12	0.005	4.79	0.80	8.51	−29.38	0.00	21.0	0.00
INFL	Inflation	1913M02–2008M12	0.003	2.37	0.62	11.34	−8.86	0.00	29.9	0.00
SVAR	Stock variance	1885M02–2008M12	0.002	1.97	6.77	64.36	−4.99	0.00	15.9	0.00
NTIS	Net equity expansion	1926M12–2008M12	0.019	1.25	1.39	10.07	−3.69	0.00	1080.0	0.00
TMS	Term spread	1920M02–2008M12	0.015	0.83	−0.13	3.13	−5.26	0.00	213.3	0.00
CORPR	Corporate bond rate	1926M01–2008M12	0.005	4.29	0.87	10.95	−21.21	0.00	39.2	0.00
DFY	Default yield spread	1919M01–2008M12	0.012	0.59	2.06	9.53	−2.75	0.07	432.3	0.00
DFR	Default return spread	1926M01–2008M12	0.000	120.00	−0.57	11.86	−37.09	0.00	22.4	0.00
DP	Dividend price	1871M01–2008M12	0.045	10.58	0.48	4.18	−3.78	0.00	2722.0	0.00
		Sample: Macroeconomic variables								
Returns	S&P500 returns	1/1/1971–5/25/2012	0.123	18.74	−0.57	8.44	−48.92	0.00	41.9	0.00
ER	Dollar–pound exchange rate	1/1/1971–5/25/2012	0.580	0.16	−0.07	3.12	0.00	0.16	50.5	0.00
OIL	US crude oil	5/13/1983–5/25/2012	36.911	0.72	1.51	4.43	−1.04	0.74	50.3	0.00

Notes: CV, ADF and ARCH refer to the coefficient of variation, the augmented Dickey–Fuller test, and a Lagrange multiplier test of the zero slope restriction is an ARCH regression of order six.

Table 4: Predictive regression diagnostics.

	Returns ARCH		Predictor ARCH		Endogeneity	
Predictor	Wald	p-value	Wald	p-value	$\hat{\gamma}$	p-value
Sample: Financial ratios						
BM	81.97	0.00	574.96	0.00	−1.00	0.00
LTY	146.59	0.00	282.10	0.00	−3.05	0.00
EP	510.52	0.00	133.75	0.00	−9.27	0.00
LTR	135.36	0.00	103.65	0.00	0.28	0.00
INFL	162.88	0.00	58.20	0.00	0.39	0.18
SVAR	259.12	0.00	71.69	0.00	−3.79	0.00
NTIS	148.87	0.00	81.69	0.00	−0.40	0.22
TMS	144.46	0.00	54.06	0.00	−0.25	0.58
CORPR	127.94	0.00	198.34	0.00	0.50	0.00
DFY	73.54	0.00	296.32	0.00	−10.06	0.00
DFR	123.65	0.00	96.87	0.00	0.55	0.00
DP	170.82	0.00	209.57	0.00	−14.32	0.00
Sample: Macroeconomic variables						
ER	222.91	0.00	284.70	0.00	−19.05	0.00
OIL	247.01	0.00	0.00	1.00	0.11	0.00

Notes: The Wald test tests the zero slope restriction in the estimated ARCH equations for $\hat{s}^2_{y.x}$ and $\hat{s}^2_x$, and γ refer to the slope coefficient in a regression of $\hat{e}_{y.x,t}$ onto $\hat{e}_{x,t}$. See Table 3 for an explanation of the various predictors.

Table 5: Subsample FQGLS 95% confidence intervals for β.

	Full sample		Post war	
Predictor	SYM	EQ	SYM	EQ
	Sample: Financial ratios			
BM	$[-0.00518, 0.00119]$	$[-0.00562, -0.00077]$	$[-0.00875, -0.00077]$	$[-0.00906, -0.00332]$
LTY	$[-0.00018, 0.00038]$	$[-0.00019, 0.00033]$	$[-0.00043, 0.00036]$	$[-0.00048, 0.00027]$
EP	$[-0.00940, -0.00114]$	$[-0.00970, -0.00375]$	$[-0.00904, 0.00396]$	$[-0.00958, 0.00052]$
LTR	$[-0.00057, 0.00003]$	$[-0.00059, 0.00000]$	$[-0.00111, -0.00003]$	$[-0.00120, -0.00027]$
INFL	$[-0.00010, 0.00013]$	$[-0.00010, 0.00013]$	$[-0.00023, 0.00000]$	$[-0.00024, -0.00004]$
SVAR	$[-0.00037, -0.00008]$	$[-0.00038, -0.00019]$	$[-0.00028, -0.00001]$	$[-0.00029, -0.00012]$
NTIS	$[-0.00055, -0.00010]$	$[-0.00056, -0.00024]$	$[-0.00059, -0.00004]$	$[-0.00060, -0.00016]$
TMS	$[-0.00022, 0.00026]$	$[-0.00016, 0.00028]$	$[-0.00010, 0.00063]$	$[0.00003, 0.00066]$
CORPR	$[-0.00042, 0.00016]$	$[-0.00041, 0.00023]$	$[-0.00088, -0.00002]$	$[-0.00092, -0.00020]$
DFY	$[-0.00021, -0.00002]$	$[-0.00023, -0.00003]$	$[0.00000, 0.00012]$	$[0.00001, 0.00012]$
DFR	$[-0.00001, 0.00034]$	$[0.00008, 0.00037]$	$[-0.00014, 0.00037]$	$[0.00001, 0.00040]$
DP	$[-0.01520, -0.00612]$	$[-0.01561, -0.01001]$	$[-0.01029, 0.00107]$	$[-0.01070, -0.00197]$
	Sample: Macroeconomic variables			
ER	$[0.00022, 0.00222]$	$[0.00011, 0.00211]$	–	–
OIL	$[-1.13736, -0.34598]$	$[-1.15718, -0.40511]$	–	–

Notes: See Table 3 for an explanation.