

TECHNICAL WORKING PAPER SERIES

ESTIMATING CONDITIONAL EXPECTATIONS
WHEN VOLATILITY FLUCTUATES

Robert F. Stambaugh

Technical Working Paper No. 140

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 1993

I am grateful for comments by Marshall Blume, Tim Bollerslev, Robert Engle, Robert Hodrik, Michael Gibbons, and workshop participants at the University of Pennsylvania. This paper is part of NBER's research program in Asset Pricing. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

ESTIMATING CONDITIONAL EXPECTATIONS
WHEN VOLATILITY FLUCTUATES

ABSTRACT

Asymptotic variances of estimated parameters in models of conditional expectations are calculated analytically assuming a GARCH process for conditional volatility. Under such heteroskedasticity, OLS estimators of parameters in single-period models can possess substantially larger asymptotic variances than GMM estimators employing additional multiperiod moment conditions - an approach yielding no efficiency gain under homoskedasticity. In estimating models of long-horizon expectations, the VAR approach provides an efficiency advantage over long-horizon regressions under homoskedasticity, but that ordering can reverse under heteroskedasticity, especially when the conditional mean and variance are both persistent. In such cases, the VAR approach maintains a slight efficiency advantage if the OLS estimator is replaced by an alternative GMM estimator. Heteroskedasticity can increase dramatically the apparent asymptotic power advantages of long-horizon regressions to reject constant expectations against persistent alternatives.

Robert F. Stambaugh
Finance Department
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104-6367
and NBER

1. Introduction

Much empirical research in finance and economics attempts to estimate models describing changes through time in conditional expectations of times series. The series of interest, such as returns on financial assets or growth rates of economic aggregates, often appear to exhibit significant fluctuations in conditional volatility. This study investigates the role of such heteroskedasticity in the estimation of conditional expectations.

Large-sample variances of estimators of parameters in models of expectations are calculated analytically under a GARCH specification for conditional heteroskedasticity.¹ Many of the estimators considered have been used often in previous empirical research, especially in finance, and in all cases the estimators use the Generalized Method of Moments (GMM) to exploit only the unpredictability of the forecast errors in linear models of conditional expectations.² The primary objective is to explore the precision of estimators obtained using this popular approach when the forecast errors possess conditional heteroskedasticity. A GARCH process for the heteroskedasticity is specified simply to allow analytic results. The investigation does not entertain methods, such as maximum likelihood, that would incorporate a model of conditional heteroskedasticity in the estimation. Such alternative approaches may offer a fruitful direction for subsequent research, however, especially given the apparently important effects of heteroskedasticity found in this study. Future research can also investigate, probably using Monte Carlo experiments, the extent to which the asymptotic results approximate the corresponding properties in finite samples. It is hoped that the design of such experiments can be guided by the analytic results obtained here, which allow precise statements about the effects of heteroskedasticity across a wide range of behaviors for conditional means and variances.

The investigation is conducted in the context of a model with two variables, x_t and y_t . Conditional expectations of these variables are assumed to obey the two equations,

$$y_{t+1} = \alpha + \rho y_t + \epsilon_{t+1}, \quad (1.1)$$

$$x_{t+1} = a + b y_t + u_{t+1}, \quad (1.2)$$

where $0 \leq \rho < 1$, and both ϵ_{t+1} and u_{t+1} have zero expectations conditioned on information available at time t . This simple model provides a tractable framework that has allowed a number of studies to investigate properties of estimators of conditional expectations in both

¹Bollerslev (1986) introduces Generalized Autoregressive Conditional Heteroskedasticity (GARCH), a generalization of the ARCH processes of Engle (1982). Bollerslev, Chou, and Kroner (1992) review the use of such models in applications in finance and economics.

²Hansen (1982) develops the asymptotic properties of GMM tests and estimators.

small and large samples.³ In a common finance application, the principal motivation for this study, x_{t+1} is an asset's rate of return from time t to $t + 1$ and y_t , observed at time t , is assumed to capture the information relevant to predicting that return. A typical choice for y_t is a dividend yield, a bond yield, or another variable that is inversely related to an asset's current price.⁴

The univariate first-order autoregression for y_t in (1.1) serves as the starting point for the investigation. After quantifying the increase in the variance of the ordinary-least-squares (OLS) estimator of ρ under heteroskedasticity, section 2 considers alternative estimators that exploit additional moment conditions using Hansen's (1982) Generalized Method of Moments (GMM). An estimator formed by using multiple lagged values of y_t as instruments—an approach with no asymptotic advantage under homoskedasticity—can provide substantial reductions in asymptotic variance under heteroskedasticity, especially when the conditional variance is persistent.⁵ The number of additional moment conditions underlying such an estimator can be large, however, and section 2 goes on to consider an estimator that replaces those multiple moment conditions with one condition based on a simple multiperiod forecast error. This estimator appears to provide efficiency gains relative to OLS that, in many cases, nearly match those delivered by the estimator computed with multiple additional moment conditions.

Section 3 extends the analysis to consider the estimation of the parameters of the second equation, the regression of x_{t+1} on y_t in (1.2). In the finance application mentioned above, this equation gives the asset's conditional expected single-period return. An estimator obtained by adding two moment conditions based on multiperiod forecast errors again provides substantial asymptotic efficiency gains relative to the OLS estimator in the presence of heteroskedasticity.

Equations (1.1) and (1.2) provide conditional expectations of y_t and x_t one period ahead, but conditional expectations of multiperiod quantities are often of interest as well. In finance applications, for example, researchers have analyzed conditional expectations of the multiperiod sum $\sum_{j=1}^K x_{t+j}$, a K -period return. Estimating this conditional expectation in the presence of heteroskedasticity is the topic of section 4. The "long-horizon-regression" approach estimates the conditional expectation by regressing observations of $\sum_{j=1}^K x_{t+j}$ directly on y_t , where the observations can either overlap by $K - 1$ periods or be spaced so as

³Examples include Mankiw and Shapiro (1986), Stambaugh (1986), Nelson and Kim (1993), Campbell (1992), and Boudoukh and Richardson (1993).

⁴See, for example, Keim and Stambaugh (1986) and Fama and French (1988).

⁵As used throughout this study, the term "persistence" describes a stationary series with high autocorrelation, as opposed to an integrated process.

not to overlap. The efficiency gains achieved by overlapping the observations are somewhat greater when the conditional variance is time-varying and persistent, although these gains become modest in any case as ρ approaches unity. An alternative approach to estimating the conditional expectation of $\sum_{j=1}^K x_{t+j}$ is first to compute OLS estimates of the parameters of the two-equation system in (1.1) and (1.2) and then to use those estimates along with the implications of that vector autoregression (VAR) in computing the multiperiod expectation. Under homoskedasticity, the VAR approach offers efficiency advantages over the long-horizon regression, but heteroskedasticity can reverse this ordering, especially when the conditional variance is persistent. When, instead of OLS, GMM with two additional multiperiod moment conditions is used to estimate the parameters in (1.1) and (1.2), it appears that the VAR approach maintains its efficiency advantage even under heteroskedasticity. The advantage is minimal, however, when conditional volatility is persistent.

Section 4 also investigates the effects of heteroskedasticity on the power of long-horizon regressions to reject the null hypothesis of constant expectations. Using the “approximate slope” measure due to Bahadur (1960) and Geweke (1981), Campbell (1992) finds that, as compared to single-period regressions, long-horizon regressions can possess asymptotic power advantages against alternative hypotheses in which conditional expectations are persistent. Campbell’s analysis assumes homoskedasticity. The investigation here reveals that the apparent asymptotic power advantages of long-horizon regressions can increase dramatically in the presence of heteroskedasticity.

2. The Univariate Case

In order to focus on the most essential aspects of the problem and to lay some groundwork for the two-equation case, we first consider estimating only the parameters in equation (1.1), the first-order autoregression for y_t . Assume that σ_t^2 , the conditional variance of ϵ_{t+1} , obeys a GARCH(1,1) process,

$$E_t\{\epsilon_{t+1}^2\} = \sigma_t^2 = \gamma_0 + \gamma_1\sigma_{t-1}^2 + \gamma_2\epsilon_t^2, \quad (2.1)$$

where γ_1 and γ_2 are nonnegative and

$$\gamma \equiv \gamma_1 + \gamma_2 < 1. \quad (2.2)$$

Let σ_ϵ^2 denote the unconditional variance of ϵ_t . As observed by Engle and Bollerslev (1986), equation (2.1) implies

$$E_t\{\sigma_{t+1}^2 - \sigma_t^2\} = \gamma(\sigma_t^2 - \sigma_\epsilon^2), \quad (2.3)$$

so γ is the autocorrelation of the conditional variance.⁶ The conditional distribution of ϵ_{t+1} need not be normal, but assume

$$E_t\{\epsilon_{t+1}^3\} = 0 \quad (2.4)$$

and that $E\{\epsilon_t^4\}$ exists. Also define

$$\kappa = \frac{E\{\epsilon_t^4\}}{\sigma_t^4} - 3, \quad (2.5)$$

the unconditional excess kurtosis of ϵ_t .⁷

2.1. The Asymptotic Variance of the Sample Autocorrelation

Let $\hat{\rho}_{(1)}$ denote the OLS estimator of ρ . (The subscript notation is explained and used later.) When the disturbances in (1.1) are homoskedastic, it is well known that $\sqrt{T}(\hat{\rho}_{(1)} - \rho)$ converges in distribution to a normal variate with zero mean and asymptotic variance $1 - \rho^2$ as T grows large.⁸

More generally, when the disturbances in (1.1) obey the GARCH (1,1) process as specified above, it can be shown that the asymptotic variance (multiplied by T) of $\hat{\rho}_{(1)}$ is given by

$$\text{var}(\hat{\rho}_{(1)}) = (1 - \rho^2) \left[1 + \frac{\gamma(1 - \rho^2)(\kappa + 2)}{1 - \rho^2\gamma} \right], \quad (2.6)$$

which simplifies to $1 - \rho^2$ in the homoskedastic case of $\gamma = 0$. (Details are provided in the Appendix.) The bracketed quantity in (2.6) is greater than or equal to unity, increasing in κ and γ , and decreasing in ρ . In other words, the relative increase in asymptotic variance is larger when ϵ_t^2 is more volatile, σ_t^2 is more persistent, but y_t is less persistent.⁹

Table 1 reports the ratio of $\hat{\rho}_{(1)}$'s asymptotic variance under homoskedasticity to that under heteroskedasticity. That is, each entry in table 1 is the reciprocal of the bracketed

⁶Although they do not address parameter estimation, Baillie and Bollerslev (1992) analyze the distribution of prediction errors from the same AR(1)-GARCH(1,1) model specified above.

⁷If the conditional excess kurtosis of ϵ_t is constant, denoted by κ_c , then

$$\kappa_c = \frac{\kappa(1 - \gamma^2 - 2\gamma_2^2) - 6\gamma_2^2}{\kappa\gamma_2^2 + 1 - \gamma^2 + 3\gamma_2^2},$$

which, given γ and κ , obeys the bounds

$$\frac{\kappa(1 - 3\gamma^2) - 6\gamma^2}{\kappa\gamma^2 + 1 + 2\gamma^2} \leq \kappa_c \leq \kappa.$$

⁸See, for example, chapter 7 of Brockwell and Davis (1987).

⁹Note that, by the definition of excess kurtosis κ , the unconditional variance of ϵ_t^2 is given by $(\kappa + 2)\sigma_t^4$, and $\kappa \geq -2$.

quantity in (2.6) for a given combination of γ , κ , and ρ . The choices of κ are intended to include a fairly wide but reasonable range of values relevant to applications in economics and finance. The values $\kappa = 3$ and $\kappa = 13$ (used in subsequent tables as well) approximate the sample excess kurtosis of the residuals of (1.1) fitted using OLS over two periods, 1/1927–12/1952 and 1/1953–12/1989, where y_t is the monthly time series of the dividend-price ratio (D/P) for the value-weighted portfolio of stocks on the New York Stock Exchange (NYSE). The dividend-price ratio for month t is computed as the sum of all dividends paid during the twelve months through month t divided by the value of the portfolio at the end of month t . It should be noted that even the higher value of $\kappa = 13$ is considerably lower than the sample excess kurtosis of 26 obtained from the residuals of (1.1) using the D/P series for the entire 1927–89 period.

As table 1 indicates, the largest relative increases in the asymptotic variance of $\hat{\rho}_{(1)}$ occur when γ is high and ρ is low, but the increases when ρ is high can be substantial as well.¹⁰ The value $\rho = 0.97$ corresponds to the sample autocorrelation of D/P over the 1927–89 period. Even for this value of ρ , the ratios in table 1 are as low as 0.161 when $\gamma = 0.9$, and a number of previous studies have estimated the persistence of the conditional variance of financial time series to be at least that high. For example, Hodrick (1992) estimates a value of $\gamma = 0.98$ in a GARCH(1,1) process for unanticipated movements in the dividend-price ratio.¹¹ In general, it appears that conditional heteroskedasticity can produce large increases in the asymptotic variances of sample autocorrelations.

2.2. Adding Multiple Moment Conditions

In the presence of conditional heteroskedasticity, one approach to estimating α and ρ is to specify a conditional density function for ϵ_t and then compute maximum likelihood estimates using both the AR(1) process for y_t in (1.1) as well as the GARCH(1,1) process for σ_t^2 in (2.1). This study investigates estimation approaches that specify neither the process for σ_t^2 nor the conditional density of ϵ_t . The estimators considered here rely only on the requirement that the forecast errors in (1.1) have zero conditional expectations. The large-sample variances of such estimators in the presence of heteroskedasticity are analyzed using the GARCH(1,1) process.

The requirement $E_t\{\epsilon_{t+1}\} = 0$ can be used to generate many moment conditions that

¹⁰Milhøj (1985) and Diebold (1986, 1988) show that the asymptotic variances of sample autocorrelations can be substantially larger in the presence of ARCH.

¹¹Hodrick specifies a GARCH(1,1) model in a somewhat different setting than assumed above. Specifically, he estimates a multivariate constant-correlation GARCH process for the disturbances of a VAR in three variables, one of which is D/P , and he assumes multivariate normality.

could be used in a GMM procedure. One set of conditions can be obtained from the $K + 1$ element vector

$$g_{K,t} = \begin{bmatrix} \epsilon_{t+1} \\ \epsilon_{t+1}y_t \\ \epsilon_{t+2}y_t \\ \vdots \\ \epsilon_{t+K}y_t \end{bmatrix}. \quad (2.7)$$

(For later tractability, these conditions are stated in terms of leading the ϵ 's rather than the equivalent representation of lagging the y 's.) Let $\hat{\theta}_{(K)} = (\hat{\alpha}_{(K)} \hat{\rho}_{(K)})'$ denote the optimal GMM estimator of θ based on the $K + 1$ moment conditions $E\{g_{K,t}\} = 0$. That is, $\hat{\theta}_{(K)}$ is the value of θ that minimizes $\bar{g}'_K \hat{S}_K^{-1} \bar{g}_K$, where $\bar{g}_K = (1/T) \sum_{t=1}^T g_{K,t}$, T is the sample size, and \hat{S}_K is a consistent estimator of

$$S_K = \sum_{t=-\infty}^{\infty} E\{g_{K,t} g'_{K,t-l}\}. \quad (2.8)$$

Note that $\hat{\theta}_{(1)}$ is the OLS estimator. Hansen (1982) shows that as T grows large $\sqrt{T}(\hat{\theta}_{(K)} - \theta)$ converges in distribution to a normally distributed vector with mean zero and covariance matrix $(D'_K S_K^{-1} D_K)^{-1}$, where

$$D_K = E \left[\frac{\partial g_{K,t}}{\partial \theta'} \right]. \quad (2.9)$$

Under the assumptions for (1.1) and the GARCH process in (2.1), D_K and S_K can be calculated analytically. Let μ_y and σ_y^2 denote the unconditional mean and variance of y_t . Then

$$D_K = - \begin{bmatrix} 1 & \mu_y \\ \mu_y & \mu_y^2 + \sigma_y^2 \\ \mu_y & \mu_y^2 + \rho\sigma_y^2 \\ \vdots & \vdots \\ \mu_y & \mu_y^2 + \rho^{K-1}\sigma_y^2 \end{bmatrix}, \quad (2.10)$$

and

$$S_K = \sigma_y^2(1 - \rho^2) \begin{bmatrix} 1 & \mu_y \iota' \\ \mu_y \iota & \Phi \end{bmatrix}, \quad (2.11)$$

where ι is a K -vector of ones and the (i, j) element of Φ is given by

$$\phi(i, j) = \mu_y^2 + \rho^{|i-j|} \sigma_y^2 \left(1 + \gamma^{\max(i,j)} \left[\frac{(1 - \rho^2)(\kappa + 2)}{1 - \rho^2 \gamma} \right] \right). \quad (2.12)$$

(Details are provided in the Appendix.)

It is useful to observe that $\hat{\theta}_{(K)}$ is the efficient instrumental variable (IV) estimator of the regression in (1.1) when the instrument vector is $(1 \ y_t \ y_{t-1} \ \cdots \ y_{t-K})$. If the regression disturbances are serially uncorrelated and homoskedastic, standard results imply that

this IV estimator has the same asymptotic covariance matrix as the OLS estimator.¹² That is, including additional lags of y_t offers no asymptotic efficiency advantage, which can be confirmed by evaluating (2.10) through (2.12) in the homoskedastic case of $\gamma = 0$. In the presence of heteroskedastic regression disturbances, it is possible that expanding the set of instruments used to compute an IV estimator can produce efficiency gains, as has been analyzed in general settings by Cragg (1983), White (1984), and others. In the model employed here, it can be verified that when $\rho = 0$, $\text{var}(\hat{\theta}_{(K)}) = \text{var}(\hat{\theta}_{(1)})$ for all $K > 1$. That is, the additional moment conditions provide no efficiency gain when y_t is serially uncorrelated, whether or not the disturbances are homoskedastic. When both ρ and γ are nonzero, however, there will in general be asymptotic efficiency gains by adding moment conditions for multiple-step-ahead forecasts or, equivalently, by adding multiple lags of y_t . Hansen and Singleton (1990) analyze GMM estimators that use multiple lagged values of an instrument to estimate a class of models that, in terms of departures from the standard i.i.d. setting, might be viewed as the complement of the models considered here. Hansen and Singleton consider models in which the disturbances are homoskedastic but serially correlated, whereas the disturbances in (1.1) and (1.2) are serially uncorrelated but heteroskedastic.

Table 2 displays the ratio $\text{var}(\hat{\rho}_{(K)})/\text{var}(\hat{\rho}_{(1)})$ for various values of K , ρ , γ , and κ . (The ratio does not depend on μ_y or σ_y^2 .) In the context of the previously described example using the monthly (D/P) series, K would denote a number of months. Although the entries in table 2 indicate that the ratio $\text{var}(\hat{\rho}_{(K)})/\text{var}(\hat{\rho}_{(1)})$ is monotone in neither ρ nor γ , many of the lowest ratios (greatest efficiency gains) tend to occur when both of those parameters are high. In such cases, asymptotic efficiency gains continue as K is increased to fairly large values. For example, when $\rho = 0.97$, $\gamma = 0.9$, and $K = 48$, the variance of the OLS estimator exceeds that of $\hat{\rho}_{(K)}$ by 33% when $\kappa = 3$ and by 92% when $\kappa = 13$. Of interest for finance applications is that, as observed earlier, researchers often obtain similarly high values for sample estimates of both ρ and γ .

Increasing K to large values is not always required to achieve substantial asymptotic efficiency gains. By construction, $\text{var}(\hat{\rho}_{(K)})/\text{var}(\hat{\rho}_{(1)})$ cannot increase as moments are added, but there are cases where most of the gain is achieved with a modest value of K . For example, gains nearly equal to those noted above are obtained with $K = 6$ when $\rho = 0.9$ and $\gamma = 0.5$. For all of the cases considered in table 2, $\text{var}(\hat{\rho}_{(K)})/\text{var}(\hat{\rho}_{(1)})$ is decreasing in κ .

¹²See, for example, proposition 4.50 in White (1984).

2.3. Adding One Long-Horizon Moment Condition

The analysis in the previous subsection indicates that, although substantial reductions in the asymptotic variance of $\hat{\rho}_{(K)}$ relative to $\hat{\rho}_{(1)}$ are possible when both γ and ρ are high, achieving those gains can require rather large values of K , and thus a large number of moment conditions. We consider here an alternative GMM estimator obtained by adding only one moment condition to the two conditions used in computing the OLS estimator.

Let $\epsilon_{t+K,K}$ denote the forecast error for the sum of K future y 's:

$$\begin{aligned}\epsilon_{t+K,K} &= \sum_{k=1}^K y_{t+k} - \mathbf{E}_t\left\{\sum_{k=1}^K y_{t+k}\right\}, \\ &= \sum_{k=1}^K y_{t+k} - \alpha \left(\frac{K - (K+1)\rho + \rho^{K+1}}{(1-\rho)^2} \right) - \left(\frac{\rho - \rho^{K+1}}{1-\rho} \right) y_t.\end{aligned}\quad (2.13)$$

(When $K = 1$, the second subscript on ϵ is suppressed.) Construct the product

$$m_{K,t} = \epsilon_{t+K,K} \cdot y_t, \quad (2.14)$$

and define the three-element vector

$$g_{K,t}^* = \begin{bmatrix} g_{1,t} \\ m_{K,t} \end{bmatrix}. \quad (2.15)$$

Let $\theta_{(K)}^* = (\alpha_{(K)}^* \rho_{(K)}^*)'$ be the optimal GMM estimator of θ based on the moment conditions $\mathbf{E}\{g_{K,t}^*\} = 0$. The asymptotic covariance matrix for $\theta_{(K)}^*$ can be evaluated analytically. The expected value of the gradient of $g_{K,t}^*$ is given by

$$D_K^* = \mathbf{E} \left[\frac{\partial g_{K,t}^*}{\partial \theta'} \right] = \begin{bmatrix} D_1 \\ \mathbf{E} \left(\frac{\partial m_{K,t}}{\partial \alpha} \right) \quad \mathbf{E} \left(\frac{\partial m_{K,t}}{\partial \rho} \right) \end{bmatrix}, \quad (2.16)$$

where

$$\mathbf{E} \left(\frac{\partial m_{K,t}}{\partial \alpha} \right) = -\frac{\mu_y}{(1-\rho)^2} [K(1-\rho) - \rho(1-\rho^K)], \quad \text{and} \quad (2.17)$$

$$\begin{aligned}\mathbf{E} \left(\frac{\partial m_{K,t}}{\partial \rho} \right) &= -\frac{\mu_y^2}{(1-\rho)^2} [K(1-\rho) - \rho(1-\rho^K)] \\ &\quad -\frac{\sigma_y^2}{(1-\rho)^2} [1 - (K+1)\rho^K + K\rho^{K+1}].\end{aligned}\quad (2.18)$$

The multiperiod forecast error $\epsilon_{t+K,K}$ is a combination of the single-period forecast errors:

$$\epsilon_{t+K,K} = \sum_{k=1}^K \frac{1 - \rho^{K-k+1}}{1-\rho} \epsilon_{t+k}. \quad (2.19)$$

This implies that the three elements of $g_{K,t}^*$ can be written as linear combinations of the $K + 1$ elements of $g_{K,t}$, defined in (2.7). Specifically,

$$g_{K,t}^* = Q_K g_{K,t}, \quad (2.20)$$

where

$$Q_K = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & \frac{1-\rho^K}{1-\rho} & \frac{1-\rho^{K-1}}{1-\rho} & \cdots & 1 \end{bmatrix}. \quad (2.21)$$

Therefore,

$$S_K^* = \sum_{l=-\infty}^{\infty} E\{g_{K,t}^* g_{K,t-l}^{*'}\} = \sum_{l=-\infty}^{\infty} E\{Q_K g_{K,t} g_{K,t-l}' Q_K'\} = Q_K S_K Q_K', \quad (2.22)$$

and the asymptotic covariance matrix of $\theta_{(K)}^*$ is equal to $(D_K^{*'}(S_K^*)^{-1}D_K^*)^{-1}$.

Table 3 displays the ratio $\text{var}(\rho_{(K)}^*)/\text{var}(\hat{\rho}_{(1)})$ for the same parameter values used to construct table 2. Here, increasing K replaces the single additional moment condition, so $\text{var}(\rho_{(K)}^*)/\text{var}(\hat{\rho}_{(1)})$ will first decline and then, unlike the entries in table 2, eventually rise as K increases. Nevertheless, two rather striking observations emerge from table 3. First, there generally exists a value of K such that the efficiency gains for $\hat{\rho}_{(K)}$ reported in table 2 are nearly matched by $\rho_{(K)}^*$. Second, even when K differs from that “best” choice, the relative reductions in asymptotic variance produced by $\rho_{(K)}^*$ can still be substantial.

Consider again the case in which $\rho = 0.97$ and $\gamma = 0.9$. When 49 separate moment conditions are used ($K = 48$), the resulting estimator $\hat{\rho}_{(48)}$ has an asymptotic variance that is either 0.748 or 0.521 times the variance of the OLS estimator $\hat{\rho}_{(1)}$, depending on whether κ equals 3 or 13. Those ratios increase to only 0.751 and 0.528 for $\rho_{(48)}^*$, the estimator computed with three moment conditions—the two conditions used in OLS plus one condition for a 48-month forecasted sum of the y 's. In the case where $\rho = 0.9$ and $\gamma = 0.5$, also considered earlier, the asymptotic variance of $\hat{\rho}_{(12)}$ is 0.553 times that of $\hat{\rho}_{(1)}$ when $\kappa = 13$, and that ratio increases to only 0.565 for $\rho_{(12)}^*$. Choosing K smaller or larger than 12 still produces substantial efficiency gains: $\text{var}(\rho_{(6)}^*)/\text{var}(\hat{\rho}_{(1)}) = 0.574$ and $\text{var}(\rho_{(24)}^*)/\text{var}(\hat{\rho}_{(1)}) = 0.666$.

In general, it appears that a significant fraction of the asymptotic efficiency gains relative to OLS obtained with $\hat{\rho}_{(K)}$, the estimator computed using $K + 1$ moment conditions, are delivered by $\rho_{(K)}^*$, an estimator computed with three moment conditions. With both estimators, the gains relative to OLS increase with kurtosis (κ).

2.4. Weights on the Multiperiod Moment Conditions

Both $\hat{\theta}_{(K)}$ and $\theta_{(K)}^*$ are solutions to a two-equation system of the form $A\bar{g}_K(\theta) = 0$, where A is a $2 \times (K + 1)$ “weighting matrix” and $\bar{g}_K(\theta)$, a function of θ , is the sample analogue of $E\{g_{K,t}\}$. The choices of A differ for the two estimators, and the differences relevant to determining asymptotic variances lie in the relative weights applied to elements 3 through $K + 1$ of $\bar{g}_K(\theta)$, which are the sample analogues of the multiperiod moment conditions $E\{\epsilon_{t+j}y_t\}$, $j = 2, \dots, K$. The weights on those conditions used in computing $\theta_{(K)}^*$ are proportional to the corresponding elements in the third row of Q_K in (2.21). In that case, the weights depend only on ρ and are determined by the identity governing the multiperiod forecast error in (2.19). In computing $\hat{\theta}_{(K)}$, where the additional $K - 1$ moment conditions enter separately, the weighting matrix converges (for large T) to $A = D'_K S_K^{-1}$ [see Hansen (1982)]. It can be verified that only the second row of that matrix contains nonzero entries for the additional $K - 1$ moment conditions, which permits a straightforward comparison of the relative weighting schemes for the estimators.

Figure 1 displays several sets of weights on the sample analogues of the moment conditions $E\{\epsilon_{t+j}y_t\}$, $j = 2, \dots, K$, for the case $\rho = 0.9$, $\gamma = 0.5$, and $\kappa = 13$. The “optimal” weights used in computing $\hat{\theta}_{(K)}$ increase through $j = 4$ and then decline monotonically. These weights are shown for $K = 48$, although it is evident from the figure that essentially the same weights would result for any $K > 15$, since the weights are nearly zero beyond that horizon. The weights in the K -period sum used to compute $\theta_{(K)}^*$, shown in the figure for both $K = 12$ and $K = 48$, decline monotonically to zero over the range $j = 2, \dots, K + 1$. Both sets of multiperiod-sum weights depart from the optimal weights, but when $K = 12$ there is at least fairly close agreement with the optimal weights at the beginning and ending points ($j = 2$ and $j = 12$). The results in table 3 are consistent with this analysis, in that $\theta_{(12)}^*$ does substantially better than $\theta_{(48)}^*$ in matching the asymptotic efficiency gains achieved by $\hat{\theta}_{(K)}$ (for $K \geq 12$). The relevant ratios in table 3 are 0.565 for $\theta_{(12)}^*$ and 0.823 for $\theta_{(48)}^*$, as compared to 0.553 for $\hat{\theta}_{(K)}$ (table 2). Although a simple analytical expression for the optimal weights remains elusive, those weights typically appear to decline less rapidly when γ and ρ are both high. That behavior is consistent with the result in table 3 that, when $\rho = 0.97$ and $\gamma = 0.9$, the greatest efficiency gain for $\theta_{(K)}^*$ then occurs with $K = 48$.

3. The Two-Equation Case

The previous section addressed the estimation of the parameters in equation (1.1), the first-order autoregression for y_t . In finance applications, the parameters of equation (1.2), which

gives conditional expected single-period returns, are often of equal or greater interest. This section considers the estimation of a and b . As in the previous section, conditional heteroskedasticity in the disturbances ϵ_{t+1} and u_{t+1} remains the central focus.

In order to keep the problem tractable enough so that asymptotic variances of the estimators can still be calculated analytically and characterized by a small number of parameters, the conditional heteroskedasticity in u_{t+1} is assumed to be driven by that in ϵ_{t+1} . Specifically, the conditional variance-covariance matrix of the disturbances is assumed to obey

$$E_t \left\{ \begin{bmatrix} \epsilon_{t+1} \\ u_{t+1} \end{bmatrix} \begin{bmatrix} \epsilon_{t+1} & u_{t+1} \end{bmatrix} \right\} = \sigma_t^2 \Omega, \quad (3.1)$$

where Ω is a (constant) positive-definite matrix with its (1, 1) element set to unity, and σ_t^2 is still defined by (2.1). This simple specification of the conditional heteroskedasticity in the bivariate process for $(\epsilon_{t+1} \ u_{t+1})$ may no doubt be viewed as too restrictive for many applications. It is hoped that the analytical results obtained here can provide some guidance for subsequent investigations, perhaps involving Monte Carlo experiments, under more general specifications of heteroskedasticity.

3.1. Adding Multiple Moment Conditions

We first consider a GMM estimator for $\beta = (a \ b)'$ analogous to the estimator $\hat{\theta}_{(K)}$ defined in the previous section. That is, define

$$f_{K,t} = \begin{bmatrix} u_{t+1} \\ u_{t+1}y_t \\ u_{t+2}y_t \\ \vdots \\ u_{t+K}y_t \end{bmatrix}, \quad (3.2)$$

and let $\hat{\beta}_{(K)} = (\hat{a}_{(K)} \ \hat{b}_{(K)})'$ denote the optimal GMM estimator of β based on the $K + 1$ moment conditions $E\{f_{K,t}\} = 0$. As before, $\hat{\beta}_{(1)}$ is the OLS estimator.

Obtaining the asymptotic covariance matrix of $\hat{\beta}_{(K)}$ involves a simple extension of the analysis from the previous section. Define the $2(K + 1)$ -element vector

$$h_{K,t} = \begin{bmatrix} g_{K,t} \\ f_{K,t} \end{bmatrix}, \quad (3.3)$$

as well as the four-element parameter vector $\delta = (\alpha \ \rho \ a \ b)'$. It is straightforward to show that

$$E \left[\frac{\partial h_{K,t}}{\partial \delta'} \right] = I_2 \otimes D_K, \quad (3.4)$$

where I_2 denotes the 2×2 identity matrix and D_K is as defined in (2.10). In addition, the specification in (3.1) implies that

$$\sum_{l=-\infty}^{\infty} E\{h_{K,t}h'_{K,t-l}\} = \Omega \otimes S_K, \quad (3.5)$$

where S_K is as defined in (2.11). It follows from (3.4) and (3.5) that the asymptotic covariance matrix of $\hat{\beta}_{(K)}$ is given by

$$\text{var}(\hat{\beta}_{(K)}) = \omega_{2,2} \text{var}(\hat{\theta}_{(K)}), \quad (3.6)$$

where $\omega_{2,2}$ is the (2,2) element of Ω . Moreover, if $\hat{\delta}_{(K)}$ is the optimal GMM estimator of δ based on the $2(K+1)$ moment conditions $E\{h_{K,t}\} = 0$, it also follows from (3.4) and (3.5) that $\text{var}(\hat{\delta}_{(K)}) = \Omega \otimes \text{var}(\hat{\theta}_{(K)})$, so there is no asymptotic gain in efficiency relative to the estimator $(\hat{\theta}'_{(K)} \hat{\beta}'_{(K)})'$.

The immediate implications of (3.6) are that (i) each value in table 1 is also the ratio of $\hat{b}_{(1)}$'s variance under homoskedasticity to its variance under heteroskedasticity and (ii) each value in table 2 is also the ratio $\text{var}(\hat{b}_{(K)})/\text{var}(\hat{b}_{(1)})$. Obviously, the same observations offered earlier about those tables apply here as well. Note also that $\text{var}(\hat{\beta}_{(K)})$ does not depend on β .

3.2. Adding Two Long-Horizon Moment Conditions

Proceeding as before, we now consider an estimator obtained by combining the two moment conditions used in OLS with moment conditions corresponding to multiperiod forecast errors for sums of future y_t 's and x_t 's. Equations (1.1) and (1.2) constitute a first-order vector autoregression (VAR) for y_t and x_t that can be used to obtain a forecast of a sum of K future x_t 's—a multiperiod return in the finance application described earlier. If $u_{t+K,K}$ denotes the error in such a forecast, one can show that

$$\begin{aligned} u_{t+K,K} &= \sum_{k=1}^K x_{t+k} - E_t\left\{\sum_{k=1}^K x_{t+k}\right\} \\ &= \sum_{k=1}^K x_{t+k} - Ka - \alpha b \left(\frac{K-1-K\rho+\rho^K}{(1-\rho)^2}\right) - b \left(\frac{1-\rho^K}{1-\rho}\right) y_t. \end{aligned} \quad (3.7)$$

(As before, the second subscript on u is suppressed when $K=1$.) Construct the product

$$n_{K,t} = u_{t+K,K} \cdot y_t, \quad (3.8)$$

and define the six-element vector

$$h_{K,t}^\dagger = \begin{bmatrix} g_{1,t} \\ m_{K,t} \\ f_{1,t} \\ n_{K,t} \end{bmatrix} = \begin{bmatrix} g_{K,t}^* \\ f_{1,t} \\ n_{K,t} \end{bmatrix}. \quad (3.9)$$

Let $\delta_{(K)}^\dagger = (\alpha_{(K)}^\dagger \ \rho_{(K)}^\dagger \ a_{(K)}^\dagger \ b_{(K)}^\dagger)'$ denote the optimal GMM estimator of δ based on the moment conditions $E\{h_{K,t}^\dagger\} = 0$.

The asymptotic covariance matrix of δ_K^\dagger can also be calculated analytically. The expected value of the gradient is given by

$$D_K^\dagger = E \left[\frac{\partial h_{K,t}^\dagger}{\partial \delta'} \right] = \begin{bmatrix} D_K^* & 0 \\ 0 & D_1 \\ E \left(\frac{\partial n_{K,t}}{\partial \alpha} \right) & E \left(\frac{\partial n_{K,t}}{\partial \rho} \right) & E \left(\frac{\partial n_{K,t}}{\partial a} \right) & E \left(\frac{\partial n_{K,t}}{\partial b} \right) \end{bmatrix}, \quad (3.10)$$

where

$$E \left(\frac{\partial n_{K,t}}{\partial \alpha} \right) = -\frac{b\mu_y}{(1-\rho)^2} [K(1-\rho) - 1 + \rho^K], \quad (3.11)$$

$$E \left(\frac{\partial n_{K,t}}{\partial \rho} \right) = -\frac{b\mu_y^2}{(1-\rho)^2} [K(1-\rho) - 1 + \rho^K] \\ -\frac{b\sigma_y^2}{(1-\rho)^2} [1 - K\rho^{K-1} + (K-1)\rho^K], \quad (3.12)$$

$$E \left(\frac{\partial n_{K,t}}{\partial a} \right) = -K\mu_y, \text{ and} \quad (3.13)$$

$$E \left(\frac{\partial n_{K,t}}{\partial b} \right) = -K\mu_y^2 - \sigma_y^2 \frac{1-\rho^K}{1-\rho}. \quad (3.14)$$

The multiperiod forecast error $u_{t+K,K}$ is a combination of the single-period forecast errors in both y_t and x_t ,

$$u_{t+K,K} = \sum_{k=1}^{K-1} \frac{b(1-\rho^{K-k})}{1-\rho} \epsilon_{t+k} + \sum_{k=1}^K u_{t+k}, \quad (3.15)$$

which implies that $n_{K,t}$ is a linear combination of the elements in $g_{K,t}$ and $f_{K,t}$. As observed earlier in (2.19), $m_{K,t}$ is a linear combination of the elements of $g_{K,t}$. Therefore, $h_{K,t}^\dagger$ can be written as

$$h_{K,t}^\dagger = P_K h_{K,t} \quad (3.16)$$

where

$$P_K = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & \frac{1-\rho^K}{1-\rho} & \frac{1-\rho^{K-1}}{1-\rho} & \dots & \frac{1-\rho^2}{1-\rho} & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & \frac{b(1-\rho^{K-1})}{1-\rho} & \frac{b(1-\rho^{K-2})}{1-\rho} & \dots & b & 0 & 0 & 1 & 1 & \dots & 1 \end{bmatrix} \quad (3.17)$$

Combining (3.5), (3.16), and (3.17), and applying the same approach as in (2.22), gives

$$S_K^\dagger = P_K(\Omega \otimes S_K)P_K'. \quad (3.18)$$

The asymptotic covariance matrix of $\delta_{(K)}^\dagger$ is then given by $(D_K^{\dagger'}(S_K^\dagger)^{-1}D_K^\dagger)^{-1}$.

Table 4 reports the ratio $\text{var}(b_{(K)}^\dagger)/\text{var}(\hat{b}_{(1)})$ for the same values of the model parameters used to construct tables 2 and 3. Constructing this and subsequent tables requires values for some additional parameters. The unconditional mean and standard deviation of y_t are specified as

$$\mu_y = 0.044 \text{ and} \quad (3.19)$$

$$\sigma_y = 0.014, \quad (3.20)$$

which are sample estimates obtained from the monthly series of (D/P) for the period 1/1927–12/1989. The unconditional mean and standard deviation of x_t are specified as

$$\mu_x = 0.010 \text{ and} \quad (3.21)$$

$$\sigma_x = 0.057, \quad (3.22)$$

and the unconditional correlation between ϵ_t and u_t is set to

$$\text{corr}\{\epsilon_t, u_t\} = -0.90. \quad (3.23)$$

The values in (3.21) and (3.22) are the sample moments of monthly returns on the value-weighted NYSE portfolio for the 1927–89 period, and the value in (3.23) is the sample correlation of the residuals obtained by an OLS fit of both (1.1) and (1.2) using D/P and the value-weighted NYSE return over that same period.

The asymptotic variance of $b_{(K)}^\dagger$ also depends on b , and table 4 displays results for both $b = 0.0$ and $b = 0.5$. The first value corresponds to the case of constant expected returns, seen as an interesting null hypothesis by many researchers in finance, whereas the second value is typical of sample estimates obtained in previous empirical studies investigating the relation between D/P and expected returns [e.g., Fama and French (1988)]. As indicated by the entries in table 4, however, the variance of $b_{(K)}^\dagger$ is virtually identical for both choices of b .

The values of $\text{var}(b_{(K)}^\dagger)/\text{var}(\hat{b}_{(1)})$ in table 4 indicate that adding two moment conditions to those used in computing the OLS estimator achieves many of the asymptotic efficiency gains produced by adding the $K - 1$ separate moment conditions used in computing $\hat{b}_{(K)}$. That is, many of the entries in table 4 are nearly as low as those in table 2. As was the

case with $\rho_{(K)}^*$, substantial gains occur over a fairly wide range for K . In the previously considered case where $\rho = 0.97$ and $\gamma = 0.90$, for example, $\text{var}(b_{(48)}^\dagger)/\text{var}(\hat{b}_{(1)}) = 0.53$ and $\text{var}(b_{(12)}^\dagger)/\text{var}(\hat{b}_{(1)}) = 0.69$. Thus, at least for the simple specification of heteroskedasticity considered here, it appears that approaches similar to those considered in the previous section for estimating the single-equation autoregression for y_t can produce similar reductions in asymptotic variance when estimating the parameters of the regression of x_{t+1} on y_t .

4. Estimating Long-Horizon Expectations

The previous sections address the estimation of the parameters in equations (1.1) and (1.2), which give conditional expectations one period ahead. In finance applications, interest often centers on estimating expectations for multiperiod horizons. Specifically, the parameters of interest are a_K and b_K in the regression

$$\sum_{k=1}^K x_{t+k} = a_K + b_K y_t + u_{t+K,K}, \quad (4.1)$$

where the dependent variable is the rate of return over a K -period investment horizon.

4.1. Long-Horizon Regressions

One approach to estimating $\beta_K = (a_K \ b_K)'$ is simply to regress $\sum_{k=1}^K x_{t+k}$ on y_t .¹³ In other words, if

$$w_{K,t} = \begin{bmatrix} u_{t+K,K} \\ u_{t+K,K} y_t \end{bmatrix}, \quad (4.2)$$

a long-horizon regression computes the GMM estimators of a_K and b_K based on the moment condition $E\{w_{K,t}\} = 0$.

If the long-horizon regression uses $w_{K,t}$ for each of T periods in the sample, then the successive observations of $\sum_{k=1}^K x_{t+k}$ overlap by $K - 1$ periods. Denote the long-horizon regression estimator obtained in this case as $\hat{\beta}_{K(REG,lap)}$. Let $\hat{\beta}_{K(REG,non)}$ denote the estimator obtained when $w_{K,t}$ is instead computed only every K periods, so that the multiperiod observations do not overlap.

To obtain the asymptotic covariance matrices of these regression estimators, first define the vector

$$\eta_{K,t} = \begin{pmatrix} 1 \\ y_t \end{pmatrix} \otimes (\epsilon_{t+1} \cdots \epsilon_{t+K} \ u_{t+1} \cdots u_{t+K})', \quad (4.3)$$

¹³See, for example, Fama and French (1988, 1989).

which is $4K \times 1$. It can be verified that

$$S_{\eta K} = \sum_{l=-\infty}^{\infty} E\{\eta_{K,t}\eta'_{K,t-l}\} = \sigma_\epsilon^2 \begin{bmatrix} \Omega \otimes (I \quad I') & \Omega \otimes (\mu_y I \quad I') \\ \Omega \otimes (\mu_y I \quad I') & \Omega \otimes \Phi \end{bmatrix}, \quad (4.4)$$

where Φ is defined as in (2.11) and (2.12). Next observe from (3.15) that

$$w_{K,t} = W_K \eta_{K,t}, \quad (4.5)$$

where

$$W_K = I_2 \otimes \begin{bmatrix} \frac{b(1-\rho^{K-1})}{1-\rho} & \frac{b(1-\rho^{K-2})}{1-\rho} & \dots & b & 0 & I' \end{bmatrix}. \quad (4.6)$$

Equations (4.4) and (4.5) imply that

$$S_{WK} = \sum_{l=-\infty}^{\infty} E\{w_{K,t}w'_{K,t-l}\} = W_K S_{\eta K} W'_K, \quad (4.7)$$

so the asymptotic covariance matrix of $\hat{\beta}_{K(REG,lap)}$ is given by $(D'_{WK} S_{WK}^{-1} D_{WK})^{-1}$, where

$$D_{WK} = E \left[\frac{\partial w_{K,t}}{\partial \beta'_K} \right] = \begin{bmatrix} 1 & \mu_y \\ \mu_y & \mu_y^2 + \sigma_y^2 \end{bmatrix}. \quad (4.8)$$

The asymptotic covariance matrix of $\hat{\beta}_{K(REG,non)}$, which does not use overlapping observations, is given by $K(D'_{WK}(S_{WK}^{-1})^{-1}D_{WK})^{-1}$, where $S_{WK} = W_K S_{\eta K} W'_K$ and

$$S_{\eta K} = E\{\eta_{K,t}\eta'_{K,t}\} = \sigma_\epsilon^2 \begin{bmatrix} \Omega \otimes I_K & \Omega \otimes \mu_y I_K \\ \Omega \otimes \mu_y I_K & \Omega \otimes \text{diag}(\Phi) \end{bmatrix}. \quad (4.9)$$

For the case in which $b = 0$ and the disturbances in (1.1) and (1.2) are homoskedastic, the ratio $\text{var}(\hat{b}_{K(REG,lap)})/\text{var}(\hat{b}_{K(REG,non)})$ is computed analytically by Boudoukh and Richardson (1993). They observe that, although this ratio equals $1/K$ when $\rho = 0$, indicating large asymptotic efficiency gains to using overlapping observations in that case, the efficiency gains are much smaller if ρ is high. If $K = 24$, for example, they report that $\text{var}(\hat{b}_{K(REG,lap)})/\text{var}(\hat{b}_{K(REG,non)})$ rises from 0.042 at $\rho = 0$ to 0.505 at $\rho = 0.9$. Table 5 reports $\text{var}(\hat{b}_{K(REG,lap)})/\text{var}(\hat{b}_{K(REG,non)})$ for the same parameter values used in table 4. In other words, the analytical computations of Boudoukh and Richardson, reproduced in the $\gamma = 0$ rows of the $b = 0$ panel in table 5, are extended here to include cases in which $b > 0$ and the model's disturbances are heteroskedastic.

The results in table 5 indicate that $\text{var}(\hat{b}_{K(REG,lap)})/\text{var}(\hat{b}_{K(REG,non)}) = 1/K$ when $\rho = 0$, whether or not b and γ are zero, but perhaps more interesting are the results regarding the asymptotic efficiency gains produced by overlapping observations when ρ is high. First, the

efficiency gains increase in γ . For example, when $b = 0$, $\rho = 0.97$, and $K = 48$, the ratio $\text{var}(\hat{b}_{K(REG,lap)})/\text{var}(\hat{b}_{K(REG,non)})$ equals 0.649 when $\gamma = 0$ but falls to 0.469 when $\gamma = 0.9$. In other words, the gains from overlapping observations are smallest in the homoskedastic case. Also observe that, in the homoskedastic case, the relative efficiency gains from overlapping are smaller when $b = 0$ than when $b = 0.5$. For example, again when $\rho = 0.97$ and $K = 48$, $\text{var}(\hat{b}_{K(REG,lap)})/\text{var}(\hat{b}_{K(REG,non)})$ equals 0.588 when $b = 0.5$, compared to 0.649. This latter difference disappears and even reverses slightly, however, as γ increases: when $b = 0.5$, $\text{var}(\hat{b}_{K(REG,lap)})/\text{var}(\hat{b}_{K(REG,non)})$ equals 0.484, compared to 0.469. In general, although the asymptotic efficiency gains obtained by overlapping observations are modest in any event when ρ is high as compared to when $\rho = 0$, it appears that such gains are smallest for the case of homoskedastic disturbances and $b = 0$.

4.2. The VAR Approach

An alternative approach to estimating $\beta_K = (a_K \ b_K)'$ is to substitute an estimator of $\delta = (\alpha \ \rho \ a \ b)'$ into the functions,

$$a_K = Ka + \alpha b \left(\frac{K - 1 - K\rho + \rho^K}{(1 - \rho)^2} \right), \quad (4.10)$$

$$b_K = b \left(\frac{1 - \rho^K}{1 - \rho} \right), \quad (4.11)$$

as given earlier on the right-hand side of equation (3.15). That is, the VAR approach estimates long-horizon expectations by using both (1.1) and (1.2).¹⁴ The alternative estimators of δ discussed in the previous section, when substituted into (4.10) and (4.11), give alternative VAR-implied estimators of β_K . Let $\hat{\beta}_{K(1,VAR)}$ denote the estimator based on the OLS estimator $\hat{\delta}_{(1)}$, and let $\beta_{K(J,VAR)}^\dagger$ denote the estimator based on the GMM estimator δ_J^\dagger .

The asymptotic covariance matrices of $\hat{\beta}_{K(1,VAR)}$ and $\beta_{K(J,VAR)}^\dagger$ can be obtained by using the first-derivative matrix,

$$D_{\beta_K} = \left[\frac{\partial \beta_K}{\partial \delta'} \right] = \left[\begin{array}{cccc} \frac{\partial a_K}{\partial \alpha} & \frac{\partial a_K}{\partial \rho} & \frac{\partial a_K}{\partial a} & \frac{\partial a_K}{\partial b} \\ \frac{\partial b_K}{\partial \alpha} & \frac{\partial b_K}{\partial \rho} & \frac{\partial b_K}{\partial a} & \frac{\partial b_K}{\partial b} \end{array} \right], \quad (4.12)$$

where $\partial a_K / \partial a = K$, $\partial b_K / \partial \alpha = \partial b_K / \partial a = 0$,

$$\frac{\partial a_K}{\partial \alpha} = b \left(\frac{K - 1 + \rho^K - K\rho}{(1 - \rho)^2} \right), \quad (4.13)$$

¹⁴Such an approach to constructing expected long-horizon returns is analyzed by Kandel and Stambaugh (1987, 1988), Campbell (1991), Hodrick (1992), and others.

$$\frac{\partial a_K}{\partial \rho} = \alpha b \left(\frac{K(\rho^{K-1} - 1)(1 - \rho) + 2(K - 1 + \rho^K - K\rho)}{(1 - \rho)^3} \right), \quad (4.14)$$

$$\frac{\partial a_K}{\partial b} = \alpha \left(\frac{K - 1 + \rho^K - K\rho}{(1 - \rho)^2} \right), \quad (4.15)$$

$$\frac{\partial b_K}{\partial \rho} = b \left(\frac{1 + (K - 1)\rho^K - K\rho^{K-1}}{(1 - \rho)^2} \right), \quad \text{and} \quad (4.16)$$

$$\frac{\partial b_K}{\partial b} = \frac{1 - \rho^K}{1 - \rho}. \quad (4.17)$$

The asymptotic covariance matrix of $\hat{\delta}_{(1)}$ is equal to $\Omega \otimes (D_1' S_1^{-1} D_1)^{-1}$, by (3.4) and (3.5). Therefore, the asymptotic covariance matrix of $\hat{\beta}_{K(1,VAR)}$ is $D_{\beta K} [\Omega \otimes (D_1' S_1^{-1} D_1)^{-1}] D_{\beta K}'$. Similarly, since the asymptotic covariance matrix of δ_J^\dagger is equal to $(D_J^\dagger (S_J^\dagger)^{-1} D_J^\dagger)^{-1}$, as shown in the previous section, the asymptotic covariance matrix of $\beta_{K(J,VAR)}^\dagger$ is equal to $D_{\beta K} [(D_J^\dagger (S_J^\dagger)^{-1} D_J^\dagger)^{-1}] D_{\beta K}'$. The analysis below sets J , the horizon used in constructing the additional VAR moments, equal to K , the multiperiod horizon in (4.1). This choice for J , although perhaps somewhat natural, is made purely to contain the scope of this study; other choices for J could prove more desirable.

We first compare the two approaches to estimating b_K that have been analyzed in a number of previous studies: (i) computing an implied b_K as a function of the OLS estimators of the single-period VAR parameters versus (ii) regressing $\sum_{k=1}^K x_{t+k}$ on y_t using overlapping observations. Table 6 reports the ratio $\text{var}(\hat{b}_{K(1,VAR)})/\text{var}(\hat{b}_{K(REG,lap)})$ for the same parameter values used in tables 4 and 5. Boudoukh and Richardson (1993) compute this ratio when $b = 0$ and the disturbances are homoskedastic, and their results are reproduced in the $\gamma = 0$ rows of the $b = 0$ panel in table 6. When both b and γ are zero, the ratio of variances equals $1/K$ at $\rho = 0$ and then approaches unity as ρ increases. In other words, the OLS-VAR approach always enjoys some efficiency advantage over the long-horizon-regression approach in this case, but, as Boudoukh and Richardson observe, that advantage grows small as ρ approaches unity. Table 6 indicates that the same statement can be made when $b = 0.5$, except that the efficiency advantage of the OLS-VAR approach is slightly greater.

The most striking result in table 6 is that persistent heteroskedasticity can render the OLS-VAR approach decidedly inferior to the regression approach, especially when ρ is high. Consider again, for example, the case where $b = 0.5$, $\rho = 0.97$, and $K = 48$. When γ goes from 0 to 0.9, the variance of the OLS-VAR estimator of b_K goes from 65% less than that of the regression estimator to 62% more. Another dramatic reversal occurs in the case where $b = 0.5$, $\rho = 0.9$, and $K = 12$: as γ goes from 0 to 0.9, the variance of the OLS-VAR estimator goes from 52% less to 73% more than that of the regression estimator, although this case demonstrates that the ratio of variances is not monotonic in γ . Similar results

hold for $b = 0$. In general, the apparent asymptotic efficiency advantages of the OLS-VAR approach derived under homoskedasticity are not robust to violations of that assumption.¹⁵

Table 7 reports the ratio $\text{var}(b_{K(K,VAR)}^\dagger)/\text{var}(\hat{b}_{K(REG,lap)})$. In other words, table 7 reports the same VAR-versus-regression comparison as table 6, except that the VAR approach replaces the OLS estimator of δ with the GMM estimator based on the two additional moment conditions. For the parameter values considered, the long-horizon-regression estimator of b_K never has an asymptotic variance less than that of the VAR-based estimator. That is, none of the entries in table 7 exceeds unity, although in several cases with $K = 2$ the ratio is equal to 1.000 (to that many decimal places). In the case considered previously, where $b = 0.5$, $\rho = 0.97$, and $K = 48$, the ratio of variances remains below unity as γ increases, although the ratio rises to 0.892 at $\gamma = 0.9$. The VAR estimator of b_K using the additional moments provides only a slight asymptotic efficiency advantage over the regression estimator when both ρ and γ are high, but this advantage, unlike that of the OLS-VAR estimator, is not reversed by the presence of autocorrelated conditional heteroskedasticity.

4.3. Approximate Slopes

Much empirical research in finance investigates the hypothesis that rates of return for all investment horizons are unpredictable. In the context of the model used here, that null hypothesis corresponds to the restriction $b = b_K = 0$ for all K . The powers of various tests to reject that null hypotheses can be analyzed using the asymptotic variances of the parameter estimators computed above. The vehicle for the analysis is “approximate slope,” a measure introduced by Bahadur (1960) and analyzed subsequently by Geweke (1981). Consider the powers of statistics z_1 and z_2 to reject the null in the presence of a fixed alternative, and let n_1 and n_2 denote the numbers of observations required for each statistic to reject with a given probability at significance level s . Geweke shows that, as T grows large and s is decreased so as to hold constant the probability of rejection, n_1/n_2 approaches c_2/c_1 , where c_1 and c_2 denote the statistics’ approximate slopes. If c_1 is twice c_2 , then asymptotically, z_1 requires only half as many observations as z_2 to reject the null with a given probability.

For any statistic distributed χ^2 under the null, Geweke (1981) also shows that the approximate slope is the probability limit of $1/T$ times the statistic under the alternative. When the null is that a single parameter p is zero, and if the estimator \hat{p} is asymptotically normal, then this probability limit is simply $c = p_\lambda^2/(T\text{var}\{\hat{p}\})$, where the variance is calculated

¹⁵The data-generating process used by Hodrick (1992) in his Monte Carlo experiments assumes a GARCH specification for heteroskedasticity. Hodrick considers both the OLS-VAR estimator and the long-horizon regression estimator, but he does not directly compare their variances.

under the alternative $p = p_A$. It follows that the ratios of asymptotic variances reported in tables 2 through 7 can be interpreted as reciprocals of asymptotic slopes under a variety of alternatives. That is, when the parameter examined is the same for both tests, which is the case in those tables, then the ratio of approximate slopes simplifies to the reciprocal of the ratio of the variances of the estimators.

Campbell (1992) analyzes the powers of various regression-based tests of predictability.¹⁶ He finds that statistics based on long-horizon regressions with overlapping observations can possess approximate slopes greater than tests based on single-period regressions. Campbell calculates approximate slopes under an assumption of homoskedasticity. The asymptotic variances computed in this study can be used to investigate the properties of the same approximate slopes under heteroskedasticity. The approximate slope for the test based on the single-period regression is given by $c(1) = b^2/(T\text{var}\{\hat{b}_{(1)}\})$. In the long-horizon regression, the approximate slope is $c(K) = b_K^2/(T\text{var}\{\hat{b}_{K(REG,lap)}\})$, where b_K is given by (4.11), and the variance in the denominator is calculated under that alternative. Note that $c(K)/c(1)$ does not simplify to a ratio of variances, since the two tests examine different parameters, b_K versus b . As Campbell observes, the long-horizon regression can achieve asymptotic power advantages if, as K increases, b_K^2 grows faster than $\text{var}\{\hat{b}_{K(REG,lap)}\}$.

Table 8 reports the ratio of approximate slopes when the alternative hypothesis is given by $b = 0.5$. In essence, the two values of b used previously in tables 4 through 7 are used here to specify the null and the alternative. The “ $\gamma = 0$ ” rows correspond to the homoskedastic setting analyzed by Campbell (1992). The entries in those rows confirm that long-horizon regressions can indeed possess asymptotic power advantages when ρ is high, although the ratios of approximate slopes exceed unity by only modest amounts for the horizons and parameter values used here. Campbell obtains higher ratios by including somewhat different parameter settings and by entertaining horizons substantially longer than 48 months. Table 8 reveals that higher ratios can also arise in the presence of heteroskedasticity. Most dramatic is the case of the 48-month regression when $\rho = 0.97$. When $\kappa = 13$, the ratio of approximate slopes goes from 1.14 to 5.25 as γ increases from 0 to 0.9. Substantial increases occur in other cases with much shorter horizons. When $\rho = 0.9$, for example, the ratio of approximate slopes in the 6-month regression goes from 0.92 to 2.30 as γ increases from 0 to 0.5. The latter case also reveals that the ratio of approximate slopes is not monotonic in γ .

Given that approximate slope is an asymptotic construct, the power advantages it ap-

¹⁶Jegadeesh (1991) and Richardson and Smith (1991, 1992) also use approximate slope to analyze power in regressions of returns on lagged returns.

pears to indicate may prove illusory in finite samples.¹⁷ Monte Carlo experiments can reveal the extent to which comparisons of approximate slopes are indicative of true power advantages, but the results in table 8 suggest that such investigations may depend importantly on the assumptions made regarding heteroskedasticity. Again, while an asymptotic analysis cannot supersede a Monte Carlo investigation, perhaps it can suggest important dimensions for designing such experiments.

5. Conclusion

Modeling conditional volatility as a GARCH(1,1) process permits the analytical calculation of asymptotic variances of parameter estimators in models of conditional expectations. Those calculations reveal that heteroskedasticity plays an important role in determining the large-sample behavior of such estimators. In estimating models for single-period-ahead conditional expectations, GMM estimators using moment conditions for multiperiod forecast horizons offer no large-sample efficiency gains under homoskedasticity. Under heteroskedasticity, however, such estimators can possess asymptotic variances that are substantially lower than those of OLS estimators, especially when both the conditional mean and the conditional variance are persistent. The addition of just one or two moment conditions based on forecasts of multiperiod sums can deliver asymptotic efficiency gains relative to OLS that are, in many cases, similar to those obtained by adding a larger number of moments based on multiperiod forecasts.

Heteroskedasticity also affects comparisons of alternative approaches to estimating models for conditional expectations of multiperiod quantities, such as long-horizon rates of return. In that application, heteroskedasticity increases somewhat the advantage of overlapping the observations in regressions of long-horizon returns on a predetermined variable, and, in sharp contrast to the homoskedastic case, heteroskedasticity can provide such regressions with an efficiency advantage over an alternative VAR-based approach considered in previous studies. Replacing OLS estimators with GMM estimators that use additional multiperiod moment conditions can restore some efficiency advantage to the VAR approach, but that advantage is minimal when the conditional mean and the conditional variance are both persistent. Campbell (1992) argues that, when compared to single-period regressions, long-horizon regressions possess greater power to reject the null hypothesis of constant expectations in favor of alternative hypotheses in which the conditional expectation is persistent. This apparent power advantage, which is computed asymptotically using approximate slope, can increase

¹⁷James Stock emphasized this point in a discussion of Campbell's paper, and Campbell also reports Monte Carlo experiments conducted to address this concern.

dramatically in the presence of heteroskedasticity.

Whether the asymptotic properties reported here serve as useful guideposts to finite-sample behavior remains to be investigated. The Monte Carlo experiments of Boudoukh and Richardson (1993), conducted in a homoskedastic setting, indicate that ratios of finite-sample variances of several of the estimators considered above in section 4 are approximated reasonably well by their asymptotic counterparts.¹⁸ In other words, the finite-sample efficiency of one estimator relative to another is, in many cases, close to the ratio of asymptotic variances. At the same time, however, the asymptotic variance of either estimator might prove less useful in conducting hypothesis tests. Monte Carlo simulations by Richardson and Stock (1989) and Hodrick (1992) reveal that standard asymptotic distributions often provide poor approximations to the sizes of tests in finite samples, especially for methods employing multiperiod forecasts in samples of sizes typically encountered in economics and finance applications.¹⁹ Although the asymptotic results presented in this study are confined largely to comparisons of variances, Monte Carlo experiments could clearly address a wider range of issues. As stated at the outset, it is hoped that the asymptotic results can aid in the design of such experiments.

Since the GMM estimators using additional multiperiod moment conditions produce overidentified models, the method for estimating the covariance matrix of the moment conditions becomes an important choice in evaluating the finite-sample behavior of such estimators. The asymptotic results in this study require only that the estimator of the covariance matrix be consistent in the presence of heteroskedasticity and autocorrelation, but the finite-sample properties of the estimators in the model of conditional expectations may be depend significantly on the choice of the covariance-matrix estimator. One approach would be to employ an estimator that accomodates heteroskedasticity and autocorrelation of unknown form.²⁰ Another approach might make use of the implied structure of the large-sample covariance matrix by specifying and estimating some features of the process governing heteroskedasticity. In the AR(1)-GARCH (1,1) model considered here, for example, the large-sample covariance matrix S_K depends on σ_ϵ^2 , γ , and κ (in addition to the parameters of the model for conditional expectations).

¹⁸For example, they find that the asymptotic ratio $\text{var}(\hat{b}_{K(REG,isp)})/\text{var}(\hat{b}_{K(REG,non)})$ is close to that obtained in samples of size $T = 720$ for $K \leq 60$.

¹⁹In fact, Richardson and Stock (1989) find that a nonstandard asymptotic approach, wherein the forecast horizon tends to a fixed fraction of the sample size (as opposed to zero in the standard approach used here and elsewhere), provides a closer approximation to finite-sample properties.

²⁰Examples include the estimators in Hansen (1982) and Newey and West (1987). Andrews (1991) investigates the asymptotic and finite-sample properties of a wide class of such estimators. Mishkin (1990) also provides evidence on the finite-sample behavior of the Newey-West estimator.

Also of interest are the potential effects of model misspecification on both the asymptotic and finite-sample properties of the various estimators considered here. Boudoukh and Richardson (1993) argue that the VAR estimator of expected long-horizon returns is likely to be more sensitive to misspecification, since, unlike the long-horizon regression, that estimator relies on the assumed dynamic behavior of the single-period expected return. A similar concern arises with the GMM estimators based on additional multiperiod moment conditions, since those conditions rely on the same assumption.

Appendix

This appendix outlines the derivation of (2.11) and (2.12). The first step is to obtain a simple expression for

$$\lambda(i) = E\{\epsilon_{t+i}^2, \epsilon_t^2\}, \quad i = 0, 1, \dots \quad (\text{A.1})$$

Define

$$\nu_{t+1} = \epsilon_{t+1}^2 - \sigma_t^2, \quad (\text{A.2})$$

so that $E_t\{\nu_{t+1}\} = 0$. Use (2.1) and (A.2) to write

$$\epsilon_{t+i+1}^2 = \gamma_0 + \gamma_1 \epsilon_{t+i}^2 - \gamma_1 \nu_{t+i} + \gamma_2 \epsilon_{t+i}^2 + \nu_{t+i+1}, \quad (\text{A.3})$$

for $i > 0$. Multiplying both sides of (A.3) by ϵ_t^2 and then taking expectations gives

$$\lambda(i+1) = \gamma_0 \sigma_t^2 + \gamma \lambda(i). \quad (\text{A.4})$$

The solution to this difference equation is

$$\lambda(i) = \frac{\gamma_0 \sigma_t^2}{1-\gamma} + [\lambda(0) - \frac{\gamma_0 \sigma_t^2}{1-\gamma}] \gamma^i, \quad (\text{A.5})$$

where, using (2.5),

$$\lambda(0) = E\{\epsilon_t^4\} = \sigma_t^4(\kappa + 3). \quad (\text{A.6})$$

It is easily verified that

$$\gamma_0 = \sigma_t^2(1-\gamma), \quad (\text{A.7})$$

so equation (A.5) can be rewritten as

$$\lambda(i) = \sigma_t^4[1 + (\kappa + 2)\gamma^i]. \quad (\text{A.8})$$

The next step is to obtain an expression for

$$\psi(i) = E\{\epsilon_{t+i}^2 y_t^2\}, \quad (\text{A.9})$$

where $i > 0$. Squaring both sides of (1.1), multiplying through by ϵ_{t+i}^2 , and then taking expectations gives

$$\begin{aligned} \psi(i) &= \rho^2 \psi(i+1) + \alpha^2 \sigma_t^2 + \lambda(i) + 2\alpha\rho\mu_y \sigma_t^2 \\ &= \rho^2 \psi(i+1) + \mu_y^2 \sigma_t^2 (1-\rho^2) + \lambda(i), \end{aligned} \quad (\text{A.10})$$

where the first line makes use of the property that $E_t\{\epsilon_{t+i}^2\epsilon_{t+1}\} = 0$ for $i \geq 0$ [implied by (2.1) and (2.4)], and the second line simplifies using $\alpha = \mu_y(1 - \rho)$. Multiplying both sides of (A.3) by y_t^l and then taking expectations gives

$$\begin{aligned}\psi(i+1) &= \gamma_0(\mu_y^2 + \sigma_y^2) + \gamma\psi(i) \\ &= (1 - \gamma)\sigma_\epsilon^2(\mu_y^2 + \sigma_y^2) + \gamma\psi(i).\end{aligned}\quad (\text{A.11})$$

Solving (A.10) and (A.11), substituting (A.8) for $\lambda(i)$, gives

$$\psi(i) = \sigma_\epsilon^2 \left[\mu_y^2 + \sigma_y^2 + \gamma^i \frac{\sigma_\epsilon^2(\kappa + 2)}{1 - \rho^2\gamma} \right]. \quad (\text{A.12})$$

Now observe that, for both k and l greater than zero,

$$\epsilon_{t+k}^2 y_t y_{t-l} = \epsilon_{t+k}^2 \rho^l y_{t-l}^2 + \epsilon_{t+k}^2 \mu_y (1 - \rho^l) y_{t-l} + \epsilon_{t+k}^2 \sum_{j=0}^{l-1} \rho^j \epsilon_{t-j} y_{t-l}, \quad (\text{A.13})$$

and that the expectation of the third term on the right-hand side is zero. Therefore,

$$E\{\epsilon_{t+k}^2 y_t y_{t-l}\} = \rho^l \psi(k+l) + \sigma_\epsilon^2 \mu_y^2 (1 - \rho^l) \quad \text{for } k > 0 \text{ and } l \geq 0. \quad (\text{A.14})$$

From (2.8) and the definition of $g_{K,t}$ in (2.7), the (i, j) element of Φ in (2.11) obeys

$$\sigma_\epsilon^2 \phi(i, j) = E\{(\epsilon_{t+i} y_t)(\epsilon_{t-l+j} y_{t-l})\} \quad \text{for } l = j - i, \quad (\text{A.15})$$

since the above expectation is zero for all other values of l . The expression on the right-hand side of (A.15) corresponds to (A.14) with $k = i$ and $l = j - i$ when $i \leq j$. When $j < i$, then the correspondence occurs with $k = j$ and $l = i - j$. In both cases, $k + l = \max(i, j)$ and $l = |i - j|$. Thus, using (A.14) and (A.12),

$$\begin{aligned}\sigma_\epsilon^2 \phi(i, j) &= \rho^{|i-j|} \psi(\max(i, j)) + \sigma_\epsilon^2 \mu_y^2 (1 - \rho^{|i-j|}) \\ &= \sigma_\epsilon^2 \left\{ \mu_y^2 + \rho^{|i-j|} \sigma_y^2 + \rho^{|i-j|} \gamma^{\max(i, j)} \left[\frac{\sigma_\epsilon^2(\kappa + 2)}{1 - \rho^2\gamma} \right] \right\} \\ &= \sigma_y^2 (1 - \rho^2) \left\{ \mu_y^2 + \rho^{|i-j|} \sigma_y^2 \left(1 + \gamma^{\max(i, j)} \left[\frac{(1 - \rho^2)(\kappa + 2)}{1 - \rho^2\gamma} \right] \right) \right\},\end{aligned}\quad (\text{A.16})$$

which verifies (2.12).

The right-hand side of equation (2.6) is the $(2, 2)$ element of $(D_1' S_1^{-1} D_1)^{-1}$. The elements of this matrix are easily computed analytically using (2.10) through (2.12) for $K = 1$.

Table 1

Effect of Heteroskedasticity on the Variance of the Sample Autocorrelation

Each entry is the ratio of the asymptotic variance of $\hat{\rho}_{(1)}$ under homoskedasticity to the asymptotic variance of $\hat{\rho}_{(1)}$ when the disturbances follow a GARCH (1,1) process, where $\hat{\rho}_{(1)}$ is the OLS estimator of the first-order autocorrelation (ρ .)

γ	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.97$
$\kappa = 1$					
0.0	1.000	1.000	1.000	1.000	1.000
0.3	0.526	0.543	0.650	0.816	0.931
0.5	0.400	0.412	0.497	0.676	0.857
0.7	0.323	0.329	0.380	0.520	0.733
0.9	0.270	0.272	0.289	0.346	0.490
$\kappa = 3$					
0.0	1.000	1.000	1.000	1.000	1.000
0.3	0.400	0.416	0.527	0.726	0.890
0.5	0.286	0.296	0.372	0.556	0.782
0.7	0.222	0.227	0.269	0.394	0.623
0.9	0.182	0.183	0.196	0.241	0.365
$\kappa = 8$					
0.0	1.000	1.000	1.000	1.000	1.000
0.3	0.250	0.263	0.358	0.570	0.802
0.5	0.167	0.173	0.228	0.385	0.642
0.7	0.125	0.128	0.155	0.246	0.452
0.9	0.100	0.101	0.109	0.137	0.224
$\kappa = 13$					
0.0	1.000	1.000	1.000	1.000	1.000
0.3	0.182	0.192	0.271	0.470	0.730
0.5	0.118	0.123	0.165	0.295	0.544
0.7	0.087	0.089	0.109	0.178	0.355
0.9	0.069	0.070	0.075	0.096	0.161

Table 2

OLS Estimator of ρ Versus GMM Estimator with $K-1$ Additional Moment Conditions

Each entry is the ratio $\text{var}(\hat{\rho}_{(K)})/\text{var}(\hat{\rho}_{(1)})$, where $\hat{\rho}_{(1)}$ is the OLS estimator and $\hat{\rho}_{(K)}$ is the GMM estimator obtained by adding $K-1$ moment conditions to those used in computing the OLS estimator.

K	γ	$\kappa = 3$					$\kappa = 13$				
		$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.97$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.97$
2	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.974	0.893	0.904	0.956	1.000	0.935	0.745	0.734	0.851
	0.5	1.000	0.982	0.904	0.878	0.928	1.000	0.969	0.830	0.753	0.812
	0.7	1.000	0.993	0.953	0.912	0.924	1.000	0.990	0.931	0.856	0.847
	0.9	1.000	0.999	0.994	0.980	0.967	1.000	0.999	0.992	0.973	0.950
6	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.973	0.880	0.888	0.949	1.000	0.931	0.684	0.666	0.813
	0.5	1.000	0.981	0.865	0.809	0.886	1.000	0.965	0.715	0.555	0.670
	0.7	1.000	0.993	0.917	0.788	0.814	1.000	0.989	0.857	0.601	0.583
	0.9	1.000	0.999	0.988	0.933	0.869	1.000	0.999	0.983	0.903	0.794
12	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.973	0.880	0.888	0.949	1.000	0.931	0.684	0.666	0.813
	0.5	1.000	0.981	0.865	0.809	0.886	1.000	0.965	0.715	0.553	0.668
	0.7	1.000	0.993	0.916	0.777	0.801	1.000	0.989	0.853	0.554	0.534
	0.9	1.000	0.999	0.987	0.909	0.793	1.000	0.999	0.983	0.859	0.650
24	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.973	0.880	0.888	0.949	1.000	0.931	0.684	0.666	0.813
	0.5	1.000	0.981	0.865	0.809	0.886	1.000	0.965	0.715	0.553	0.668
	0.7	1.000	0.993	0.916	0.777	0.801	1.000	0.989	0.853	0.554	0.533
	0.9	1.000	0.999	0.987	0.903	0.752	1.000	0.999	0.983	0.841	0.541
48	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.973	0.880	0.888	0.949	1.000	0.931	0.684	0.666	0.813
	0.5	1.000	0.981	0.865	0.809	0.886	1.000	0.965	0.715	0.553	0.668
	0.7	1.000	0.993	0.916	0.777	0.801	1.000	0.989	0.853	0.554	0.533
	0.9	1.000	0.999	0.987	0.903	0.748	1.000	0.999	0.983	0.840	0.521

Table 3

OLS Estimator of ρ Versus GMM Estimator with One Additional Moment Condition

Each entry is the ratio $\text{var}(\hat{\rho}_{(K)}^*)/\text{var}(\hat{\rho}_{(1)})$, where $\hat{\rho}_{(1)}$ is the OLS estimator and $\hat{\rho}_{(K)}^*$ is the GMM estimator obtained by adding one long-horizon moment condition to those used in computing the OLS estimator.

K	γ	$\kappa = 3$					$\kappa = 13$				
		$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.97$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.97$
2	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.974	0.893	0.904	0.956	1.000	0.935	0.745	0.734	0.851
	0.5	1.000	0.982	0.904	0.878	0.928	1.000	0.969	0.830	0.753	0.812
	0.7	1.000	0.993	0.953	0.912	0.924	1.000	0.990	0.931	0.856	0.847
	0.9	1.000	0.999	0.994	0.980	0.967	1.000	0.999	0.992	0.973	0.950
6	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.989	0.903	0.895	0.950	1.000	0.961	0.705	0.668	0.814
	0.5	1.000	0.989	0.870	0.810	0.887	1.000	0.976	0.715	0.574	0.687
	0.7	1.000	0.996	0.917	0.808	0.837	1.000	0.993	0.860	0.666	0.658
	0.9	1.000	1.000	0.988	0.945	0.904	1.000	0.999	0.984	0.923	0.853
12	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.995	0.951	0.925	0.961	1.000	0.982	0.817	0.721	0.832
	0.5	1.000	0.995	0.919	0.831	0.892	1.000	0.986	0.773	0.565	0.671
	0.7	1.000	0.998	0.934	0.778	0.803	1.000	0.995	0.863	0.567	0.561
	0.9	1.000	1.000	0.990	0.917	0.839	1.000	1.000	0.985	0.877	0.746
24	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.998	0.981	0.960	0.975	1.000	0.991	0.921	0.822	0.872
	0.5	1.000	0.998	0.966	0.895	0.918	1.000	0.992	0.885	0.666	0.709
	0.7	1.000	0.999	0.967	0.825	0.814	1.000	0.997	0.912	0.584	0.538
	0.9	1.000	1.000	0.994	0.904	0.771	1.000	1.000	0.990	0.843	0.616
48	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.999	0.992	0.985	0.987	1.000	0.996	0.965	0.920	0.921
	0.5	1.000	0.999	0.986	0.955	0.950	1.000	0.996	0.946	0.823	0.787
	0.7	1.000	0.999	0.985	0.913	0.866	1.000	0.998	0.953	0.734	0.603
	0.9	1.000	1.000	0.997	0.931	0.751	1.000	1.000	0.993	0.859	0.528

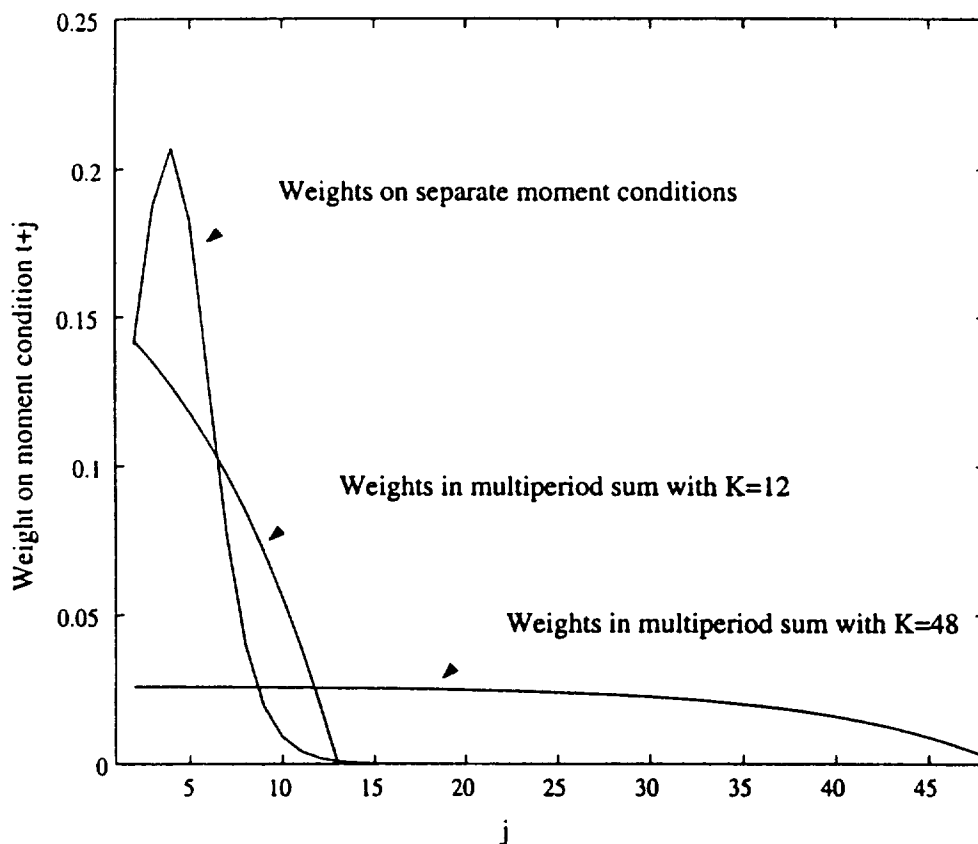


Figure 1. Weights on the moment conditions in computing $\hat{\rho}_{(K)}$, the estimator that uses $K - 1$ additional separate moment conditions, versus the weights implicit in the multiperiod sum used to obtain the one additional moment condition used in computing $\rho_{(K)}^*$. The figure plots the weights on the sample analogues of $E\{\epsilon_{t+j}y_t\}$, $j = 2, \dots, 48$, when $\rho = 0.9$, $\gamma = 0.5$, and $\kappa = 13$. All weights are normalized to sum to unity.

Table 4

OLS Estimator of b Versus GMM Estimator with Two Additional Moment Conditions

Each entry is the ratio $\text{var}(b_{(K)}^{\dagger})/\text{var}(b_{(1)})$, where $b_{(1)}$ is the OLS estimator and $b_{(K)}^{\dagger}$ is the GMM estimator obtained by adding two long-horizon moment conditions to those used in computing the OLS estimator.

K	γ	$\kappa = 3$					$\kappa = 13$				
		$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.97$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.97$
$b = 0.0$											
2	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.974	0.893	0.904	0.956	1.000	0.935	0.745	0.734	0.851
	0.5	1.000	0.982	0.904	0.878	0.928	1.000	0.969	0.830	0.753	0.812
	0.7	1.000	0.993	0.953	0.912	0.924	1.000	0.990	0.931	0.856	0.847
	0.9	1.000	0.999	0.994	0.980	0.967	1.000	0.999	0.992	0.973	0.950
12	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.996	0.959	0.937	0.966	1.000	0.983	0.844	0.755	0.849
	0.5	1.000	0.995	0.931	0.852	0.902	1.000	0.986	0.799	0.594	0.684
	0.7	1.000	0.998	0.942	0.787	0.803	1.000	0.995	0.873	0.555	0.537
	0.9	1.000	1.000	0.991	0.910	0.809	1.000	1.000	0.986	0.863	0.691
24	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.998	0.984	0.968	0.979	1.000	0.992	0.930	0.851	0.892
	0.5	1.000	0.998	0.970	0.913	0.931	1.000	0.992	0.896	0.713	0.741
	0.7	1.000	0.999	0.971	0.850	0.834	1.000	0.997	0.920	0.623	0.560
	0.9	1.000	1.000	0.995	0.910	0.754	1.000	1.000	0.991	0.843	0.563
48	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.999	0.993	0.987	0.989	1.000	0.996	0.967	0.933	0.935
	0.5	1.000	0.999	0.986	0.962	0.959	1.000	0.996	0.949	0.849	0.821
	0.7	1.000	0.999	0.986	0.927	0.888	1.000	0.998	0.955	0.770	0.651
	0.9	1.000	1.000	0.997	0.940	0.769	1.000	1.000	0.994	0.871	0.527
$b = 0.5$											
2	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.974	0.893	0.904	0.956	1.000	0.935	0.745	0.734	0.851
	0.5	1.000	0.982	0.904	0.878	0.928	1.000	0.969	0.830	0.753	0.812
	0.7	1.000	0.993	0.953	0.912	0.924	1.000	0.990	0.931	0.856	0.847
	0.9	1.000	0.999	0.994	0.980	0.967	1.000	0.999	0.992	0.973	0.950
12	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.996	0.959	0.937	0.967	1.000	0.983	0.845	0.756	0.850
	0.5	1.000	0.995	0.932	0.853	0.902	1.000	0.986	0.801	0.596	0.684
	0.7	1.000	0.998	0.942	0.788	0.803	1.000	0.995	0.874	0.555	0.537
	0.9	1.000	1.000	0.991	0.910	0.809	1.000	1.000	0.987	0.863	0.689
24	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.998	0.984	0.968	0.979	1.000	0.992	0.930	0.852	0.892
	0.5	1.000	0.998	0.970	0.914	0.931	1.000	0.992	0.898	0.714	0.741
	0.7	1.000	0.999	0.971	0.851	0.834	1.000	0.997	0.921	0.625	0.561
	0.9	1.000	1.000	0.995	0.910	0.754	1.000	1.000	0.991	0.843	0.562
48	0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	0.999	0.993	0.987	0.989	1.000	0.996	0.968	0.933	0.935
	0.5	1.000	0.999	0.987	0.963	0.960	1.000	0.996	0.950	0.850	0.821
	0.7	1.000	0.999	0.986	0.927	0.889	1.000	0.998	0.956	0.772	0.652
	0.9	1.000	1.000	0.997	0.940	0.770	1.000	1.000	0.994	0.871	0.527

Table 5
Efficiency Gains from Overlapping Observations in Long-Horizon Regressions

Each entry is the ratio $\text{var}(b_{K(REG,lap)})/\text{var}(b_{K(REG,non)})$, where $b_{K(REG,lap)}$ is the OLS estimator obtained with overlapping observations, and $b_{K(REG,non)}$ is the OLS estimator obtained without overlapping (using only every K th observation).

K	γ	$\kappa = 3$					$\kappa = 13$				
		$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.97$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.97$
$b = 0.0$											
2	0.0	0.500	0.650	0.850	0.950	0.985	0.500	0.650	0.850	0.950	0.985
	0.3	0.500	0.611	0.781	0.902	0.966	0.500	0.591	0.730	0.847	0.934
	0.5	0.500	0.618	0.785	0.894	0.957	0.500	0.608	0.758	0.854	0.923
	0.7	0.500	0.630	0.807	0.905	0.956	0.500	0.626	0.796	0.887	0.933
	0.9	0.500	0.644	0.835	0.932	0.969	0.500	0.643	0.833	0.929	0.964
12	0.0	0.083	0.146	0.366	0.686	0.889	0.083	0.146	0.366	0.686	0.889
	0.3	0.083	0.140	0.343	0.663	0.879	0.083	0.131	0.308	0.622	0.858
	0.5	0.083	0.137	0.323	0.632	0.860	0.083	0.128	0.277	0.553	0.810
	0.7	0.083	0.137	0.312	0.588	0.822	0.083	0.132	0.279	0.501	0.731
	0.9	0.083	0.143	0.338	0.611	0.802	0.083	0.142	0.331	0.588	0.759
24	0.0	0.042	0.075	0.209	0.504	0.795	0.042	0.075	0.209	0.504	0.795
	0.3	0.042	0.073	0.202	0.495	0.790	0.042	0.070	0.189	0.477	0.780
	0.5	0.042	0.072	0.194	0.480	0.780	0.042	0.068	0.173	0.439	0.751
	0.7	0.042	0.072	0.185	0.448	0.751	0.042	0.069	0.163	0.379	0.679
	0.9	0.042	0.073	0.191	0.428	0.685	0.042	0.073	0.184	0.395	0.611
48	0.0	0.021	0.038	0.111	0.318	0.649	0.021	0.038	0.111	0.318	0.649
	0.3	0.021	0.038	0.109	0.315	0.647	0.021	0.037	0.105	0.309	0.643
	0.5	0.021	0.037	0.107	0.310	0.643	0.021	0.036	0.099	0.294	0.629
	0.7	0.021	0.037	0.103	0.296	0.627	0.021	0.036	0.093	0.263	0.589
	0.9	0.021	0.037	0.102	0.270	0.558	0.021	0.037	0.097	0.238	0.469
$b = 0.5$											
2	0.0	0.500	0.649	0.849	0.949	0.985	0.500	0.649	0.849	0.949	0.985
	0.3	0.500	0.614	0.784	0.904	0.966	0.500	0.594	0.736	0.851	0.935
	0.5	0.500	0.620	0.788	0.896	0.958	0.500	0.610	0.762	0.857	0.924
	0.7	0.500	0.631	0.809	0.906	0.956	0.500	0.628	0.799	0.889	0.934
	0.9	0.500	0.643	0.835	0.932	0.969	0.500	0.642	0.833	0.929	0.964
12	0.0	0.083	0.145	0.360	0.674	0.880	0.083	0.145	0.360	0.674	0.880
	0.3	0.083	0.139	0.340	0.657	0.872	0.083	0.131	0.310	0.628	0.857
	0.5	0.083	0.136	0.324	0.634	0.858	0.083	0.128	0.282	0.572	0.818
	0.7	0.083	0.137	0.314	0.598	0.826	0.083	0.132	0.282	0.522	0.749
	0.9	0.083	0.142	0.336	0.614	0.807	0.083	0.141	0.330	0.595	0.769
24	0.0	0.042	0.075	0.206	0.486	0.768	0.042	0.075	0.206	0.486	0.768
	0.3	0.042	0.073	0.199	0.480	0.765	0.042	0.070	0.188	0.468	0.759
	0.5	0.042	0.072	0.192	0.470	0.759	0.042	0.068	0.173	0.442	0.742
	0.7	0.042	0.071	0.184	0.448	0.740	0.042	0.068	0.163	0.394	0.693
	0.9	0.042	0.073	0.190	0.431	0.691	0.042	0.073	0.184	0.403	0.631
48	0.0	0.021	0.038	0.110	0.304	0.588	0.021	0.038	0.110	0.304	0.588
	0.3	0.021	0.038	0.108	0.302	0.587	0.021	0.037	0.105	0.298	0.585
	0.5	0.021	0.037	0.106	0.298	0.585	0.021	0.036	0.099	0.287	0.579
	0.7	0.021	0.037	0.103	0.288	0.579	0.021	0.036	0.093	0.263	0.560
	0.9	0.021	0.037	0.102	0.269	0.541	0.021	0.037	0.097	0.241	0.484

Table 6
Long-Horizon Regression Versus the OLS-VAR Approach

Each entry is the ratio $\text{var}(b_{K(1,VAR)})/\text{var}(b_{K(REG,lap)})$, where $b_{K(REG,lap)}$ is the estimator of b_K obtained in a long-horizon regression with overlapping observations, and $b_{K(1,VAR)}$ is the value of b_K implied by the parameters of a VAR estimated using OLS.

K	γ	$\kappa = 3$					$\kappa = 13$				
		$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.97$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.97$
b = 0.0											
2	0.0	0.500	0.650	0.850	0.950	0.985	0.500	0.650	0.850	0.950	0.985
	0.3	0.633	0.868	1.109	1.106	1.045	0.701	0.997	1.329	1.308	1.147
	0.5	0.609	0.830	1.092	1.136	1.072	0.642	0.890	1.205	1.284	1.187
	0.7	0.566	0.758	1.006	1.097	1.076	0.579	0.781	1.048	1.161	1.151
	0.9	0.521	0.684	0.901	1.006	1.034	0.524	0.689	0.909	1.018	1.051
12	0.0	0.083	0.097	0.205	0.521	0.813	0.083	0.097	0.205	0.521	0.813
	0.3	0.177	0.209	0.375	0.710	0.911	0.298	0.376	0.681	1.079	1.106
	0.5	0.206	0.251	0.487	0.899	1.027	0.315	0.412	0.890	1.569	1.440
	0.7	0.191	0.235	0.512	1.085	1.212	0.247	0.316	0.761	1.768	1.860
	0.9	0.124	0.148	0.328	0.843	1.211	0.133	0.160	0.361	0.955	1.437
24	0.0	0.042	0.047	0.092	0.292	0.652	0.042	0.047	0.092	0.292	0.652
	0.3	0.096	0.107	0.172	0.400	0.732	0.181	0.210	0.324	0.615	0.892
	0.5	0.121	0.139	0.234	0.517	0.831	0.218	0.265	0.476	0.949	1.186
	0.7	0.126	0.148	0.281	0.686	1.024	0.195	0.240	0.509	1.327	1.719
	0.9	0.084	0.097	0.201	0.646	1.244	0.098	0.114	0.244	0.841	1.758
48	0.0	0.021	0.023	0.043	0.135	0.438	0.021	0.023	0.043	0.135	0.438
	0.3	0.050	0.054	0.082	0.185	0.492	0.101	0.112	0.156	0.286	0.600
	0.5	0.066	0.073	0.113	0.241	0.560	0.135	0.155	0.243	0.449	0.803
	0.7	0.075	0.085	0.146	0.331	0.699	0.139	0.163	0.303	0.691	1.210
	0.9	0.059	0.068	0.130	0.398	1.028	0.080	0.092	0.186	0.636	1.812
b = 0.5											
2	0.0	0.442	0.611	0.837	0.947	0.985	0.442	0.611	0.837	0.947	0.985
	0.3	0.577	0.835	1.104	1.106	1.045	0.650	0.972	1.334	1.313	1.148
	0.5	0.552	0.795	1.086	1.137	1.073	0.586	0.858	1.204	1.288	1.188
	0.7	0.508	0.720	0.997	1.097	1.076	0.522	0.744	1.040	1.162	1.153
	0.9	0.463	0.645	0.890	1.005	1.034	0.466	0.651	0.898	1.016	1.051
12	0.0	0.082	0.093	0.182	0.479	0.798	0.082	0.093	0.182	0.479	0.798
	0.3	0.174	0.202	0.334	0.654	0.895	0.294	0.366	0.612	0.998	1.087
	0.5	0.203	0.244	0.438	0.833	1.010	0.312	0.404	0.819	1.479	1.422
	0.7	0.189	0.230	0.470	1.025	1.199	0.246	0.312	0.719	1.730	1.866
	0.9	0.122	0.144	0.299	0.799	1.207	0.132	0.156	0.331	0.915	1.446
24	0.0	0.041	0.046	0.085	0.241	0.607	0.041	0.046	0.085	0.241	0.607
	0.3	0.095	0.105	0.159	0.332	0.681	0.179	0.207	0.301	0.511	0.831
	0.5	0.120	0.136	0.217	0.430	0.774	0.216	0.262	0.445	0.795	1.107
	0.7	0.125	0.146	0.263	0.579	0.959	0.194	0.238	0.485	1.156	1.632
	0.9	0.084	0.096	0.191	0.569	1.204	0.097	0.114	0.234	0.768	1.762
48	0.0	0.021	0.023	0.042	0.109	0.352	0.021	0.023	0.042	0.109	0.352
	0.3	0.050	0.054	0.079	0.149	0.395	0.101	0.111	0.151	0.231	0.482
	0.5	0.066	0.073	0.109	0.195	0.450	0.134	0.154	0.234	0.364	0.645
	0.7	0.075	0.085	0.141	0.269	0.563	0.138	0.162	0.294	0.568	0.980
	0.9	0.059	0.067	0.127	0.339	0.863	0.079	0.092	0.184	0.570	1.624

Table 7

Long-Horizon Regression Versus the VAR Approach Using Two Additional Moment Conditions

Each entry is the ratio $\text{var}(b_{K(K,VAR)}^1)/\text{var}(b_{K(REG,lap)}^1)$, where $b_{K(REG,lap)}^1$ is the estimator of b_K obtained in a long-horizon regression with overlapping observations, and $b_{K(K,VAR)}^1$ is the value of b_K implied by the parameters of a VAR estimated using GMM with two long-horizon moment conditions in addition to those used in OLS.

K	γ	$\kappa = 3$					$\kappa = 13$				
		$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.97$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.97$
b = 0.0											
2	0.0	0.500	0.650	0.850	0.950	0.985	0.500	0.650	0.850	0.950	0.985
	0.3	0.633	0.846	0.991	1.000	0.999	0.701	0.933	0.989	0.960	0.976
	0.5	0.609	0.815	0.987	0.997	0.995	0.642	0.862	1.000	0.967	0.964
	0.7	0.566	0.753	0.959	1.000	0.994	0.579	0.774	0.975	0.994	0.975
	0.9	0.521	0.684	0.896	0.986	1.000	0.524	0.689	0.902	0.990	0.998
12	0.0	0.083	0.097	0.205	0.521	0.813	0.083	0.097	0.205	0.521	0.813
	0.3	0.177	0.208	0.360	0.665	0.881	0.298	0.370	0.575	0.815	0.940
	0.5	0.206	0.250	0.453	0.766	0.927	0.315	0.406	0.710	0.931	0.985
	0.7	0.191	0.235	0.482	0.854	0.973	0.247	0.315	0.665	0.981	0.999
	0.9	0.124	0.148	0.325	0.767	0.980	0.133	0.160	0.356	0.825	0.992
24	0.0	0.042	0.047	0.092	0.292	0.652	0.042	0.047	0.092	0.292	0.652
	0.3	0.096	0.107	0.169	0.387	0.717	0.181	0.209	0.302	0.524	0.795
	0.5	0.121	0.138	0.227	0.472	0.773	0.218	0.263	0.427	0.676	0.878
	0.7	0.126	0.148	0.273	0.584	0.854	0.195	0.239	0.468	0.826	0.962
	0.9	0.084	0.097	0.199	0.587	0.938	0.098	0.114	0.241	0.708	0.990
48	0.0	0.021	0.023	0.043	0.135	0.438	0.021	0.023	0.043	0.135	0.438
	0.3	0.050	0.054	0.081	0.183	0.487	0.101	0.111	0.151	0.266	0.561
	0.5	0.066	0.073	0.112	0.232	0.537	0.135	0.154	0.231	0.382	0.659
	0.7	0.075	0.085	0.144	0.307	0.621	0.139	0.163	0.289	0.532	0.788
	0.9	0.059	0.068	0.130	0.374	0.790	0.080	0.092	0.185	0.554	0.954
b = 0.5											
2	0.0	0.442	0.611	0.837	0.947	0.985	0.442	0.611	0.837	0.947	0.985
	0.3	0.577	0.813	0.986	1.000	0.999	0.650	0.909	0.993	0.964	0.977
	0.5	0.552	0.780	0.982	0.998	0.995	0.586	0.831	0.999	0.970	0.965
	0.7	0.508	0.716	0.950	1.000	0.995	0.522	0.737	0.968	0.995	0.976
	0.9	0.463	0.645	0.884	0.985	1.000	0.466	0.650	0.891	0.989	0.999
12	0.0	0.082	0.093	0.182	0.479	0.798	0.082	0.093	0.182	0.479	0.798
	0.3	0.174	0.201	0.322	0.617	0.866	0.294	0.360	0.526	0.770	0.928
	0.5	0.203	0.243	0.411	0.718	0.914	0.312	0.398	0.666	0.900	0.978
	0.7	0.189	0.230	0.445	0.814	0.964	0.246	0.311	0.632	0.964	0.999
	0.9	0.122	0.144	0.297	0.727	0.974	0.132	0.156	0.327	0.789	0.990
24	0.0	0.041	0.046	0.085	0.241	0.607	0.041	0.046	0.085	0.241	0.607
	0.3	0.095	0.105	0.157	0.323	0.669	0.179	0.205	0.281	0.449	0.750
	0.5	0.120	0.136	0.211	0.400	0.727	0.216	0.260	0.402	0.599	0.841
	0.7	0.125	0.146	0.256	0.506	0.811	0.194	0.237	0.448	0.760	0.937
	0.9	0.084	0.096	0.190	0.521	0.908	0.097	0.114	0.232	0.652	0.979
48	0.0	0.021	0.023	0.042	0.109	0.352	0.021	0.023	0.042	0.109	0.352
	0.3	0.050	0.054	0.078	0.148	0.393	0.101	0.110	0.146	0.219	0.460
	0.5	0.066	0.072	0.108	0.189	0.437	0.134	0.153	0.223	0.321	0.558
	0.7	0.075	0.085	0.139	0.254	0.516	0.138	0.162	0.281	0.461	0.695
	0.9	0.059	0.067	0.127	0.322	0.689	0.079	0.092	0.182	0.504	0.892

Table 8

Comparison of Approximate Slopes: Long-Horizon Regression versus Single-Period Regression

Each entry is the ratio of approximate slopes, $c(K)/c(1)$, where the approximate slope of the K -horizon regression is $c(K) = b_K^2 / (\text{Tvar}\{b_{K(REG,lap)}\})$, and the approximate slope of the single-period regression is $c(1) = b^2 / (\text{Tvar}\{b_{(1)}\})$. The null hypothesis is that $b = 0$, and the alternative hypothesis is that $b = 0.5$.

K	γ	$\kappa = 3$					$\kappa = 13$				
		$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.97$	$\rho = 0.0$	$\rho = 0.3$	$\rho = 0.7$	$\rho = 0.9$	$\rho = 0.97$
2	0.0	0.56	0.72	0.92	1.00	1.01	0.56	0.72	0.92	1.00	1.01
	0.3	0.73	0.99	1.21	1.16	1.07	0.82	1.15	1.47	1.38	1.18
	0.5	0.70	0.94	1.19	1.20	1.10	0.74	1.02	1.32	1.36	1.22
	0.7	0.64	0.85	1.10	1.15	1.11	0.66	0.88	1.14	1.22	1.18
	0.9	0.58	0.76	0.98	1.06	1.06	0.59	0.77	0.99	1.07	1.08
6	0.0	0.20	0.26	0.57	0.92	1.04	0.20	0.26	0.57	0.92	1.04
	0.3	0.38	0.51	0.99	1.24	1.16	0.55	0.82	1.68	1.82	1.39
	0.5	0.39	0.56	1.18	1.48	1.28	0.51	0.77	1.82	2.30	1.70
	0.7	0.34	0.47	1.08	1.56	1.40	0.39	0.55	1.34	2.06	1.85
	0.9	0.25	0.33	0.74	1.20	1.29	0.26	0.34	0.77	1.27	1.40
12	0.0	0.10	0.13	0.31	0.79	1.08	0.10	0.13	0.31	0.79	1.08
	0.3	0.22	0.28	0.57	1.08	1.21	0.37	0.50	1.05	1.65	1.47
	0.5	0.26	0.34	0.75	1.38	1.37	0.39	0.56	1.41	2.45	1.92
	0.7	0.24	0.32	0.81	1.70	1.62	0.31	0.43	1.23	2.86	2.53
	0.9	0.15	0.20	0.51	1.32	1.63	0.17	0.22	0.57	1.51	1.96
24	0.0	0.05	0.06	0.15	0.57	1.13	0.05	0.06	0.15	0.57	1.13
	0.3	0.12	0.15	0.29	0.78	1.27	0.23	0.29	0.54	1.20	1.55
	0.5	0.15	0.19	0.39	1.01	1.44	0.27	0.36	0.80	1.87	2.07
	0.7	0.16	0.20	0.47	1.36	1.79	0.24	0.33	0.87	2.72	3.04
	0.9	0.11	0.13	0.34	1.34	2.25	0.12	0.16	0.42	1.81	3.29
48	0.0	0.03	0.03	0.07	0.33	1.14	0.03	0.03	0.07	0.33	1.14
	0.3	0.06	0.07	0.14	0.45	1.28	0.13	0.15	0.27	0.69	1.56
	0.5	0.08	0.10	0.20	0.58	1.45	0.17	0.21	0.42	1.09	2.08
	0.7	0.09	0.12	0.25	0.81	1.82	0.17	0.22	0.53	1.70	3.16
	0.9	0.07	0.09	0.23	1.02	2.79	0.10	0.13	0.33	1.71	5.24

References

- Andrews, Donald W. K., 1991, "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.
- Bahadur, R. R., 1960, "Stochastic Comparison of Tests," *Annals of Mathematical Statistics* 31, 276-294.
- Baillie, Richard T. and Tim Bollerslev, 1992, "Prediction in Dynamic Models with Time-Dependent Conditional Variances," *Journal of Econometrics* 52, 91-113.
- Bollerslev, Tim, 1986, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* 31, 307-327.
- Bollerslev, Tim, Ray Y. Chou, and Kenneth F. Kroner, 1992, "ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics* 52, 5-59.
- Boudoukh, Jacob and Matthew Richardson, 1993, "The Statistics of Long Horizon Regressions Revisited," working paper, New York University, New York, NY, and University of Pennsylvania, Philadelphia, PA.
- Brockwell, Peter J. and Richard A. Davis, 1987, *Time Series: Theory and Methods*, Springer-Verlag, New York.
- Campbell, John Y., 1991, "A Variance Decomposition for Stock Returns," *Economic Journal* 101, 157-179.
- Campbell, John Y., 1992, "Why Long Horizons? Power Against Persistent Alternatives," working paper, Princeton University, Princeton, NJ.
- Cragg, J. G., 1983, "More Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form," *Econometrica* 51, 751-763.
- Diebold, Francis X., 1986, "Testing for Serial Correlation in the Presence of ARCH," *Proceedings of the American Statistical Association Business and Economics Statistics Section, 1986*, 323-328.
- Diebold, Francis X., 1988, *Empirical Modeling of Exchange Rate Dynamics*, Springer-Verlag, Berlin.
- Engle, Robert F., 1982, "Autoregressive Conditional Heteroskedasticity with Estimates of United Kingdom Inflation," *Econometrica* 50, 987-1007.

- Engle, Robert F. and Tim Bollerslev, 1986, "Modelling the Persistence of Conditional Variance." *Econometric Reviews* 5, 1-50.
- Fama, Eugene F. and Kenneth R. French, 1988, "Dividend Yields and Expected Stock Returns." *Journal of Financial Economics* 22, 3-25.
- Fama, Eugene F. and Kenneth R. French, 1989, "Business Conditions and Expected Returns on Stocks and Bonds," *Journal of Financial Economics* 25, 23-49.
- Geweke, John, 1981, "The Approximate Slope of Econometric Tests," *Econometrica* 49, 1427-1442.
- Hansen, Lars Peter, 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029-1054.
- Hansen, Lars Peter and Kenneth J. Singleton, 1990, "Efficient Estimation of Linear Asset Pricing Models with Moving-Average Errors," working paper, University of Chicago, Chicago, IL and Stanford University, Stanford, CA.
- Hodrick, Robert, 1992, "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies* 5, 357-386.
- Jegadeesh, Narasimhan, 1991, "Seasonality in Stock Price Mean Reversion: Evidence from the U.S. and the U.K.," *Journal of Finance* 46, 1427-1444.
- Kandel, Shmuel and Robert F. Stambaugh, 1987, "Long-Horizon Returns and Short-Horizon Models," CRSP working paper no. 222, University of Chicago, Chicago, IL.
- Kandel, Shmuel and Robert F. Stambaugh, 1988, "Modeling Expected Stock Returns for Long and Short Horizons," working paper, University of Chicago, Chicago, IL, and University of Pennsylvania, Philadelphia, PA.
- Keim, Donald B. and Robert F. Stambaugh, 1986, "Predicting Returns in the Stock and Bond Markets," *Journal of Financial Economics* 17, 357-390.
- Mankiw, N. Gregory and Matthew D. Shapiro, 1986, "Do We Reject Too Often? Small Sample Properties of Rational Expectations Models," *Economics Letters* 20, 243-247.
- Milhøj, Anders, 1985, "The Moment Structure of ARCH Processes," *Scandinavian Journal of Statistics* 12, 281-292.
- Mishkin, Frederic S., 1990, "Does Correcting for Heteroskedasticity Help?" *Economics Letters* 34, 351-356.

- Nelson, Charles R. and Myung J. Kim, 1993, "Predictable Stock Returns: The Role of Small Sample Bias," *Journal of Finance* 48, 641-661.
- Newey, Whitney K. and Kenneth D. West, 1987, "A Simple, Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.
- Richardson, Matthew and Tom Smith, 1991, "Tests of Financial Models in the Presence of Overlapping Observations," *Review of Financial Studies* 4, 227-254.
- Richardson, Matthew and Tom Smith, 1992, "Robust Power Calculations with Tests for Serial Correlation in Stock Returns." working paper, University of Pennsylvania, Philadelphia, PA and Duke University, Durham, NC.
- Richardson, Matthew and James H. Stock, "Drawing Inferences from Statistics Based on Multiyear Asset Returns," *Journal of Financial Economics* 25, 323-348.
- Stambaugh, Robert F., 1986, "Bias in Regressions with Lagged Stochastic Regressors," CRSP working paper no. 156, University of Chicago, Chicago, IL.
- White, Halbert, 1984, *Asymptotic Theory for Econometricians*, Academic Press, Orlando.

</ref_section>