Econometric Theory, **30**, 2014, 60–93.
doi:10.1017/S0266466613000121

ESTIMATING THE PERSISTENCE AND THE AUTOCORRELATION FUNCTION OF A TIME SERIES THAT IS MEASURED WITH ERROR

PETER R. HANSEN
European University Institute and CREATES

ASGER LUNDE
Aarhus University and CREATES

An economic time series can often be viewed as a noisy proxy for an underlying economic variable. Measurement errors will influence the dynamic properties of the observed process and may conceal the persistence of the underlying time series. In this paper we develop instrumental variable (IV) methods for extracting information about the latent process. Our framework can be used to estimate the autocorrelation function of the latent volatility process and a key persistence parameter. Our analysis is motivated by the recent literature on realized volatility measures that are imperfect estimates of actual volatility. In an empirical analysis using realized measures for the Dow Jones industrial average stocks, we find the underlying volatility to be near unit root in all cases. Although standard unit root tests are asymptotically justified, we find them to be misleading in our application despite the large sample. Unit root tests that are based on the IV estimator have better finite sample properties in this context.

1. INTRODUCTION

Many economic time series are constructed from survey statistics or composed of estimates that involve sampling error. It is natural to view such time series as proxies for the underlying population quantities. In this paper we develop instrumental variable (IV) methods that facilitate the analysis of two features of the underlying time series: the persistence of the latent time series and its autocorrelation function. The instrumental variables are lagged value of the observed time series that

The authors thank participants at the 2006 Zeuthen Lectures at University of Copenhagen, the 2008 SOFIE inaugural conference, and the 2008 workshop at Stanford Institute for Theoretical Economics. The Ox language (Doornik, 2006) was used to perform the calculations reported here. Center for Research in Econometric Analysis of Time Series (CREATES) is a research center at Aarhus University funded by the Danish National Research Foundation. Finally, we thank Professor Jun Yu (guest co-editor) and two anonymous referees for many suggestions that improved our manuscript. Address correspondence to Peter R. Hansen, European University Institute, Department of Economics, Villa San Paolo, Via della Piazzuola 43, 50133 Florence, Italy; e-mail: peter.hansen@eiu.eu.

take advantage of moment restrictions, known as multiperiod conditional moment restrictions, see Hansen (1985).

To take a concrete example, consider a daily time series of realized variances. Each element of this time series can be viewed as a noisy estimate of the latent volatility. Much progress has recently been made in estimating financial volatility from high-frequency data using realized measures, such as the realized variance. Despite this progress, it is important to discriminate between the realized measure of volatility and the underlying population quantity. The most accurate estimators of daily volatility that utilize thousands of high-frequency prices produce estimates for which the standard error is rarely less than 10% of the point estimate; see, e.g., Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). Measurement errors of this magnitude cause the autocorrelation function of the observed time series to look distinctively different from that of the underlying time series.

The methods used in this paper are best suited for situations where the latent process is persistent, because nonzero autocorrelations are needed for these instruments to be valid. In fact, the more persistent is the latent time series, the "stronger" will these instruments be, other things being equal. We focus on two aspects of the problem. First, we show that the IV methods provide an effective way to assess the degree of persistence for the latent time series. An implication is that unit root tests based on IV regression methods have better finite sample properties than conventional tests. Second, we propose an alternative estimator of the autocorrelation function (ACF) for the latent time series. This estimator is more informative about the ACF of the latent process, in particular if the process is persistent.

Our basic framework is simple: We model the latent time series, y_t, as an ARMA(p,q) process and the measurement error as white noise. Our first objective is to estimate a key persistence parameter that we denote by π. The IV estimator of π bypasses the need for estimating all autoregressive moving average (ARMA) parameters, which is useful as these cannot be fully identified from the contaminated time series, without additional assumptions. Our second objective is to derive an approximate estimator of the autocorrelation function for the latent process. The empirical problem that has motivated this analysis is the situation where the latent time series is daily volatility, whereas the observed time series is a sequence of realized measures, such as the realized variance, that are computed with high-frequency data. Because each of these estimates is computed with different high-frequency data (data from distinct days), it is reasonable to assume that their sampling errors are uncorrelated. This structure arises in some rational expectations model, and our instrumental variables are exploiting what Hansen (1985) called multiperiod conditional moment restrictions; see also Hansen, Heaton, and Ogaki (1988), Hansen and Singleton (1996), and West (2001). For a survey of this literature, see Anatolyev (2001). In the context of volatility models, Meddahi and Renault (2004) derived these moment conditions (for squared returns) that characterize a state space model where latent volatility is a linear combination of factors that follow a VAR(1) process. However, Meddahi and Renault did not

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

explore estimation and assumed the underlying factors to be stationary. See also Bollerslev and Zhou (2002) and Meddahi (2003), who utilized similar restrictions in the context of continuous time volatility models.

The analysis that we present in this paper is related to the literature on unit root test in the context of moving average innovations. The unit root test based on instrumental variables (that is very similar to ours) was proposed by Hall (1989), and generalized by Phillips and Hansen (1990). Perron and Ng (1996) have proposed another unit root test that is also robust to the sort of moving average innovations that arises in this context. Our analysis is also related to the literature on unobserved component models that deals with signal extraction, detrending, and filtering of noisy time series (see, for example, Ashley and Vaughan, 1986; Watson, 1986; Harvey, 2001; Harvey and Proietti, 2005; and Harvey and De Rossi, 2006). In this strand of the literature the statistical treatment of the underlying component is usually carried out by setting up the state space form and applying the associated (Kalman) filter and smoothing algorithms. These methods have been applied to realized measures of volatility by Barndorff-Nielsen and Shephard (2002), Barndorff-Nielsen, Nielsen, Shephard, and Ysusi (2004), Hansen and Lunde (2005b), and Koopman, Jungbacker, and Hol (2005). These papers show that the measurement error (sampling error) is a nontrivial component of the realized variance.

Our analysis contributes with new theoretical results that focus on the persistence of the underlying process. The IV methods we develop in this paper compliment existing methods and offer some advantages. For instance, the IV-based autocorrelation function shares the simplicity and nonparametric nature of the conventional empirical autocorrelation.

We make the following contributions: First, we propose simple instrumental variable estimators of a key parameter that captures the persistence of ARMA processes. For an AR(1) process the persistence parameter is simply the autoregressive coefficient. The persistence parameter can be estimated with simple IV estimators, and for the ARMA(1,1) case we derive the optimal IV estimator that exploits the particular covariance structure in this framework. Second, we propose an approximate ACF for the latent time series. This is important because many economic time series can be viewed as noisy proxies of the fundamental underlying process, and measurement errors can cause the ACF of the observed time series to look distinctively different from that of the latent process. Third, in our empirical analysis of realized measures of volatility, we find that actual volatility is very close to having a unit root. In fact the largest autoregressive root is typically in the range between 0.98 and 1.00. In this context, we show that standard unit root tests can be very misleading. Conventional unit root tests may be asymptotically justified in this context, but their finite sample properties can be quite poor—even with a sample size that is well over a thousand observations. The main reason is that the measurement errors are relatively large in these time series. Fourth, we make some remarks on fractionally integrated processes. The popularity of fractionally integrated processes for the modeling of volatility

is to some extent driven by two empirical observations: The shape of the ACF for observed volatility, and the apparent rejection of the unit root hypothesis using conventional unit root tests. In this paper we argue that neither can be taken as evidence of fractional integration. The reason is that a unit root process, or a local-to-unit root process, can also induce these empirical features when the magnitude of measurement errors is sufficiently large—a feature that resembles the famous aggregation result by Granger (1980). In the context of daily measures of volatility and other time series where the degree of measurement error is pronounced, we believe it is important to account for the measurements errors. In the context of autoregressive fractionally integrated moving average (ARFIMA) models this amounts to a serious modeling of the moving average (MA) part, which is sometimes neglected in this literature. We do not, however, dismiss the fractionally integrated model as a good model of daily volatility. In fact, our approximate ACF estimates have features that are consistent with long memory processes, and the instrumental variable unit root tests do reject the unit root hypothesis for most of the volatility time series. So in this regard we arrive at the same conclusion as Wright (1999).

This paper is organized as follows. In Section 2 we describe the theoretical framework, introduce the instrumental variable estimator, and derive its asymptotic properties. In Section 3 we introduce a novel estimator of the autocorrelation function for the latent time series and illustrate some of its advantages. In Section 4 we present an empirical analysis with realized measures of volatility, two macroeconomic time series of inflation, and two long time series of absolute returns. We estimate the ACF for the underlying time series for all time series. For the time series of realized measures, we show that the underlying volatility is highly persistent and close to unit root in all cases. Concluding remarks are given in Section 4, and an Appendix contains all proofs.

2. AN INSTRUMENTAL VARIABLE APPROACH TO ASSESSING THE PERSISTENCE OF A LATENT TIME SERIES

In this section we study some methods for assessing the persistence of a time series that is measured with error. We consider a class of simple instrumental variable estimators and show that these are consistent for the parameter that measures the persistence.

We use an ARMA(p,q) specification for the latent time series, y_t, and treat the observed volatility, x_t, as a noisy and possibly biased estimate of y_t.. So our model is

$$\varphi(L)(y_t - \delta) = \theta(L)\varepsilon_t, \tag{1}$$

$$x_t = y_t + \xi + \eta_t, \tag{2}$$

where $\varphi(L)$ and $\theta(L)$ are lag polynomial of orders p and q, respectively, δ and ξ are constants, ε_t the innovations to the underlying process, and η_t represents the

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

stochastic part of the measurement error. Precise assumptions about ε_t and η_t will be stated in Assumption 1. In the context of time series of volatility, the integrated variance for day t and a corresponding realized measure of volatility can taken to be the latent process, y_t, and the observed process, x_t, respectively. In our empirical analysis we study two types of realized measures, specifically, the realized variance and the realized kernel.

The model has the following implication for the observed time series.

LEMMA 1. *Given (1) and (2), we have*

$$\varphi(L)\,(x_t - \delta - \xi) = \theta(L)\varepsilon_t + \varphi(L)\eta_t. \tag{3}$$

The lemma shows that x_t is an ARMA process with exactly the same autoregressive polynomial. In the context of time series of volatility, this result was noted in Barndorff-Nielsen and Shephard (2002) and Meddahi (2003); see also Andersen, Bollerslev, and Meddahi (2004).

We make the following assumptions.

Assumption 1. The characteristic polynomials

$$\varphi(z) = 1 - \varphi_1 z - \cdots - \varphi_p z^p \quad \text{and} \quad \theta(z) = 1 - \theta_1 z - \cdots - \theta_q z^q$$

do not have any roots in common and are such that

$$\varphi(z)\big/(1-z) = 0 \;\Rightarrow\; |z| > 1 \quad \text{and} \quad \theta(1) \neq 0,$$

and $\{\varepsilon_t, \eta_t\}$ is a sequence of independent and identically distributed random variables with zero mean and $\sigma_\varepsilon^2 = \operatorname{var}(\varepsilon_t)$, $\sigma_\eta^2 = \operatorname{var}(\eta_t)$, and $\operatorname{cov}(\varepsilon_t, \eta_t) = 0$.

With Assumption 1 we ensure that y_t is either integrated of order zero, I(0), or integrated of order one, I(1). The first part of the assumption allows $\varphi(z)$ to have a single unit root, $\varphi(1) = 0$, but the multiplicity of this unit root is at most one. This rules out integration of an order higher than one. For example, y_t is I(2) when the multiplicity of the unit root is two. The requirement that $\theta(1) \neq 0$ ensures that $\theta(L)\varepsilon_t$ is I(0). Without this condition we would not have the previous relation between the roots of $\varphi(z)$ and the order of integration.

A key parameter for our analysis is the *persistence parameter* that is defined by

$$\pi = \max_{i=1,\ldots,p} \frac{1}{|z_i^*|},$$

where $z_1^*, \ldots, z_p^*$ are the roots of the characteristic polynomial, i.e., $\varphi(z_i^*) = 0$, $i = 1, \ldots, p$. We note that $\pi = 1$ when $\varphi(z)$ has a unit root, and for persistent processes, the implicit function theorem (applied to $f(\varphi_1, \ldots, \varphi_p, z) = 1 - \varphi_1 z - \cdots - \varphi_p z^p = 0$ about $z = 1$) shows that π is roughly equal to

$$\varphi_\bullet = \varphi_1 + \cdots + \varphi_p.$$

The persistence parameter can also be defined from the companion form for y_t. Without loss of generality, consider the case with $\delta = 0$ and $\theta(z) = 1$. Then

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

$Y_t = (y_t, \ldots, y_{t-p+1})'$, can be expressed as a VAR(1) process, $Y_t = \Phi Y_{t-1} + \varepsilon_t^*$, where $\varepsilon_t^* = (\varepsilon_t, 0, \ldots, 0)'$, and

$$\Phi = \begin{pmatrix} \varphi_1 & \varphi_2 & \cdots & \varphi_{p-1} & \varphi_p \\ 1 & 0 & \cdots & & 0 \\ 0 & 1 & \cdots & & \\ \vdots & & \ddots & & \\ 0 & & & 1 & 0 \end{pmatrix}.$$

The persistence parameter, π, is simply given as the spectral radius of Φ (the largest eigenvalue as measured in absolute value).

Much of our analysis can be understood from the simplest case, where $p = 1$ and $q = 0$. This case is outlined in the following example.

Example 1
Suppose that Assumption 1 holds with $p = 1$ and $q = 0$. Then y_t is an *AR(1)* process, and by manipulating the two expressions,

$$y_t = \pi y_{t-1} + (1-\pi)\delta + \varepsilon_t \quad \text{and} \quad x_t = y_t + \xi + \eta_t,$$

we have that $x_t = \pi x_{t-1} + (1-\pi)(\delta + \xi) + \varepsilon_t + \eta_t - \pi \eta_{t-1}$.

Note that we have parameterized the constant in such a way that it vanishes whenever $\pi = 1$. This ensures that y_t does not have a deterministic trend in the unit root case.

2.1. Instrumental Variable Estimators

We consider the class instrumental variable estimators of the persistence parameter π, which we defined in (3). These estimators have the form

$$\hat{\pi}_{\mathrm{IV}_z} = \frac{\sum_{t=1}^n z_t x_{t+1}}{\sum_{t=1}^n z_t x_t}, \tag{4}$$

where we refer to z_t as an instrumental variable or simply an instrument. The expression (4) defines a large class of estimators that includes the least squares estimator and instrumental variable estimators including the two-stage least squares estimator. For example, when the instrumental variable is a lagged value of the observed time series (less its sample average), we have

$$\hat{\pi}_{\mathrm{IV}_j} = \frac{\sum_{t=1}^n (x_{t-j} - \bar{x}_j) x_{t+1}}{\sum_{t=1}^n (x_{t-j} - \bar{x}_j) x_t}, \qquad j = 0, 1, 2, \ldots,$$

where $\bar{x}_j = n^{-1} \sum_{t=1}^n x_{t-j}$. When $j = 0$ this estimator simplifies to the least squares estimator

$$\hat{\pi}_{\mathrm{LS}} = \frac{\sum_{t=1}^n (x_t - \bar{x}_0) x_{t+1}}{\sum_{t=1}^n (x_t - \bar{x}_0) x_t}.$$

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

The two-stage least squares (TSLS) estimator, which is based on multiple instruments, $\tilde{Z}_t = (x_{t-J_1} - \bar{x}_{J_1}, \ldots, x_{t-J_2} - \bar{x}_{J_2})'$ with $0 \leq J_1 \leq J_2$, can also be expressed in the form of (4). In this case we have

$$z_t = \tilde{Z}_t' \hat{\alpha}_{\mathrm{TSLS}}, \quad \text{where} \quad \hat{\alpha}_{\mathrm{TSLS}} = \left(\sum_{t=1}^{n} \tilde{Z}_t \tilde{Z}_t' \right)^{-1} \sum_{t=1}^{n} \tilde{Z}_t x_t. \tag{5}$$

This is not surprising, because the first step in a TSLS procedure amounts to a dimension reduction, where a vector of instruments is mapped into a vector of instruments that has a dimension that matches that of the regression parameters. The present problem has a particular covariance structure that we can utilize to determine the optimal linear combination, α^*, of the set of instrumental variables. This will lead to an IV estimator that is more efficient than the TSLS estimator.

2.2. Properties of Estimators: The AR(1) Case

Initially we establish the properties of our estimators assuming that y_t is a simple AR(1) process. Later we consider more general ARMA specifications and evaluate the extent to which the IV estimators are robust.

LEMMA 2. *Suppose that Assumption 1 holds with $p = 1$ and $q = 0$, so that $\pi = \varphi_1$. When $|\pi| < 1$ and y_0 is assigned the stationary distribution for y_t, we have*

$$\mathrm{var}(y_t) = \sigma_y^2 = \frac{\sigma_\varepsilon^2}{1-\pi^2} \quad \textit{and} \quad \mathrm{cov}(x_t, x_{t-h}) = \begin{cases} \sigma_y^2 + \sigma_\eta^2 & \textit{for } h = 0, \\ \pi^h \sigma_y^2 & \textit{for } h \neq 0. \end{cases}$$

When $|\pi| < 1$, we have the usual errors-in-variable problem, because x_{t-1} and $u_t = \varphi(L)\eta_t + \theta(L)\varepsilon_t$ are correlated. For instance, under the assumptions of Lemma 2 we have $u_t = \eta_t - \pi\eta_{t-1} + \varepsilon_t$, so that $\mathrm{cov}(x_{t-1}, u_t) = -\pi\sigma_\eta^2$. This correlation causes the well-known attenuation bias of the least squares estimator, which has previously been discussed in the context of realized measures by Barndorff-Nielsen and Shephard (2002).

THEOREM 1 (Least squares estimator). *Suppose that Assumption 1 holds with $p = 1$ and $q = 0$. Let $\lambda = \sigma_\eta^2/\sigma_\varepsilon^2$.*

(i) We have

$$\hat{\pi}_{\mathrm{LS}} \xrightarrow{p} \pi \frac{1}{1 + \lambda(1-\pi^2)}.$$

(ii) When $\pi = 1$, we have

$$n(\hat{\pi}_{\mathrm{LS}} - 1) \xrightarrow{d} \frac{\int_0^1 (W_u - \bar{W})\mathrm{d}W_u - \lambda}{\int_0^1 (W_u - \bar{W})^2 \mathrm{d}u}.$$

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

The first part of Theorem 1 shows that the least squares estimator is inconsistent unless $\lambda = 0$ or $\pi = 1$. Although measurement errors, $\lambda > 0$, do not render the least squares estimator inconsistent when $\pi = 1$, they do affect the asymptotic distribution, because λ shows up in the limit distribution. Despite the consistency of $\hat{\pi}_{\mathrm{LS}}$ in the unit root case, the stochastic bias,

$$-\frac{\lambda}{n\int_0^1 (W_u - \bar{W})^2 \mathrm{d}u}, \tag{6}$$

need not be negligible in finite samples. This will be illustrated in our empirical application, where the bias is sizable despite a large sample size.

When $p = 1$ such that $\varphi(L) = 1 - \pi L$, we have the decomposition of our estimators,

$$\hat{\pi}_{\mathrm{IV}_z} = \pi + \frac{\sum_{t=1}^n z_t u_{t+1}}{\sum_{t=1}^n z_t x_t}, \quad \text{where} \quad u_{t+1} = \varepsilon_{t+1} + \eta_{t+1} - \pi \eta_t.$$

This shows the key to consistency whenever $|\pi| < 1$ is an instrument that is uncorrelated with u_{t+1}. For example, when $y_t \sim$ARMA(1,1) as in Meddahi (2002, 2003), the instrument $z_t = x_{t-j} - \bar{x}_j$ is valid for any $j \geq 1$, because $\mathrm{cov}(x_{t-j}, u_{t+1}) = 0$ for $j \geq 1$.

THEOREM 2 (Instrumental variable estimator). *Suppose that Assumption 1 holds with $p = 1$ and $q = 0$.*

(i) When $|\pi| < 1$, we have, for $j > 0$,

$$n^{1/2}(\hat{\pi}_{\mathrm{IV}_j} - \pi) \xrightarrow{d} N\left[0, \pi^{-2j}\left(1-\pi^2\right)\left\{1 + 2\left(1-\pi^2\right)\lambda + \left(1-\pi^4\right)\lambda^2\right\}\right],$$

where $\lambda = \sigma_\eta^2/\sigma_\varepsilon^2$.

(ii) (Hall, 1989) When $\pi = 1$, we have, for $j > 0$,

$$n(\hat{\pi}_{\mathrm{IV}_j} - \pi) \xrightarrow{d} \frac{\int_0^1 (W_u - \bar{W})\mathrm{d}W_u}{\int_0^1 (W_u - \bar{W})^2 \mathrm{d}u}.$$

For the case where $y_t \sim$AR(1), Theorem 2 shows that the instrumental variable estimator is consistent for π, when $j \geq 1$. The first part of the theorem shows that x_{t-1} is the most efficient instrumental variable, among $x_{t-1}, x_{t-2}, \ldots$, when $|\pi| < 1$, because $\hat{\pi}_{\mathrm{IV}_1}$ has the smallest asymptotic variance. This is intuitive because the autocovariance function is for $j \geq 1$ given by $\mathrm{corr}(x_t, x_{t-j}) = \pi^j$, so that x_{t-1} is more correlated with x_t than is x_{t-j}, for $j \geq 2$. The asymptotic distribution for the case where $\pi = 1$ is due to Hall (1989), who emphasized the benefits of using $\hat{\pi}_{\mathrm{IV}_j}$ to test for unit roots rather than $\hat{\pi}_{\mathrm{LS}}$, because the former has an asymptotic distribution that is free of the nuisance parameter, $\lambda = \sigma_\eta^2/\sigma_\varepsilon^2$. cf. Theorem 1$(ii)$ and Theorem 2(ii).

Having an asymptotic distribution that depends on nuisance parameters is obviously inconvenient, but the conventional unit root test has another flaw that is

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

more problematic in our empirical application. The standard OLS-based unit root test is known to be highly size distorted in the presence of a large moving average root, see Schwert (1989) and Perron and Ng (1996), and this size distortion is largely due to the stochastic bias that we defined in (6). Despite the fact that the stochastic bias is only of order $O_p(n^{-1})$, which suggests it vanishes quickly as $n \to \infty$, it can play a major role even if n is large. The reason is that λ can be large, so that the bias is not negligible even with a fairly large sample size. This is indeed the case in our application with realized measures of volatility, where n is close to 2,000. For the time series of realized variances, we find the least squares estimator to be about 30% smaller than the IV estimator, and for the more accurate estimator of volatility, the realized kernel, we find the downward bias to be about 15%.

The instruments in Theorem 2 are single variable instruments, in the sense that they are based on a single lag of x_t. When $|\pi| < 1$, we can construct a more efficient IV by taking a linear combination of multiple instruments, $(x_{t-1}, x_{t-2}, \ldots)$.

THEOREM 3 (Optimal instrument). *Suppose that Assumption 1 holds with $p = 1$ and $q = 0$, and consider the case where $|\pi| < 1$. Let $Z_t = (x_{t-1} - \bar{x}_1, \ldots x_{t-J} - \bar{x}_J)'$, where $J \geq 1$ is the dimension of Z_t. Then*

$$\operatorname{avar}\left(n^{-1/2}\sum_{t=1}^{n} Z_t u_{t+1}\right) = \sigma_\varepsilon^4 M_{\pi,\lambda} \quad \text{and} \quad \operatorname*{plim}_{n\to\infty}\left(n^{-1}\sum_{t=1}^{n} Z_t x_t\right) = \sigma_y^2 V_\pi,$$

where $V_\pi = \left(\pi, \pi^2, \ldots, \pi^J\right)'$, and

$$M_{\pi,\lambda} = \tfrac{1}{1-\pi^2} B_1 + 2\lambda I + \lambda^2 B_2,$$

where I is the $J \times J$ identity matrix, and B_1 and B_2 are symmetric band matrices given by

$$B_1 = \begin{pmatrix} 1 & \pi & \pi^2 & \cdots & \pi^{J-1} \\ \pi & 1 & \pi & \ddots & \vdots \\ \pi^2 & \pi & 1 & \ddots & \pi^2 \\ \vdots & \ddots & \ddots & \ddots & \pi \\ \pi^{J-1} & \cdots & \pi^2 & \pi & 1 \end{pmatrix},$$

$$B_2 = \begin{pmatrix} 1+\pi^2 & -\pi & 0 & \cdots & 0 \\ -\pi & 1+\pi^2 & -\pi & \ddots & \vdots \\ 0 & -\pi & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -\pi \\ 0 & \cdots & 0 & -\pi & 1+\pi^2 \end{pmatrix}.$$

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

Within the class of instrumental variables, $\{z_t : z_t = Z_t'\alpha \text{ for some } \alpha \in \mathbb{R}^J\}$*, the smallest asymptotic variance is achieved with*

$$z_t^* = Z_t'\alpha_{\pi,\lambda}^*, \quad \text{where} \qquad \alpha_{\pi,\lambda}^* = M_{\pi,\lambda}^{-1} V_\pi,$$

and the asymptotic variance of this optimal IV estimator, $\hat{\pi}^* = \hat{\pi}_{\mathrm{IV}_{z^*}}$*, is*

$$\begin{aligned}\operatorname{avar}\left\{n^{-1/2}(\hat{\pi}^* - \pi)\right\} &= \frac{(1-\pi^2)^2}{V_\pi' M_{\pi,\lambda}^{-1} V_\pi} \\ &= \frac{1-\pi^2}{V_\pi' \left\{B_1 + 2\lambda(1-\pi^2)I + \lambda^2(1-\pi^2)B_2\right\}^{-1} V_\pi}.\end{aligned}$$

Naturally, the optimal linear combination is scale invariant, in the sense that cz_t^* is also an optimal instrument for any $c \neq 0$. The estimator obviously has a generalized method of moments (GMM) interpretation, since it is deduced from moment conditions. In relation to GMM, a contribution of the theorem is the closed-form expression for the optimal weighting matrix.

2.2.1. Implementation of Multiple-Variable IV. The optimal linear combination depends on unknown parameters, so in our empirical application we will use a two-step estimation procedure, which may be iterated if needed. In the first step we obtain preliminary estimates of π and λ. For example, one can estimate π by a the TSLS estimator and then estimate λ, by

$$\hat{\lambda}_{\hat{\pi}} = -\frac{\hat{\rho}_{\Delta x,1} + \frac{1-\hat{\pi}}{2}}{\left(1+\hat{\pi}\right)\hat{\rho}_{\Delta x,1} + \frac{1+\hat{\pi}}{2}}, \tag{7}$$

where

$$\rho_{\Delta x,1} = \frac{\operatorname{cov}(\Delta x_t, \Delta x_{t-1})}{\operatorname{var}(\Delta x_t)}.$$

In the second step the instrument $z_t = Z_t'\alpha_{\hat{\pi},\hat{\lambda}}$ is computed and used to obtain new estimates of π and λ.[1] If necessary, the second step can be iterated until the estimates have converged. In our empirical application the estimates converged in just two iterations. In GMM terminology, the resulting estimate is the continuously updated GMM (CUGMM) estimator.

Note that the estimator in (7) simplifies to

$$\hat{\lambda}_{\mathrm{LL}} = -\frac{\hat{\rho}_{\Delta x,1}}{2\hat{\rho}_{\Delta x,1} + 1}$$

when $\hat{\pi} = 1$, which is the estimator of $\lambda = \sigma_\eta^2/\sigma_\varepsilon^2$, in the local level model

$$y_t = y_{t-1} + \varepsilon_t \quad \text{and} \quad x_t = y_t + \eta_t,$$

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

which is motivated by the fact that $\rho_{\Delta x,1} = -\frac{\lambda}{2\lambda+1}$ in this model; see, e.g., Harvey (1993). When $|\pi| < 1$, $p = 1$ and $q = 0$, the estimator in (7) can be motivated by the identity

$$\rho_{\Delta x,1} = \frac{\operatorname{cov}(\Delta x_t, \Delta x_{t-1})}{\operatorname{var}(\Delta x_t)} = -\frac{\sigma_\varepsilon^2\left(\lambda + \frac{1-\pi}{1+\pi}\right)}{2\sigma_\varepsilon^2(\lambda + \frac{1}{1+\pi})}$$
$$= -\frac{1}{2}\frac{(1+\pi)\lambda + (1-\pi)}{(1+\pi)\lambda + 1} = \frac{1}{2}\left\{\frac{\pi}{(1+\pi)\lambda + 1} - 1\right\}.$$

In our application we found this multiple-variable IV estimator to be insensitive to the estimate of λ, and the point estimates of π were always very similar to the TSLS estimates that do not require an estimate of λ.

2.2.2. A Comment on the Inefficiency of TSLS. In the stationary case $|\pi| < 1$, this optimal IV estimator, $\hat{\pi}^*$, will be similar to the TSLS estimator when λ is small. This follows from the fact that

$$n^{-1}\sum_{t=1}^{n} Z_t Z_t' \xrightarrow{p} \sigma_\varepsilon^2\left(\tfrac{1}{1-\pi^2}B_1 + \lambda I\right) = \sigma_\varepsilon^2\left(M_{\pi,\lambda} - \lambda I - \lambda^2 B_2\right),$$

when $|\pi| < 1$. The linear combination of instruments that is implied by the TSLS estimator converges in probability to $\alpha_{\mathrm{TSLS}} = c\left(M_{\pi,\lambda} - \lambda I - \lambda^2 B_2\right)^{-1} V_\pi$, as $n \to \infty$, for some constant $c \neq 0$. This will be similar to $\alpha^* = M_{\pi,\lambda}^{-1} V_\pi$ when λ is small. Still, the asymptotic variance of the TSLS estimator exceeds that of the optimal instrument when $\lambda > 0$ in the stationary case $|\pi| < 1$.

It is interesting (and perhaps surprising) that the TSLS is inefficient in the present context. In a textbook setting, the TSLS estimator is asymptotically efficient when Z_t is a vector of valid instruments that meets certain rank conditions, and we have $\mathbb{E}(u_{t+1}^2 Z_t Z_t') = \mathbb{E}(u_{t+1}^2)\mathbb{E}(Z_t Z_t')$. These conditions are all satisfied here. So how can it be that z_t^* is a better linear combination of Z_t than that implied by the TSLS estimator? The TSLS is asymptotically efficient if $\mathbb{E}(Z_t Z_t')$ is proportional to the asymptotic variance of $n^{-1/2}\sum_{t=1}^n Z_t u_{t+1}$, where the latter, in a textbook setup, equals $\mathbb{E}(u_{t+1}^2 Z_t Z_t')$. However, this is not true in the present context, because $\{Z_t u_{t+1}\}$ is not a martingale difference sequence. The asymptotic variance of $n^{-1/2}\sum_{t=1}^n Z_t u_{t+1}$ is $\sigma_\varepsilon^4 M_{\pi,\lambda}$, which does not equal $\mathbb{E}(u_{t+1}^2 Z_t Z_t')$, and this is the underlying reason that TSLS is inefficient in the present context.

The IV estimator computed with z_t^* is based on the solution to an eigenvalue problem. This is a feature that is shared by the well-known limited information maximum likelihood (LIML) estimator; see Anderson and Rubin (1949). Yet the two estimators are different, because the optimal estimator take full advantage of the particular covariance structure in this model.

2.2.3. Quantifying the Asymptotic Variance. In Figure 1 we have plotted the asymptotic variance for several estimators in the situation where $\pi = 0.975$ and

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

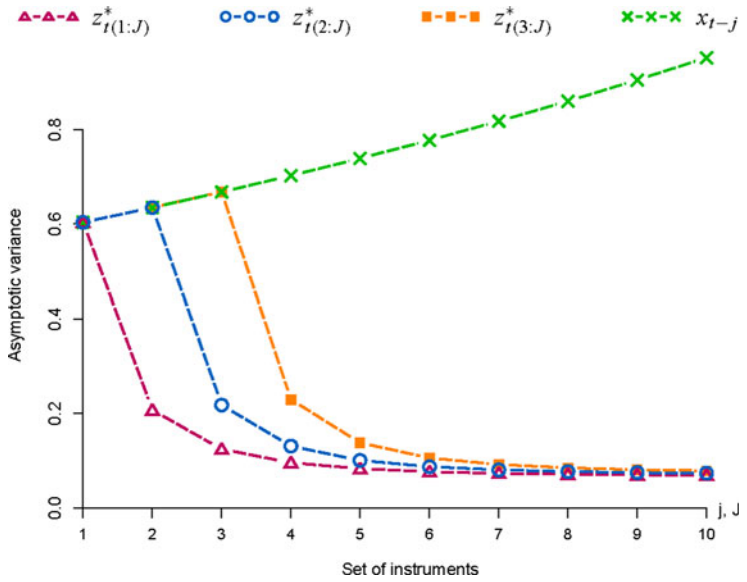

FIGURE 1. The asymptotic variance of the IV estimator based on different subset of instrumental variables. Crosses correspond to the single-variable instrument estimators where $z_t = x_{t-j} - \bar{x}_j$; triangles represent the optimal combinations of $(x_{t-1} - \bar{x}_1, \ldots, x_{t-J} - \bar{x}_J)$, for $J = 1, \ldots, 10$; circles and squares denote the optimal combinations of $(x_{t-2} - \bar{x}_2, \ldots, x_{t-J} - \bar{x}_J)$ and $(x_{t-3} - \bar{x}_3, \ldots, x_{t-J} - \bar{x}_J)$, respectively. The results are for the case where $\pi = 0.975$ and $\lambda = 10$. We see a substantial gain in efficiency by constructing instruments as a combination of multiple lagged values of x_{t-j}. Omitting the first few lags of x_{t-j} is fairly innocuous when J is sufficiently large.

$\lambda = 10$. This configuration is motivated by our empirical analysis that is presented in Section 4.

All of these estimators are linear combinations of $Z_t = (x_{t-1} - \bar{x}_1, \ldots, x_{t-10} - \bar{x}_{10})'$, so the asymptotic variance is simply given by

$$\text{avar}(\hat{\pi}_{\alpha' Z_t}) = \frac{\alpha' M_{\pi,\lambda} \alpha}{(\alpha' V_\pi)^2}.$$

The line with the crossed symbols presents the asymptotic variance of the single-variable instrument, $\hat{\pi}_{\text{IV}_j}$, which has $z_t = x_{t-j} - \bar{x}_j$, for $j = 1, \ldots, 10$. In the notation given above, these estimators correspond to α-vectors that have one nonzero element, e.g., $\alpha = (1, 0, \ldots, 0)'$.

We also present results for the estimator that optimally combines a subset of the 10 variables, specifically the optimal combination of the $J - j + 1$ instruments $(x_{t-j} - \bar{x}_j, \ldots, x_{t-J} - \bar{x}_J)$, where $j = 1, 2, 3$ and $j \leq J \leq 10$. Obviously, the smallest asymptotic variance (in this class of estimators) is achieved by the estimator that has $j = 1$ and $J = 10$. However, by increasing j the estimator

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

becomes robust to measurement errors that are m-dependent, see, e.g., White (2000), so long as $m < j$. So we are interested in the loss of efficiency by dropping the first few lags as instruments. We are also interested in the marginal gains from increasing J. The reason is that theory suggests that J be as large as possible, but the practical implication of increasing J has drawbacks. It is therefore useful to know that the efficiency gain from increasing J beyond 10, say, is small.

Figure 1 shows that there are substantial gains from combining multiple instruments, but that the loss of efficiency by dropping the first few instruments is modest in this configuration once J is chosen large enough. Increasing J beyond 10 only leads to very minor gains in efficiency. So our preferred IV estimation in our empirical analysis will be the one that combines lags four through 10.

2.3. Properties of Estimators: The ARMA(p,q) Case

Consider now the case where y_t is an ARMA(p,q) and where we allow for a more general specification for the measurement errors.

THEOREM 4. *Suppose that Assumption 1 holds.* (*i*) *When* $\pi < 1$ *and* $j \geq \max(p, q)$, *we have*

$$\hat{\pi}_{\mathrm{IV}_j} \xrightarrow{p} \frac{\gamma(j+1)}{\gamma(j)}, \quad \text{as } n \to \infty,$$

where $\gamma(h) = \mathrm{cov}(y_t, y_{t+h})$, $h = 0, 1, \ldots$. (*ii*) *When* $\pi = 1$ *we have* $\hat{\pi}_{\mathrm{IV}_j} \xrightarrow{p} 1$, *as* $n \to \infty$, *for all* $j \geq 0$, *and for* $j \geq \max(p, q)$, *we have*

$$n(\hat{\pi}_{\mathrm{IV}_j} - \pi) \xrightarrow{d} \frac{\int_0^1 (W_u - \bar{W}) \mathrm{d}W_u}{\int_0^1 (W_u - \bar{W})^2 \mathrm{d}u}.$$

Previously, we established the consistency of $\hat{\pi}_{\mathrm{IV}_j}$ for π when $p = 1$, $y_t \sim$AR(1), and the consistency holds in general in the unit root case $\pi = 1$. When $p \geq 2$ but $\pi < 1$, we see that the IV estimator, $\hat{\pi}_{\mathrm{IV}_j}$, is consistent for $\gamma(j+1)/\gamma(j)$. So the question is whether this ratio is related to the persistence parameter π. This is addressed next, where we recall that a matrix is said to be positive if all its entries are positive.

LEMMA 3. *Suppose that* Φ^m *is positive for some integer* m *and* $\pi < 1$. *Then* $\frac{\gamma(j+1)}{\gamma(j)} \to \pi$ *as* $j \to \infty$.

The lemma is a consequence of the Perron-Frobenius theorem, and the result shows, in conjunction with Theorem 4, that $\hat{\pi}_{\mathrm{IV}_j} \xrightarrow{d} \pi$ as $n, j \to \infty$. The convergence $\frac{\gamma(j+1)}{\gamma(j)} \to \pi$ occurs at a fast exponential rate that is defined by the second largest eigenvalue of Φ, which suggests that j does not have to be very large in practice. The assumption that Φ^m is positive for some m is stronger than necessary, yet it is a reasonable assumption for the type of time series we consider in

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

this paper. The assumption rules out cases where the largest eigenvalue is negative, which could induce cyclical behavior in the autocorrelation function.

2.4. Multivariate Extension

In some cases it may be desirable to estimate all of the autoregressive parameters, $(\varphi_1, \ldots, \varphi_p)$, simultaneously. This is possible with multivariate IV estimators, $\hat{\varphi}_{\mathrm{IV}Z} = \left(\sum_{t=1}^{n} Z_t X_t'\right)^{-1} \sum_{t=1}^{n} Z_t x_{t+1}$, where $X_t = (x_t, x_{t-1}, \ldots, x_{t-p+1})'$ and $Z_t = (x_{t-j} - \bar{x}_j, \ldots, x_{t-j-p+1} - \bar{x}_{j-p+1})'$. This estimator is known as the Yule-Walker estimator, as it simply relies on the Yule-Walker equations. It is simple to verify that $\hat{\varphi}_{\mathrm{IV}Z} \xrightarrow{p} (\varphi_1, \ldots, \varphi_p)'$ for $j \geq \max(p, q)$ when Assumption 1 holds. Naturally, it is also possible to use a TSLS estimator that utilizes more than p instrumental variables.

It is, however, our experience that the multivariate IV estimators are very sensitive to measurement errors. When measurement errors are of the magnitude we have in our empirical application with realized volatility measures, it appears that an extremely large sample size is needed in order to get reliable estimates of all autoregressive parameters when $p \geq 2$. Since we are mainly concerned with the persistence parameter and the autocorrelation function, we do not pursue these multivariate estimators further.

3. AN APPROXIMATE AUTOCORRELATION FUNCTION OF A LATENT TIME SERIES

In this section we introduce an approximate estimator of the autocorrelation function that is based on a variant of the IV estimator we studied in the previous section. It is well known that measurement errors cause the population ACF of the observed process to look different from that of the underlying time series. With simple measurement errors, the autocorrelations of the observed process are simple those of the underlying process, scaled by a constant. In the context of realized measures, this has been noted in Taylor (2005, p. 337).

Before we define the new approximate autocovariance function for the latent y-process, we define the traditional empirical autocovariances for the latent process, y, and the observed process, x.

If y_t is observed, then we can estimate the autocorrelations by the empirical autocorrelations that are defined by

$$\mathrm{ACF}_y(h) = \frac{\sum_{t=1}^{n} y_{t+h}(y_t - \bar{y}_0)}{\sum_{t=1}^{n} y_t(y_t - \bar{y}_0)}, \qquad h = 0, 1, \ldots.$$

The probability limit $\rho_y(h) = \mathrm{plim}_{n\to\infty} \mathrm{ACF}_y(h)$ is well defined whether $\pi = 1$ or $|\pi| < 1$. Naturally, $\mathrm{ACF}_y(h)$ is simply the least squares estimator in the regression

$$y_{t+h} = \rho_y(h) y_t + \mu_y + u_{t,h},$$

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

where μ_y is a constant. The corresponding regression equation for the observed time series is

$$x_{t+h} = \rho_x(h)x_t + \mu_x + v_{t,h},$$

where the errors-in-variables problem will cause the least squares estimator of $\rho_x(h)$, defined by

$$\mathrm{ACF}_x(h) = \frac{\sum_{t=1}^n x_{t+h}(x_t - \bar{x}_0)}{\sum_{t=1}^n x_t(x_t - \bar{x}_0)}, \qquad h = 0, 1, \ldots,$$

to be inconsistent for $\rho_y(h)$. In the presence of measurement errors, it is therefore tempting to estimate $\rho_y(h)$ using an IV estimator, analogous to the way we have estimated the persistence parameter π. Consider

$$\mathrm{ACF}_x^*(h) = \frac{\sum_{t=1}^n z_t x_{t+h}}{\sum_{t=1}^n z_t x_t},$$

where the instrument z_t could be a lagged value of $x_{t-j} - \bar{x}_j$ or a linear combination of such $z_t = Z_t'\alpha$ with $Z_t = (x_{t-j} - \bar{x}_j, \ldots, x_{t-J} - \bar{x}_J)'$. The latter corresponds to a TSLS estimator, where α is determined in the first stage by regressing x_t onto Z_t. In our empirical application, we use $j = 4$ and $J = 10$.

We refer to $\mathrm{ACF}_x^*(h)$, $h = 1, 2, \ldots$, as the *approximate autocorrelation* function, where the nomenclature " approximate" is due to the fact that $\mathrm{ACF}_x^*(h)$ need not be consistent for $\rho_y(h)$ in general. The potential inconsistency can be understood from the following simple example. Suppose that y_t is an AR(2) process with $|\pi| < 1$ and $z_t = x_{t-j} - \bar{x}_j$. Then $\mathrm{ACF}_x^*(h) \xrightarrow{p} \gamma(h+j)/\gamma(j)$, which need not equal $\rho(h) = \gamma(h)/\gamma(0)$. The ratio of the two equals $\gamma(h+j)/\gamma(h)$ divided by $\rho(j)$, which will be close to one for many persistent processes. For instance, if y_t is an AR(1) process with $\pi \neq 0$, then $\gamma(h+j)/\gamma(j) = \gamma(h)/\gamma(0) = \pi^h$, making $\mathrm{ACF}_x^*(h)$ consistent for $\rho_y(h)$ in this special case. More generally, we have that $z_t = Z_t'\alpha$ with $Z_t = (x_{t-j} - \bar{x}_j, \ldots, x_{t-J} - \bar{x}_J)'$ and $j \geq \max(p, q)$ that $\mathrm{ACF}_x^*(h)$ converges to $\sum_j^J \alpha_i \gamma(i+h)/\sum_j^J \alpha_i \gamma(i)$ in probability. A drawback of the approximate ACF is the fact that its probability limit depends on the choice of instruments. Fortunately, the approximation error will be small when $\gamma(h+i)/(\gamma(h)\rho(i))$ is close to unity, which is the case for persistent processes.

We illustrate the merits of the approximate ACF by considering two persistent AR(2) processes, measured with error. Specifically, we consider the process

$$y_t = \varphi_\bullet(\tfrac{3}{4}y_{t-1} + \tfrac{1}{4}y_{t-2}) + \varepsilon_t,$$

where $\varepsilon_t \sim \mathrm{iid}\, N(0,1)$ and where $\varphi_\bullet$ is either $\varphi_\bullet = 1$ (unit root), $\varphi_\bullet = 1 - \frac{c}{n}$ with $c = 20$ (local to unit root), or $\varphi_\bullet = 0.975$ (stationary). In the local-to-unit design, this translates into $\pi = 0.936$, $\pi = 0.984$, and $\pi = 0.998$, for $n = 250$, $n = 1,000$, and $n = 10,000$, respectively, and in the stationary design $\varphi_\bullet = 0.975$ translates into $\pi = 0.980$.

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

We consider the ACFs for both y_t and $x_t = y_t + \eta_t$, where $\eta_t \sim \text{iid}\, N(0, \sigma_\eta^2)$ with $\sigma_\eta^2 = 4$. The corresponding results for $\sigma_\eta^2 = 1$ and $\sigma_\eta^2 = 10$ are in the Supplementary Appendix, which is available upon request. The simulation results reported here are based on 1,000 simulations, where the initial values were set to $y_0 = y_{-1} = 0$.

The solid thin black lines in Figure 2 represent the population ACF for the underlying process. Naturally, for the unit root process in the left panels, the population autocorrelation is constant and equal to one. The solid thick gray lines represent the empirical ACF, ACF_x, which is computed with the observed time series, x_t, and the short-dashed lines represent the new approximate ACF, ACF_x^*. For comparison we also include the empirical ACF, ACF_y, which is computed with underlying y-process (long-dashed lines). The difference between the empirical ACF_y and the population ACF reflects the well-known bias that vanishes as $n \to \infty$. The empirical autocorrelations reported are the averages over 1,000 simulations.

For the unit root case in the left-hand panels, it is interesting to note that the (infeasible) empirical ACF_y, which is computed with the (in practice unobserved)

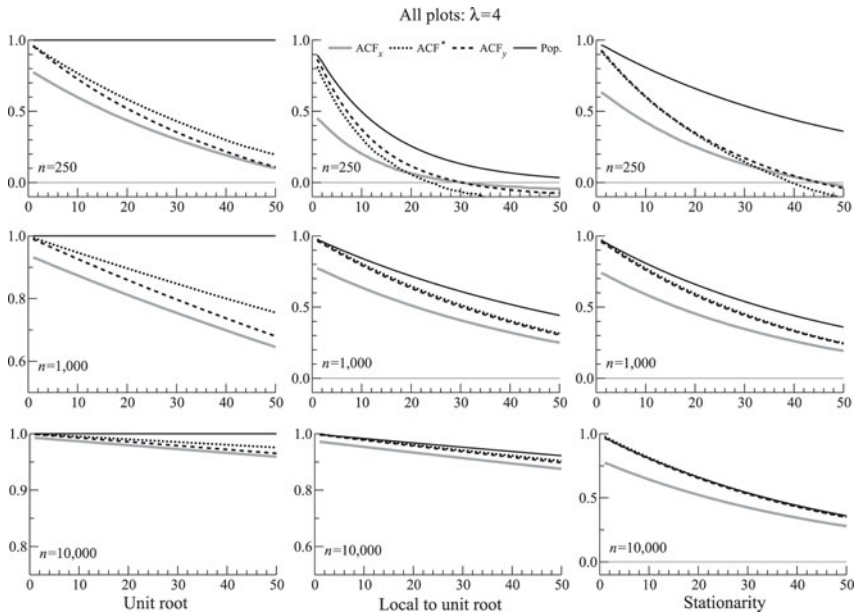

FIGURE 2. The solid thin black lines represent the population autocorrelation function for the underlying autoregressive y-process (the object of interest). The solid thick gray lines are the empirical autocorrelation functions for the observed x-process, ACF_x. The short-dashed line is the approximate autocorrelation function, ACF_x^*, also computed with the observed x-process. For comparison we include ACF_y, which is the (infeasible) empirical acf for the y-process (long-dashed lines).

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

y_t process, is more biased than the approximate ACF. This result is even true with much higher levels of measurement errors (see the Supplementary Appendix for the corresponding results with $\sigma_\eta^2 = 10$). For the local-to-unit root process we see that ACF_x^* is nearly as good as the infeasible ACF_y, in particular for large sample sizes, and it clearly dominates the traditional ACF_x.

4. EMPIRICAL ANALYSIS OF REALIZED MEASURES, INFLATION, AND ABSOLUTE RETURNS

We present empirical results using three data sets. The first data set consists of realized measures of volatility for stocks in the DJIA (as of medio 2008). We consider both the realized variance and the realized kernel. The second data set consists of two macroeconomic variables measuring U.S. inflation. The third data set consists of two long time series of absolute returns.

We present two types of results. For the realized measures of volatility, we estimates the persistence parameter, π, and compute the unit root tests based on both OLS and IV estimates. We compute the approximate ACFs for all time series in the three data sets and compare them to the empirical ACF for the observed time series.

High-frequency-based estimators of volatility are far more precise of volatility than squared returns, which is valuable for a number of reasons. For the purpose of evaluating volatility models, Andersen and Bollerslev (1998) documented that the realized variance strongly dominates squared returns when generalized autoregressive conditional heteroskedasticity (GARCH) models are evaluated with Minzer-Zarnowitch regressions; see also Hansen and Lunde (2005a). On a related issue, Hansen and Lunde (2006) have shown that a precise proxy of the latent volatility is critical for the empirical ranking of volatility models to be consistent for the population ranking unless the loss function has certain properties. See also Patton (2011) and Patton and Sheppard (2009), who provide further insight about this issue.

A drawback of the realized variance (RV) is that it is sensitive to market microstructure noise, and it is therefore not sensible to compute the RV with ultra-high-frequency returns such as tick-by-tick returns. There are several robust estimators that can utilize the entire database, including those by Zhang, Mykland, and Aït-Sahalia (2005), Barndorff-Nielsen et al. (2008, 2009), and Hansen and Horel (2009). However, even these estimators are in practice found to have a sizable sampling error, as is evident from the confidence intervals that are reported for these estimators; see, e.g., Barndorff-Nielsen et al. (2008) and Hansen and Horel. Any of these point estimates should therefore not be taken to be the true volatility. The sampling error must be accounted for when the objective is to learn about the dynamics of actual volatilities. Naturally, the measurement error can have other sources besides sampling error, such as those induced by market microstructure noise that has not been accounted for properly. The measurement error will be different for different realized measures, but the conclusions that one

draws about y_t from each of them should not contradict one another because they are all proxies for the same latent variable.

The importance of accounting for the measurement errors in the observed realized variances have been stressed by Andersen, Bollerslev, Diebold, and Labys (2003) and Andersen, Bollerslev, and Meddahi (2005). They emphasize the downward bias in the R^2 that measurement errors induce in predictive regressions. See also Meddahi (2001), Meddahi (2002), Andersen et al. (2004), and Andersen, Bollerslev, and Meddahi (2011).

4.1. Data Description

4.1.1. Realized Measures of Volatility. We analyze daily volatility estimates primarily based on high-frequency asset prices for 29 assets in the DJIA. The sample period runs from January 3, 2002, to July 31, 2009, with a total of 1,907 trading days for most of the series. The high-frequency data used to compute the realized measures of volatility are the trades and quotes recorded on the NYSE. These high-frequency data were extracted from the TAQ database through the Wharton Research Data Services (WRDS) system. Both the realized kernel and the realized variance are computed with transaction prices. However, quote prices are being used to clean the transaction data for anomalies. We follow the step-by-step cleaning procedure proposed by Barndorff-Nielsen et al. (2009).

We did not include realized measures that were computed with high-frequency data that spanned less than six hours. For each of the assets there were about 18 such days, primarily days where the market closed at noon, such as the day after Thanksgiving. These data were removed in order to eliminate obvious outliers that would arise from realized measures that correspond to just half a day of volatility. However, removing these data points barely affected any of our estimates.

4.1.2. Other Data: Macroeconomic Time Series and Absolute Return Series. We present the approximate autocorrelation functions for two time series of inflation and two time series of absolute returns. The two time series of inflation are derived from the Consumer Price Index for All Urban Consumers: All Items (CPIAUCSL). It is a monthly time series drawn from the data available online in Federal Reserve Economic Data (FRED) at the Federal Reserve Bank of St. Louis. The sample period we consider is 1947:M1 to 2010:M11; shorter samples periods (1959:M1 to 1997:M9) have previously been analyzed in Stock and Watson (1999) and Hansen, Lunde, and Nason (2011). The two time series of absolute returns are based on daily returns on the DJIA and the Standard & Poor's (S&P) 500 downloaded from Yahoo Finance. These span the period October 1, 1928, to September 16, 2009, and January 3, 1950, to September 16, 2009, respectively.

4.2. The Persistence of the Underlying Volatility

In this section we estimate the persistence parameter, π, using the IV estimators that were introduced in Section 2. The persistence parameter has been estimated in

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

earlier work. For instance, Maheu and McCurdy (2002) estimated an ARMA(1,1) model with a daily time series of the realized variance for an exchange rate, and they reported $\pi = 0.9$. Meddahi (2003) deduced the value, $\pi = 0.95$, from Bollerslev and Zhou (2002), who estimated stochastic volatility diffusions using empirical moments of the integrated variance. Interestingly, Bollerslev and Zhou did mention the potential use of IVS in the context of additive measurement errors, but they did not pursue this estimation strategy. In an application with time-varying betas that were computed from daily returns, Ghysels and Jacquier (2006) used the first lag as an IV to estimate an augmented AR(1) model for the noisy time series of betas.

We use $y_t = \log \mathrm{IV}_t$ and $x_t = \log \mathrm{RM}_t$, where $\mathrm{IV}_t = \int_{t-1}^{t} \sigma_s^2 \mathrm{d}s$ is the integrated variance and RM_t is a corresponding realized measure for day t. We prefer the logarithmic transformed variables for two reasons. First, it gets around the problem that the ARMA model does not prevent volatility from being negative. Second, in the Barndorff-Nielsen and Shephard framework the asymptotic variance of $\mathrm{RV}_t - \mathrm{IV}_t$ is proportional to $\mathrm{IQ}_t = \int_{t-1}^{t} \sigma_s^4 \mathrm{d}s$, whereas the asymptotic variance of $\log \mathrm{RV}_t - \log \mathrm{IV}_t$ is proportional to $\mathrm{IQ}_t / \mathrm{IV}_t^2$. The latter varies less with t, so the log-transformation leads to less heteroskedasticity.

In Tables 1–2 we present least squares and IV estimates of the persistence parameter. The estimates in Table 1 are those for the time series with realized kernel estimates. For each of the 29 assets, we compute the least squares estimator and eight IV estimators. The first four IV estimates are single-variable instruments based on the instrument $z_t = x_{t-j} - \bar{x}_j$, where $j = 1, \ldots, 4$. The last four IV estimates are based on multiple lags of the observed process, $x_{t-j} - \bar{x}_j, \ldots, x_{t-10} - \bar{x}_{10}$, $j = 1, \ldots, 4$, and these are computed with the procedure described in Section 2.2.1.

The least squares estimates in the first column are 10–20% smaller than the IV estimators in most cases. This shows that stochastic bias is important despite the large sample size. Thus, interpreting the asymptotic result that the stochastic bias is of order $O_p(n^{-1})$ to mean that this bias is negligible is very misleading in this context. We see that the persistence parameter tends to be large when the first few lags of x_{t-j} are not used as instruments. This indicates some serial dependence in the measurements errors that a larger value of j offers robustness to. So $j = 1$ (and $j = 2$ in some cases) appears to be too small to properly account for the noise. The multiple IV estimators with $j = 3$ and $j = 4$ are very similar and close to one in all cases, which shows that the underlying time series is highly persistent and close to unit root.

In Table 2 we present empirical results that are analogous to those in Table 1. The only difference is that these are based on the realized variance computed with 30-minute returns instead of the realized kernel. The realized variance is expected to be a less accurate estimator of the quadratic variation than the realized kernel, which translates into a larger measurement error variance. This is indeed found to be the case, because the bias of the least squares estimator is about twice as large as that we observed with the realized kernel. The IV estimates based on the

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

TABLE 1. Autoregressive persistence parameter, π . $x = \log(RK_t)$

		Single-variable IV				*Multiple-variables IV*			
	OLS	IV_1	IV_2	IV_3	IV_4	$IV_{1:10}$	$IV_{2:10}$	$IV_{3:10}$	$IV_{4:10}$
AA	0.854	0.966	0.976	0.986	0.981	0.977	0.983	0.985	0.986
AXP	0.926	0.988	0.986	0.987	0.991	0.989	0.990	0.991	0.993
BA	0.832	0.960	0.972	0.987	0.994	0.976	0.984	0.988	0.989
BAC	0.942	0.978	0.985	0.992	0.994	0.984	0.989	0.992	0.993
C	0.938	0.978	0.985	0.991	0.987	0.983	0.988	0.990	0.989
CAT	0.845	0.949	0.975	0.987	0.994	0.971	0.983	0.987	0.987
CVX	0.850	0.953	0.967	0.984	0.989	0.968	0.977	0.981	0.980
DD	0.847	0.959	0.973	0.988	0.986	0.975	0.982	0.985	0.984
DIS	0.864	0.959	0.976	0.990	0.991	0.975	0.984	0.988	0.987
GE	0.902	0.975	0.984	0.992	0.991	0.984	0.989	0.991	0.991
GM	0.865	0.952	0.984	0.986	0.984	0.974	0.988	0.991	0.996
HD	0.850	0.964	0.968	0.992	0.989	0.977	0.983	0.988	0.986
HPQ	0.819	0.943	0.967	0.979	0.985	0.965	0.978	0.983	0.984
IBM	0.857	0.968	0.970	0.992	0.993	0.978	0.983	0.986	0.984
INTC	0.858	0.945	0.962	0.998	0.993	0.964	0.980	0.990	0.984
JNJ	0.827	0.944	0.982	0.986	0.982	0.970	0.984	0.986	0.987
JPM	0.927	0.972	0.982	0.993	0.989	0.981	0.987	0.991	0.990
KO	0.836	0.965	0.967	0.984	0.991	0.976	0.981	0.985	0.985
MCD	0.751	0.933	0.955	0.988	0.976	0.963	0.977	0.986	0.986
MMM	0.809	0.946	0.946	0.994	0.967	0.962	0.972	0.982	0.978
MRK	0.751	0.899	0.980	0.972	0.980	0.946	0.977	0.976	0.979
MSFT	0.866	0.965	0.977	0.990	0.981	0.978	0.984	0.986	0.985
PFE	0.793	0.935	0.942	0.991	1.002	0.959	0.975	0.991	0.989
PG	0.800	0.928	0.964	0.973	0.992	0.955	0.973	0.979	0.981
T	0.832	0.940	0.958	0.978	0.984	0.959	0.973	0.983	0.986
UTX	0.825	0.955	0.967	0.972	0.988	0.969	0.976	0.980	0.983
VZ	0.848	0.962	0.971	0.982	0.989	0.975	0.982	0.987	0.989
WMT	0.813	0.952	0.961	0.991	0.980	0.970	0.979	0.986	0.985
XOM	0.850	0.954	0.967	0.987	0.985	0.969	0.977	0.982	0.978

Point estimates of the persistence parameter π . The first column contains the least squares estimator. The next four columns are IV estimates based on single-variable instruments: $x_{t-2} - \bar{x}_2, \ldots, x_{t-5} - \bar{x}_5$, respectively. The next four columns are estimates using multiple-instrumental variables, $x_{t-i} - \bar{x}_i, \ldots, x_{t-10} - \bar{x}_{10}$, for $i = 2, \ldots, 5$.

realized variance are strikingly similar to those we obtained with the realized kernel. This is further evidence that the latent volatility is highly persistent and close to being a unit root process. The average difference between the estimates based on the realized kernel and those based on the realized variance are reported in the last row of Table 2, where the average is taken over all assets. We observed that the OLS estimator based on the realized variance is, on average, 0.127 smaller than that based on the realized kernel. In contrast, the IV estimates are in agreement in all cases. These observations reflect that the IV estimators are estimating

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

TABLE 2. Autoregressive persistence parameter, π . $x = \log(RV_t)$

		Single-variable IV				*Multiple-variables IV*			
	OLS	IV_1	IV_2	IV_3	IV_4	$IV_{1:10}$	$IV_{2:10}$	$IV_{3:10}$	$IV_{4:10}$
AA	0.722	0.953	0.974	0.991	0.982	0.978	0.984	0.986	0.987
AXP	0.853	0.982	0.990	0.989	0.984	0.990	0.991	0.992	0.993
BA	0.659	0.967	0.969	0.965	0.985	0.980	0.984	0.989	0.994
BAC	0.887	0.964	0.985	0.990	0.993	0.980	0.990	0.992	0.993
C	0.881	0.975	0.991	0.990	0.985	0.986	0.990	0.990	0.990
CAT	0.701	0.976	0.948	0.999	0.983	0.981	0.983	0.987	0.985
CVX	0.698	0.948	0.971	0.965	0.976	0.970	0.975	0.978	0.982
DD	0.716	0.941	0.998	0.989	0.965	0.978	0.987	0.984	0.985
DIS	0.746	0.953	0.980	0.983	1.002	0.980	0.988	0.990	0.991
GE	0.820	0.973	0.979	0.986	0.992	0.985	0.988	0.991	0.993
GM	0.758	0.945	0.966	0.974	0.983	0.973	0.985	0.993	1.001
HD	0.706	0.976	0.952	0.999	0.976	0.981	0.984	0.988	0.987
HPQ	0.670	0.927	0.952	0.967	0.971	0.960	0.972	0.979	0.983
IBM	0.708	0.973	0.968	0.999	0.982	0.982	0.984	0.985	0.984
INTC	0.725	0.935	0.971	0.994	0.977	0.970	0.982	0.984	0.983
JNJ	0.715	0.936	0.978	0.977	0.965	0.969	0.980	0.982	0.985
JPM	0.842	0.969	0.979	0.991	0.989	0.983	0.988	0.991	0.991
KO	0.708	0.967	0.959	0.989	0.990	0.980	0.984	0.988	0.989
MCD	0.626	0.896	0.966	0.985	0.969	0.956	0.979	0.984	0.986
MMM	0.639	0.931	0.948	1.027	0.930	0.968	0.979	0.984	0.977
MRK	0.611	0.895	0.991	0.955	0.975	0.957	0.976	0.973	0.980
MSFT	0.750	0.955	0.986	0.977	0.976	0.978	0.983	0.984	0.986
PFE	0.651	0.910	0.916	1.004	1.037	0.958	0.981	0.999	0.993
PG	0.639	0.916	0.939	0.997	0.999	0.960	0.976	0.984	0.980
T	0.709	0.941	0.948	0.969	0.998	0.967	0.978	0.987	0.991
UTX	0.685	0.962	0.957	0.986	0.982	0.976	0.980	0.984	0.984
VZ	0.719	0.960	0.973	0.974	0.990	0.978	0.983	0.986	0.989
WMT	0.650	0.941	0.975	0.977	0.986	0.974	0.982	0.984	0.986
XOM	0.709	0.963	0.953	0.973	0.985	0.972	0.975	0.980	0.981
Average difference between π-estimates based on RK and RV									
Ave.	0.127	0.005	0.003	0.002	0.004	−0.003	−0.001	−0.000	−0.001

Point estimates of the persistence parameter π. The first column contains the least squares estimator. The next four columns are IV estimates based on a single-variable instruments: $x_{t-2} - \bar{x}_2, \ldots, x_{t-5} - \bar{x}_5$, respectively. The next four columns are estimates using multiple instrumental variables, $x_{t-i} - \bar{x}_i, \ldots, x_{t-10} - \bar{x}_{10}$, for $i = 2, \ldots, 5$. The last row displays the average difference (across assets) between the estimates obtained with the realized kernel and the realized variance.

the persistence of the same underlying time series, whereas the least squares estimators are affected by the variance of the measurement errors that is larger for the realized variance than the realized kernel.

The unit root test statistics, $n(\hat{\pi} - 1)$, that arise from our estimates of π using the realized kernel estimates are reported in Table 3. The first column is the

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

TABLE 3. Unit root test statistics: $n(\hat{\pi} - 1)$. $x = \log(RK_t)$

		Single-variable IV				*Multiple-variables IV*			
	OLS	IV_1	IV_2	IV_3	IV_4	$IV_{1:10}$	$IV_{2:10}$	$IV_{3:10}$	$IV_{4:10}$
AA	−276	−65.0	−46.0	−25.6	−35.8	−42.5	−32.2	−27.5	−26.5
AXP	−140	−23.5	−26.9	−25.2	**−16.3**	−21.2	**−19.3**	**−16.3**	**−13.3**
BA	−317	−75.2	−52.4	−25.1	**−12.0**	−44.2	−30.1	−22.3	−21.6
BAC	−110	−42.3	−27.9	**−16.0**	**−11.5**	−30.1	**−20.2**	**−14.6**	**−13.9**
C	−117	−42.4	−28.5	**−16.1**	−24.5	−31.1	−22.6	**−18.9**	**−20.0**
CAT	−293	−95.6	−46.4	−24.6	**−11.6**	−54.8	−31.3	−24.2	−25.1
CVX	−283	−89.1	−61.6	−30.7	−21.5	−60.6	−43.4	−34.9	−38.1
DD	−289	−76.8	−50.9	−23.0	−27.0	−47.7	−34.1	−27.8	−29.4
DIS	−257	−77.6	−46.0	**−19.1**	**−16.1**	−47.6	−30.2	−23.2	−25.2
GE	−185	−47.7	−30.4	**−15.1**	**−17.7**	−29.9	−21.1	**−17.4**	**−17.7**
GM	−247	−87.7	−28.6	−26.5	−29.0	−48.2	−22.2	**−16.5**	**−7.93**
HD	−282	−68.1	−60.0	**−14.6**	**−20.3**	−44.1	−32.3	−23.0	−26.0
HPQ	−326	−102	−58.6	−37.0	−27.5	−62.5	−39.8	−31.4	−29.6
IBM	−270	−59.7	−55.8	**−15.8**	**−13.6**	−40.8	−32.4	−25.9	−30.3
INTC	−268	−103	−71.4	**−3.92**	**−13.6**	−67.4	−38.2	**−18.7**	−29.9
JNJ	−326	−106	−34.8	−26.4	−33.1	−56.3	−29.2	−26.3	−25.2
JPM	−138	−52.5	−34.4	**−12.6**	**−20.1**	−36.5	−23.7	**−16.8**	**−19.5**
KO	−309	−65.8	−62.5	−30.9	**−17.8**	−45.5	−36.1	−27.6	−27.6
MCD	−469	−127	−84.6	−23.5	−44.4	−69.9	−42.3	−26.2	−25.6
MMM	−361	−103	−102	**−11.8**	−62.1	−72.3	−53.2	−34.0	−40.6
MRK	−469	−191	−37.2	−53.3	−38.0	−101	−42.4	−44.9	−40.2
MSFT	−254	−66.8	−42.7	**−19.2**	−35.0	−42.2	−30.5	−26.1	−27.8
PFE	−391	−123	−109	**−16.4**	**2.94**	−77.6	−46.3	**−17.5**	−20.9
PG	−378	−136	−68.1	−50.9	**−14.3**	−83.7	−50.1	−39.1	−35.1
T	−315	−113	−79.4	−41.4	−30.1	−77.0	−50.4	−31.9	−26.5
UTX	−330	−85.0	−62.3	−52.0	−23.3	−58.8	−45.5	−36.8	−31.5
VZ	−286	−72.4	−55.5	−33.3	**−20.7**	−46.5	−33.7	−24.6	−20.8
WMT	−352	−90.9	−73.2	**−16.2**	−36.8	−56.1	−38.7	−26.2	−28.9
XOM	−284	−86.6	−62.9	−23.9	−28.0	−58.9	−42.4	−34.3	−40.6

The 1% and 5% critical values are −20.7 and −14.1, respectively (see, e.g., (Fuller, 1996, Tab. 10.A.1, p. 641)). Test statistics in bold font are those that are insignificant at the 1% level.

traditional Dickey-Fuller t-statistic. These typically range between −150 and −400, which suggests overwhelming evidence against the unit root hypothesis. However, as we have seen earlier, the least squares estimates of the persistence parameter, π, are very biased, and this bias causes these test statistics to be misleading. The test statistics based on the IV estimates offer a more accurate picture of the evidence against the unit root hypotheses.

The test statistics in bold font are those for which we fail to reject the unit root hypothesis at the 1% level. While the unit root hypothesis is rejected for most

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

series, it is evident that the empirical evidence against the unit root hypothesis is less clear-cut than suggested by the OLS-based unit root tests. The conclusion we draw from our estimates of the persistence parameter is that the underlying process is highly persistent. This may be attributed to the underlying process being local to unit root, fractionally integrated, or some other form of persistent process. Naturally, the usual suspect, "structural change," cannot be ruled out either.

4.3. Empirical ACFs for Realized Measures

We consider two time series with estimates of daily volatility. The first time series is based on realized variance, computed with 30-minute returns, while the other is computed with the realized kernel estimator implemented as detailed in Barndorff-Nielsen et al. (2009). Both the realized variance and the realized kernel are estimates of the quadratic variation. Thus, we can view both as noisy proxies of the same population quantity, the underlying quadratic variation. We compute the approximate autocorrelations ACF_x^* using the TSLS estimator discussed in Section 3, where seven lags, $x_{t-4}, \ldots, x_{t-10}$, are used as IVs.

Figure 3 Panel (a) displays the estimated ACFs for realized measures of volatility computed with high-frequency returns on The Boeing Company (BA). The two approximate autocorrelation functions, ACF^*, based on the realized kernel and the realized variance, are similar and lie above the convention ACFs.
The fact that the two ACF^* estimates are similar is precisely what one would expect because the realized variance and the realized kernel are both estimates of the same underlying quantity, and the approximate ACF is designed to reflect the persistence of the latent time series. In contrast, we see that the conventional empirical ACFs for the realized kernel and realized variance produce distinctly different estimates. The difference between the two ACF_xs simply reflects that the measurement errors in the two series are different. Figure 3 Panel (b) gives another example of the estimated ACFs for another asset, Merck & Co., Inc. (MRK), and the results are very similar; the same applies for all other assets.

In a survey of the existing literature, Taylor (2005) noted that the ACF for the realized variance is typically estimated to be between 0.60 and 0.65 (for the autocorrelation of order one) with a slow decay for higher-order autocorrelations, precisely as is the case for the conventional autocorrelation estimates for the realized variance in Figure 3. Taylor (Sec. 12.9.4) discusses the downward bias that measurement errors induce on ACF_x, and he speculates that the first-order autocorrelation of the underlying volatility may be 0.70 or larger. We estimate the first-order autocorrelation of the underlying volatility to be very close to unity in all cases, so the downward bias is far more severe than may have been thought.

The traditional ACFs suggest that the realized kernel is somewhat more persistent than the realized variance for both BA and MRK. A point we want to emphasize here is that this discrepancy between the two ACFs is induced by the realized variance being a less accurate estimator of the latent volatility, and that neither of the conventional ACFs properly reflects the persistence of the

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

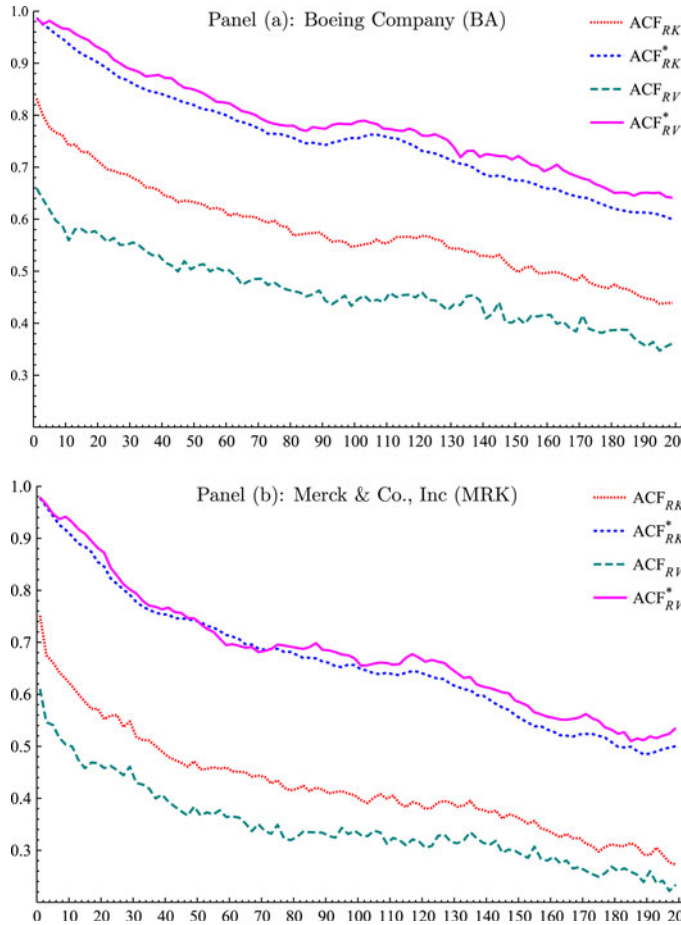

FIGURE 3. The empirical and approximate autocorrelation functions computed with realized measures of volatility for BA (top) and MRK (bottom). The approximate ACF* better reflects the autocorrelation function of latent volatility. While the conventional ACF is quite different, the two ACF* estimates are in agreement.

population measure of volatility. The persistence is better assessed with our approximate estimation of the autocorrelation function, ACF*, which produces very similar estimates for the realized kernel and the realized variance.

Given the results we reported in Figures 2 and 3, we see that the evidence against the unit root hypothesis is less clear-cut than suggested by the conventional ACF. Since the estimated autocorrelations are downward biased when the underlying population quantity is close to one, the estimated ACF*s are by no means strong evidence against the unit root hypothesis. When taking the finite sample bias and sampling error into account, the estimated ACF* could be

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

consistent with a unit root process, a fractionally integrated process, as well as many other types of processes.

4.4. Autocorrelation Functions for Time Series of Inflation

We consider two inflation series, the year-over-year inflation, π_t^{12}, and the month-over-month inflation, π_t^1. These are defined by $\pi_t^h = (1200/h)\ln(P_t/P_{t-h})$, so that both are measured at an annual rate. The inflation series are plotted in the top panel of Figure 4, where the year-over-year series naturally emerges as a smoothed version of the monthly series. In the bottom panel of Figure 4 we present the empirical autocorrelation function, ACF_x, and the approximate

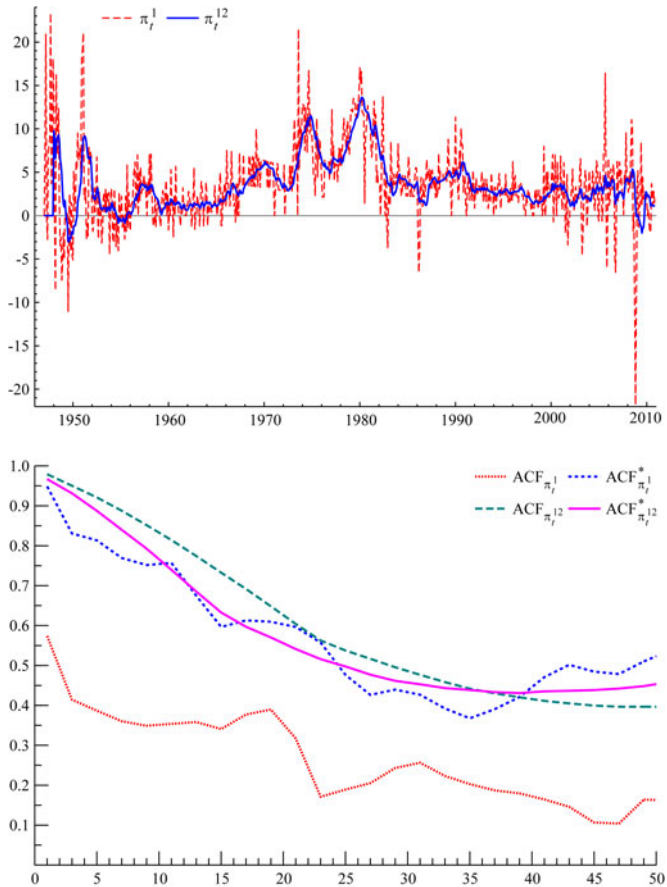

FIGURE 4. Top panel: Month-over-month inflation, π_t^1, and year-over-year inflation, π_t^{12}. Bottom panel: The conventional and approximate autocorrelation functions for the two inflation series. Note that the conventional autocorrelation function for the month-over-month inflation is severely influenced by the "noisy" features of this series.

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

estimator, ACF_x^*, for these two time series of inflation. As in the previous section we use the seven lags, $x_{t-4}, \ldots, x_{t-10}$, as instruments for computing ACF_x^*. We observe that the conventional ACF is distinctly lower for the month-over-month series, π_t^1, than that of year-over-year inflation, π_t^{12}. Also the ACF_x for π_t^1 is much lower than ACF_x^*. In contrast, the two ACFs for π_t^{12} are similar, which suggests that the measurement errors in this time series are relatively small. From the top panel of Figure 4, it is evident that both inflation series are more volatile toward the end of the sample. To ensure that our results are not driven by this feature, we have computed the ACFs using a pre-2007 sample, and the results (which are presented in the Supplementary Appendix) are quite similar. Again, ACF_x for the π_t^1 series is distinctly lower than the three other autocorrelation functions, which are virtually identical for the pre-2007 sample.

4.5. Empirical Analysis of Absolute Returns

Absolute returns are often used as an example of a process with properties that resemble those of a fractionally integrated process; see Ding, Granger, and Engle (1993). Squared returns are a simple one-to-one transformation of absolute returns, so if absolute returns have long memory features, then so will squared returns. In fact, the ACFs for log-absolute returns and log-squared returns are identical. It is perhaps puzzling that the case of long memory is rarely made about squared returns, even though the order of fractional integration, d, is the same for absolute returns and squared returns; see, e.g., Andersen and Bollerslev (1997), Harvey (1998), and Bollerslev and Wright (2001). We believe that the explanation for this is that either series can be viewed as a noisy measurement of volatility, and that the noise is simply more pronounced in squared returns, which conceal the persistence to a larger extent than is the case for absolute returns. The eigenfunction analysis, see Meddahi (2001) and Andersen et al. (2011), provides the deeper theoretical explanation for this. For instance, in the context of stochastic volatility models, this phenomenon would naturally arise if the instantaneous volatility equals the first eigenfunction, as is the case in Forsberg and Ghysels (2007).

In Figure 5 we have computed the empirical ACF for absolute returns as well as the approximate ACF, and these are, as to be expected, distinctly different. The empirical ACF is often interpreted as evidence of long memory and sometimes considered to be evidence against a unit root hypothesis. The new estimation of the autocorrelation of the underlying process, ACF_x^*, reveals that the choice between long memory and unit root is less clear-cut than is suggested by the conventional ACF for the observed process, ACF_x.

Relying on the empirical ACF for the observed series is perhaps not the best way to classify the long-dependence properties of a time series, because it is influenced by the short-run dynamics. A better classification scheme of persistent processes is that of Muller and Watson (2008). Their method is explicitly designed to filter the effect of short-run dynamics and focus on the variation

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

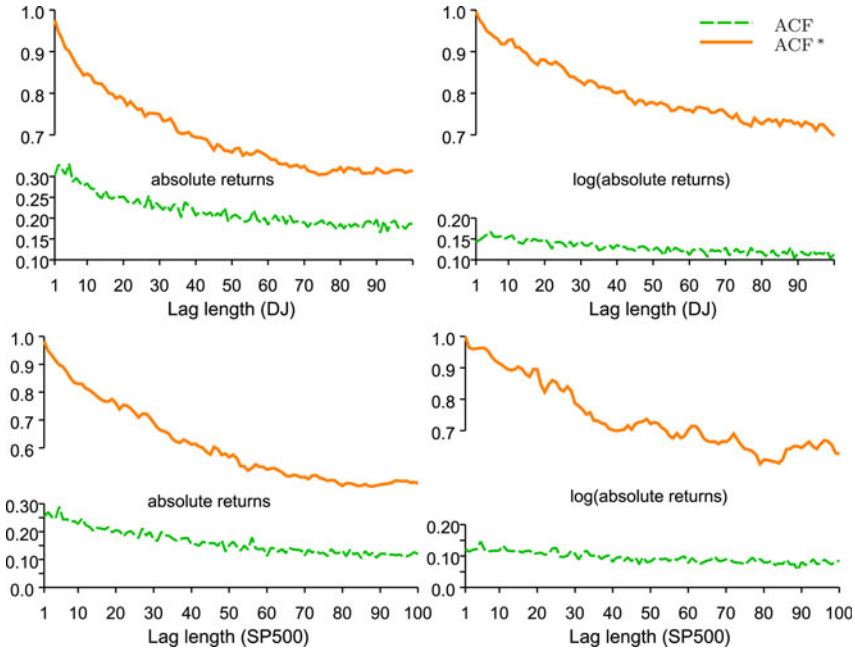

FIGURE 5. ACF for daily absolute return.

at low frequencies. In their empirical analysis of absolute returns, they do find empirical evidence that supports a fractionally integrated model for absolute returns.

Long-memory models that explicitly account for noise in the manner we discuss in this paper are a relatively unexplored topic. A few papers, such as Chong and Lui (1999), Sun and Phillips (2003), Hurvich, Moulines, and Soulier (2005), and Haldrup and Nielsen (2007), consider estimation of the memory parameter in an ARFIMA setting, where the time series of interest is perturbed by an additive noise term. In the context of volatility measures, Bollerslev and Wright (2000) showed that high-frequency-based volatility measures lead to more accurate estimates of the long-memory parameter. This is quite intuitive, because the use of realized volatility measures effectively amounts to reducing the measurement error.

5. SUMMARY AND CONCLUDING REMARKS

In a situation where a time series is observed with measurement errors, we have shown that the persistence of the underlying time series can be assessed by instrumental variable methods. When the latent time series is an ARMA(p,q) process, it is possible to estimate the autoregressive parameters consistently, hence

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

a consistent estimate of the persistence parameter. The instrumental variables we employ are lagged values of the observed time series, and we derived the optimal linear IV estimator in a special case.

Serial dependence makes the lagged values of the observed time series useful as instruments. So a highly persistent latent time series offers an ideal framework for the IV estimators. On the other hand, a time series with little autocorrelation causes the lagged values to be weak instruments, and the IV estimators may be unreliable in such circumstances.

We have also proposed a novel estimator of the autocorrelation function for the underlying time series. This estimator also relies on lagged values of the observed process being good and valid instruments. So this estimator is best suited for the case where the underlying time series is persistent.

We have shown that measurement errors can conceal the persistence of the underlying time series, and that unit root tests are unreliable unless the measurement errors are accounted for. So the empirical evidence against the unit root hypothesis may, in some cases, not have been as clear-cut as may have been believed. Our findings are also relevant for multivariate time series. For instance, the fact that a unit root process with measurement errors can be confused with a fractionally integrated process begs the following question: Can some of the fractional cointegration results that have been documented in the literature be attributed to measurement errors? Another possible explanation is that the underlying time series are individually integrated of order one and cointegrate in the traditional sense, but that measurement errors make the individual time series appear to be fractionally integrated. In any case, we believe it is important to account for measurement errors in applications with realized measures of volatility, and other persistent time series with similar levels of measurement errors. For instance, in ARFIMA modeling, most attention is often devoted to estimation of the order of (fractional) integration, while the MA part is largely neglected. Yet the moving average component deserves serious attention, because the extent of measurement errors can influence this part of the process in important ways.

NOTE

1. In the event that the estimator of π exceeds one we suggest to substitute 1 for $\hat{\pi}$ in the expression (7).

REFERENCES

Anatolyev, S. (2001) Optimal instrumentals in time series: A survey. *Journal of Economic Surveys* 21, 143–173.

Andersen, T.G. & T. Bollerslev (1997) Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *Journal of Finance* 52(3), 975–1005.

Andersen, T.G. & T. Bollerslev (1998) Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39(4), 885–905.

Andersen, T.G., T. Bollerslev, F.X. Diebold, & P. Labys (2003) Modeling and forecasting realized volatility. *Econometrica* 71(2), 579–625.

Andersen, T.G., T. Bollerslev, & N. Meddahi (2004) Analytic evaluation of volatility forecasts. *International Economic Review* 45, 1079–1110.

Andersen, T.G., T. Bollerslev, & N. Meddahi (2005) Correcting the errors: A note on volatility forecast evaluation based on high-frequency data and realized volatilities. *Econometrica* 73, 279–296.

Andersen, T.G., T. Bollerslev, & N. Meddahi, (2011) Market microstructure noise and realized volatility forecasting. *Journal of Econometrics* 160, 220–234.

Anderson, T. & H. Rubin (1949) Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics* 20, 46–63.

Ashley, R. & D. Vaughan (1986) Measuring measurement error in economic time series. *Journal of Business and Economic Statistics* 4, 95–103.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, & N. Shephard (2008) Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–1536.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, & N. Shephard (2009) Realised kernels in practice: Trades and quotes. *Econometrics Journal* 12, 1–33.

Barndorff-Nielsen, O.E., B. Nielsen, N. Shephard, & C. Ysusi (2004) Measuring and forecasting financial variability using realised variance with and without a model. In A.C. Harvey, S.J. Koopman, & N. Shephard (eds.), *State Space and Unobserved Components Models: Theory and Applications*', pp. 205–235. Cambridge University Press.

Barndorff-Nielsen, O.E. & N. Shephard (2002) Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society* B 64, 253–280.

Bollerslev, T. & J.H. Wright (2000) Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data. *Journal of Econometrics* 98, 81–106.

Bollerslev, T. & J.H. Wright (2001) High-frequency data, frequency domain inference, and volatility forecasting. *Review of Economics & Statistics* 83(4), 596–602.

Bollerslev, T. & H. Zhou (2002) Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics* 109, 33–65.

Chong, T.T. & G.C. Lui (1999) Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data. *Economics Letters* 63, 285–194.

Ding, Z., C.W.J. Granger, & R.F. Engle, (1993) A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.

Doornik, J.A. (2006) *Ox: An Object-Orientated Matrix Programming Langguage*, 5th ed. Timberlake Consultants Ltd.

Forsberg, L. & E. Ghysels (2007) Why do absolute returns predict volatility so well? *Journal of Financial Econometrics* 5, 31–67.

Fuller, W.A. (1996) *Introduction to Time Series*, 2nd ed. Wiley.

Ghysels, E. & E. Jacquier (2006) Market Beta Dynamics and Portfolio Efficiency. Working paper. Chapel Hill, NC: University of North Carolina, Department of Economics.

Granger, C.W.J. (1980) Long memory relationships and the aggregation of dynamic models. *Jounal of Econometrics* 14, 227–238.

Haldrup, N. & M.O. Nielsen (2007) Estimation of fractional integration in the presence of data noise. *Computational Statistics & Data Analysis* 51, 3100–3114.

Hall, A. (1989) Testing for a unit roor in the presence of moving average errors. *Biometrika* 76, 49–56.

Hansen, L.P. (1985) A method for calculating bounds on asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics* 30, 203–238.

Hansen, L.P., J.C. Heaton, & M. Ogaki (1988) Efficiency bounds implied by multiperiod conditional moment restrictions. *Journal of the American Statistical Association* 83, 863–871.

Hansen, L.P. & K.J. Singleton (1996) Efficient estimation of linear asset-pricing models with moving average errors. *Journal of Business & Economic Statistics* 14, 53–68.

Hansen, P.R. & G. Horel, (2009) Quadratic Variation by Markov Chains. CREATES Research Paper 2009–13.
Hansen, P.R. & A. Lunde (2005a) A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *Journal of Applied Econometrics* 20, 873–889.
Hansen, P.R. & A. Lunde (2005b) A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics* 3, 525–554.
Hansen, P.R. & A. Lunde (2006) Consistent ranking of volatility models. *Journal of Econometrics* 131, 97–121.
Hansen, P.R., A. Lunde, & J.M. Nason (2011) The model confidence set. *Econometrica* 79, 456–497.
Harvey, A.C. (1993) *Time Series Models*, 2nd ed. Harvester Wheatsheaf.
Harvey, A.C. (1998) Long memory in stochastic volatility. In J. Knight & S. Satchell (eds.), *Forecasting Volatility in Financial Markets*, pp. 307–320. Butterworth-Heineman.
Harvey, A.C. (2001) Testing in unobserved components models. *Journal of Forecasting* 20, 1–19.
Harvey, A.C. & T. Proietti (eds.) (2005) *Readings in Unobserved Components Models*. Oxford University Press.
Harvey, O. & G. De Rossi (2006) Signal extraction. In T.C. Mills & K. Patterson (eds.), *Palgrave Handbook of Econometrics*, vol. 1, Econometric Theory. Ch. 27, pp. 970–1000. Palgrave Macmillan.
Hurvich, C.M., E. Moulines, & P. Soulier (2005) Estimating long memory in volatility. *Econometrica* 73, 1283–1328.
Koopman, S.J., B. Jungbacker & E. Hol (2005) Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance* 12, 445–475.
Maheu, J.M. & T.H. McCurdy (2002) Nonlinear features of realized FX volatility. *Review of Economics & Statistics* 84, 668–681.
Meddahi, N. (2001) An Eigenfunction Approach for Volatility Modeling. Cirano Working Paper 2001s-70.
Meddahi, N. (2002) A theoretical comparison between integrated and realized volatilities. *Journal of Applied Econometrics* 17, 479–508.
Meddahi, N. (2003) ARMA representation of integrated and realized variances. *The Econometrics Journal* 6, 334–355.
Meddahi, N. & E. Renault (2004) Temporal aggregation of volatility models. *Journal of Econometrics* 119, 355–379.
Muller, U.K. & M.W. Watson (2008) Testing models of low-frequency variablility. *Econometrica* 76, 979–1016.
Patton, A. (2011) Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160, 246–256.
Patton, A.J. & K.K. Sheppard (2009) Evaluating volatility forecasts. In T.G. Andersen, R.A. Davis, J.P. Kreiss, & T. Mikosch (eds.), *Handbook of Financial Time Series*, pp. 801–838. Springer Verlag.
Perron, P. & S. Ng (1996) Useful modifications of some unit root tests with dependent errors and their local asymptotic properties. *Review of Economic Studies* 63, 435–463.
Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regression with I(1) processes. *Review of Economic Studies* 57, 99–125.
Schwert, G.W. (1989) Tests for unit roots: A Monte-Carlo investigation. *Journal of Business and Economic Statistics* 7, 147–160.
Stock, J.H. & M.W. Watson (1999) Forecasting inflation. *Journal of Monetary Economics* 44, 293–335.
Sun, Y. & P.C.B. Phillips (2003) Nonlinear log-periodogram regression for perturbed fractional processes. *Journal of Econometrics* 115, 355–389.
Taylor, S.J. (2005) *Asset Price Dynamics, Volatility, and Prediction*, Princeton University Press.
Watson, M.W. (1986) Univariate detrending methods with stochastic trends. *Journal of Monetary Economics* 18, 49–75.

West, K.D. (2001) On optimal instrumental variable estimation of stationary time series model. *International Economic Review* 42, 1043–1050.

White, H. (2000) *Asymptotic Theory for Econometricians*, rev. ed. Academic Press.

Wright, J.H. (1999) Testing for a unit root in the volatility of asset returns. *Journal of Applied Econometrics* 14, 309–318.

Zhang, L., P.A. Mykland, & Y. Aït-Sahalia (2005) A tale of two time scales: Determining integrated volatility with noisy high frequency data. *Journal of the American Statistical Association* 100, 1394–1411.

APPENDIX OF PROOFS

The proofs of Lemmas 1 and 2 are straightforward. See the Supplementary Appendix for details.

Proof of Theorem 1. The case with $|\pi| < 1$ follows from Lemma 2. When $\pi = 1$, we have $y_{[un]}/\sqrt{n} \xrightarrow{w} \sigma_\varepsilon W_u$, where W_u is a standard Brownian motion, $u \in [0,1]$, and we consider

$$\begin{aligned} n^{-1}\sum_{t=1}^{n}(x_t-\bar{x}_0)\left(\varepsilon_{t+1}+\eta_{t+1}-\eta_t\right) &= n^{-1}\sum_{t=1}^{n}(y_t-\bar{y}_0+\eta_t-\bar{\eta})\left(\varepsilon_{t+1}+\eta_{t+1}-\eta_t\right) \\ &= n^{-1}\sum_{t=1}^{n}(y_t-\bar{y}_0)\varepsilon_{t+1}-n^{-1}\sum_{t=1}^{n}(\eta_t-\bar{\eta})^2+n^{-1}\sum_{t=1}^{n}y_t\eta_{t+1}-n^{-1}\sum_{t=0}^{n-1}y_{t+1}\eta_{t+1} \\ &\quad -\bar{y}_0 n^{-1}\sum_{t=1}^{n}\left(\eta_{t+1}-\eta_t\right)+n^{-1}\sum_{t=1}^{n}(\eta_t-\bar{\eta})\left(\varepsilon_{t+1}+\eta_{t+1}\right), \end{aligned}$$

where

$$\begin{aligned} n^{-1}\sum_{t=1}^{n}y_t\eta_{t+1}-n^{-1}\sum_{t=0}^{n-1}y_{t+1}\eta_{t+1} &= -n^{-1}\sum_{t=0}^{n-1}(y_{t+1}-y_t)\eta_{t+1}+n^{-1}(y_n\eta_0-y_0\eta_0) \\ &= -n^{-1}\sum_{t=1}^{n}\varepsilon_t\eta_t+O_p(n^{-1})=o_p(1), \end{aligned}$$

and $\bar{y}_0 n^{-1}\sum_{t=1}^{n}\left(\eta_{t+1}-\eta_t\right)+n^{-1}\sum_{t=1}^{n}(\eta_t-\bar{\eta})\left(\varepsilon_{t+1}+\eta_{t+1}\right)=o_p(1)$.

Similarly, $n^{-2}\sum_{t=1}^{n}(x_t-\bar{x}_0)^2 = n^{-2}\sum_{t=1}^{n}(y_t-\bar{y}_0)^2+o_p(1)$, so that

$$\begin{aligned} n(\hat{\pi}_{\mathrm{LS}}-1) &= \frac{n^{-1}\sum_{t=1}^{n}(x_t-\bar{x}_0)\left(\varepsilon_{t+1}+\eta_{t+1}-\eta_t\right)}{n^{-2}\sum_{t=1}^{n}(x_t-\bar{x}_0)x_t} \\ &= \frac{n^{-1}\sum_{t=1}^{n}(y_t-\bar{y}_0)\varepsilon_{t+1}+n^{-1}\sum_{t=1}^{n}\eta_t^2-n^{-1}\sum_{t=1}^{n}\varepsilon_{t+1}\eta_{t+1}+o_p(1)}{n^{-2}\sum_{t=1}^{n}(y_t-\bar{y}_0)^2+o_p(1)} \\ &\xrightarrow{d} \left[\int_0^1(W_u-\bar{W})\mathrm{d}W_u-\sigma_\eta^2/\sigma_\varepsilon^2\right]\Big/\int_0^1(W_u-\bar{W})^2\mathrm{d}u. \end{aligned}$$

■

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

LEMMA A.1. *Let Σ_t be short for $n^{-1/2}\sum_{t=1}^{n}$. Given Assumption 1 with $p=1$, $q=0$, and $\pi<1$, we have, for $j\geq 1$ and $d\geq 0$,*

$$\operatorname{cov}\left(\Sigma_t x_{t-j}u_{t+1},\Sigma_t x_{t-j-d}u_{t+1}\right)\to\sigma_\varepsilon^4\times\begin{cases}\frac{1}{1-\pi^2}+2\lambda+(1+\pi^2)\lambda^2 & d=0,\\ \frac{\pi}{1-\pi^2}-\pi\lambda^2 & d=1,\\ \frac{\pi^d}{1-\pi^2} & d\geq 2,\end{cases}$$

as $n\to\infty$

Proof. Without loss of generality we set $\delta=\xi=0$. In our proof we use that

$$\Sigma_t y_{t-j}\eta_t=\pi\,\Sigma_t y_{t-j-1}\eta_t+\Sigma_t\varepsilon_{t-j}\eta_t=\pi\,\Sigma_t y_{t-j}\eta_{t+1}+\Sigma_t\varepsilon_{t-j}\eta_t+O_p(n^{-1/2}). \tag{A.1}$$

We have

$$\begin{aligned}\Sigma_t x_{t-j}u_{t+1} &= \Sigma_t(y_{t-j}+\eta_{t-j})\left(\varepsilon_{t+1}+\eta_{t+1}-\pi\eta_t\right)\\ &= \Sigma_t y_{t-j}\varepsilon_{t+1}+\Sigma_t\eta_{t-j}\varepsilon_{t+1}+\Sigma_t y_{t-j}\eta_{t+1}-\pi\,\Sigma_t y_{t-j}\eta_t\\ &\quad+\Sigma_t\eta_{t-j}(\eta_{t+1}-\pi\eta_t)\\ &\overset{(A.1)}{=} \Sigma_t y_{t-j}\varepsilon_{t+1}+\Sigma_t\eta_{t-j}\varepsilon_{t+1}\\ &\quad+\Sigma_t y_{t-j}\eta_{t+1}-\pi^2\Sigma_t y_{t-j}\eta_{t+1}-\pi\,\Sigma_t\varepsilon_{t-j}\eta_t\\ &\quad+\Sigma_t\eta_{t-j}(\eta_{t+1}-\pi\eta_t)+O_p(n^{-1/2})\\ &= \Sigma_t y_{t-j}\varepsilon_{t+1}+\Sigma_t\eta_{t-j}\varepsilon_{t+1}-\pi\,\Sigma_t\eta_t\varepsilon_{t-j}+(1-\pi^2)\Sigma_t y_{t-j}\eta_{t+1}\\ &\quad+\Sigma_t\eta_{t-j}(\eta_{t+1}-\pi\eta_t)+O_p(n^{-1/2})\\ &= \Sigma_t y_{t-j}\left\{\varepsilon_{t+1}+(1-\pi^2)\eta_{t+1}\right\}\\ &\quad+\Sigma_t\eta_{t-j+1}(\varepsilon_{t+1}+\eta_{t+1}-\pi\eta_t)-\pi\,\Sigma_t\varepsilon_{t-j}\eta_t+O_p(n^{-1/2}),\end{aligned}$$

which is a sum of asymptotically uncorrelated terms. So the asymptotic variance is

$$\begin{aligned}\operatorname{avar}(\Sigma_t x_{t-j}u_{t+1}) &= \sigma_y^2\left\{\sigma_\varepsilon^2+(1-\pi^2)^2\sigma_\eta^2\right\}+\sigma_\eta^2\left\{\sigma_\varepsilon^2+(1+\pi^2)\sigma_\eta^2\right\}+\pi^2\sigma_\eta^2\sigma_\varepsilon^2\\ &= \tfrac{\sigma_\varepsilon^4}{1-\pi^2}+(1-\pi^2)\sigma_\varepsilon^2\sigma_\eta^2+(1+\pi^2)\sigma_\eta^2\sigma_\varepsilon^2+(1+\pi^2)\sigma_\eta^4\\ &= \tfrac{\sigma_\varepsilon^4}{1-\pi^2}+2\sigma_\varepsilon^2\sigma_\eta^2+(1+\pi^2)\sigma_\eta^4\\ &= \sigma_\varepsilon^4\left\{\left(1-\pi^2\right)^{-1}+2\lambda+\left(1+\pi^2\right)\lambda^2\right\}.\end{aligned}$$

Next we analyze the covariances. The σ_ε^4-terms are given by

$$\begin{aligned}&\operatorname{cov}(\Sigma_t y_{t-j}\varepsilon_{t+1},\Sigma_t y_{t-j-d}\varepsilon_{t+1})\\ &\quad=\operatorname{cov}\left\{\Sigma_t\left(\pi^d y_{t-j-d}+\varepsilon_{t-j}+\cdots+\pi^{d-1}\varepsilon_{t-j-d+1}\right)\varepsilon_{t+1},\Sigma_t y_{t-j-d}\varepsilon_{t+1}\right\}\\ &\quad=\pi^d\operatorname{var}(\Sigma_t y_{t-j-d}\varepsilon_{t+1})\\ &\qquad+\operatorname{cov}\left\{\Sigma_t\left(\varepsilon_{t-j}+\cdots+\pi^{d-1}\varepsilon_{t-j-d+1}\right)\varepsilon_{t+1},\Sigma_t y_{t-j-d}\varepsilon_{t+1}\right\}\\ &\quad=\pi^d\operatorname{var}(\Sigma_t y_{t-j-d}\varepsilon_{t+1})\to\pi^d(1-\pi^2)^{-1}\sigma_\varepsilon^4,\quad\text{as } n\to\infty.\end{aligned}$$

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

The $\sigma_\eta^2\sigma_\varepsilon^2$-terms are given by $\Sigma_t\eta_{t-j}\varepsilon_{t+1} - \pi\,\Sigma_t\eta_t\varepsilon_{t-j} + (1-\pi^2)\Sigma_t y_{t-j}\eta_{t+1} = \Sigma_t\eta_{t-j}\varepsilon_{t+1} - \pi\,\Sigma_t\eta_t\varepsilon_{t-j} + (1-\pi^2)\Sigma_t\left(\pi^d y_{t-j-d} + \varepsilon_{t-j} + \cdots + \pi^{d-1}\varepsilon_{t-j-d+1}\right)\eta_{t+1}$ such that

$$\begin{aligned}
&\operatorname{cov}\Big\{\Sigma_t\eta_{t-j}\varepsilon_{t+1} - \pi\,\Sigma_t\eta_t\varepsilon_{t-j} + (1-\pi^2)\Sigma_t y_{t-j}\eta_{t+1},\, \Sigma_t\eta_{t-j-d}\varepsilon_{t+1} - \pi\,\Sigma_t\eta_t\varepsilon_{t-j-d} \\
&\quad + (1-\pi^2)\Sigma_t y_{t-j-d}\eta_{t+1}\Big\} \\
&\quad = (1-\pi^2)^2\pi^d\operatorname{var}(\Sigma_t y_{t-j-d}\eta_{t+1}) \\
&\qquad + \operatorname{cov}\Big\{-\pi\,\Sigma_t\eta_t\varepsilon_{t-j-d}, (1-\pi^2)\Sigma_t\pi^{d-1}\varepsilon_{t-j-d+1}\eta_{t+1}\Big\} \\
&\quad = (1-\pi^2)\pi^d\sigma_\eta^2\sigma_\varepsilon^2 - (1-\pi^2)\pi^d\sigma_\eta^2\sigma_\varepsilon^2 = 0.
\end{aligned}$$

Finally, when $d = 1$, the σ_η^4-term is simply

$$\begin{aligned}
&\operatorname{cov}(\Sigma_t\eta_{t-j}\eta_{t+1} - \pi\,\Sigma_t\eta_{t-j}\eta_t,\, \Sigma_t\eta_{t-j-1}\eta_{t+1} - \pi\,\Sigma_t\eta_{t-j-1}\eta_t) \\
&\quad = \operatorname{cov}(\Sigma_t\eta_{t-j}\eta_{t+1}, -\pi\,\Sigma_t\eta_{t-j-1}\eta_t) = -\pi\sigma_\eta^4 + o(1),
\end{aligned}$$

whereas this term is zero when $d \geq 2$. ■

Proof of Theorem 2. With an AR(1) specification for y_t, we have

$$x_t = \pi x_{t-1} + (1-\pi)(\delta+\xi) + \varepsilon_t + \eta_t - \pi\eta_{t-1}.$$

Without loss of generality, we set $\delta = \xi = 0$. By repeated substitution, $x_t = \pi^j y_{t-j} + \varepsilon_t + \pi\varepsilon_{t-1} + \cdots + \pi^{j-1}\varepsilon_{t-j+1} + \eta_t$, and it follows that $\operatorname{cov}(x_t, x_{t-j}) = \pi^j\operatorname{cov}(y_{t-j}, x_{t-j}) = \pi^j\operatorname{var}(y_{t-j}) = \pi^j\sigma_y^2$. Next, we consider the decomposition

$$\hat\pi_{\mathrm{IV}_j} = \frac{\sum_{t=1}^n (x_{t-j}-\bar x_j)x_{t+1}}{\sum_{t=1}^n (x_{t-j}-\bar x_j)x_t} = \pi + \frac{\sum_{t=1}^n (x_{t-j}-\bar x_j)\left(\varepsilon_{t+1}+\eta_{t+1}-\pi\eta_t\right)}{n^{-1}\sum_{t=1}^n (x_{t-j}-\bar x_j)x_t}.$$

From Lemma A.1 it follows that

$$\operatorname{avar}\left\{n^{-1/2}\sum_{t=1}^n \left(x_{t-j}-\bar x_j\right)\left(\varepsilon_{t+1}+\eta_{t+1}-\pi\eta_t\right)\right\} = \sigma_\varepsilon^4\left\{\tfrac{1}{1-\pi^2} + 2\lambda + (1+\pi^2)\lambda^2\right\},$$

and that $n^{-1}\sum_{t=1}^n (x_{t-j}-\bar x_j)x_t \xrightarrow{p} \pi^j\sigma_y^2 = \pi^j\frac{\sigma_\varepsilon^2}{1-\pi^2}$, so that $n^{1/2}(\hat\pi_{\mathrm{IV}_j} - \pi) \xrightarrow{d} N(0, \sigma^2_{\hat\pi_{\mathrm{IV}_j}})$, where

$$\begin{aligned}
\sigma^2_{\hat\pi_{\mathrm{IV}_j}} &= \sigma_\varepsilon^4\left\{\tfrac{1}{1-\pi^2} + 2\lambda + (1+\pi^2)\lambda^2\right\}\left(\frac{\pi^j}{1-\pi^2}\sigma_\varepsilon^2\right)^{-2} \\
&= \pi^{-2j}\left\{1-\pi^2+2\left(1-\pi^2\right)^2\lambda + \left(1-\pi^2\right)(1-\pi^4)\lambda^2\right\}.
\end{aligned}$$

The proof for the case where $\pi = 1$ is given by Hall (1989). ■

Proof of Theorem 3. The structure of the $J \times J$ matrix $M_{\pi,\lambda}$ follows from Lemma A.1, and V_π follows from the law of large numbers and the fact that $\operatorname{cov}(x_{t-j}, x_t) = \pi^j\sigma_y^2$ for $j \geq 1$.

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press

We seek a vector, α, that solves

$$\min_{\alpha \in \mathbb{R}^J} \frac{\alpha' M_{\pi,\lambda} \alpha}{(\alpha' V_\pi)^2}.$$

This problem is clearly invariant to rescaling of α, so we can reformulate the problem as $\min_\alpha \alpha' M_{\pi,\lambda}\alpha$, s.t. $\alpha' V_\pi = 1$. The first-order conditions are simply $2M_{\pi,\lambda}\alpha - \rho V_\pi = 0$, where ρ is the Lagrange multiplier, so that $\alpha^\star = (V_\pi' M_{\pi,\lambda}^{-1} V_\pi)^{-1} M_{\pi,\lambda}^{-1} V_\pi$, is the solution to the constrained problem. By the scale invariance of α, we have that $\alpha^* = cM_{\pi,\lambda}^{-1} V_\pi$ yields an optimal instrument for any $c \neq 0$. ■

Proof of Theorem 4. (i) When $\pi < 1$, the result follows from

$$\begin{aligned} n^{-1}\textstyle\sum_{t=1}^n (x_{t-j} - \bar{x}_j)x_t &= n^{-1}\textstyle\sum_{t=1}^n (y_{t-j} - \bar{y}_j + \bar{\eta})(y_t + \xi + \eta_t) \\ &= n^{-1}\textstyle\sum_{t=1}^n (y_{t-j} - \bar{y}_j)y_t + o_p(1) = \gamma(j) + o_p(1). \end{aligned}$$

(ii) For the unit root case, the result follows from $n^{-2}\sum_{t=1}^n (x_{t-j} - \bar{x}_j)x_t = \int_0^1 (W(u) - \bar{W})^2 \mathrm{d}u + o_p(1)$, and

$$\begin{aligned} n^{-2}\textstyle\sum_{t=1}^n (x_{t-j} - \bar{x}_j)x_{t+1} - n^{-2} \; & \textstyle\sum_{t=1}^n (x_{t-j} - \bar{x}_j)x_t = n^{-2} \\ & \textstyle\sum_{t=1}^n (x_{t-j} - \bar{x}_j)\Delta x_{t+1} = o_p(1), \end{aligned}$$

for any $j \geq 0$, and for $j \geq \max(p,q)$, $n^{-1}\sum_{t=1}^n (x_{t-j} - \bar{x}_j)x_{t+1} \xrightarrow{d} \int_0^1 (W(u) - \bar{W})\mathrm{d}W(u)$; see Hall (1989). ■

Proof of Lemma 3. For a positive matrix, A, with spectral radius r, we know from the Perron-Frobenius theorem that $A^k/r^k \to ab'$ as $k \to \infty$, where a and b are the (left and right) eigenvectors associated with the largest eigenvalue of A, which equals the spectral radius, r. Moreover, the elements of the vectors, a and b, are all strictly positive.

Define the vector of p consecutive autocovariances, $\gamma_j^* = (\gamma_j, \ldots, \gamma_{j-p+1})'$. Since π is the spectral radius of Φ, then $(\pi^{-1}\Phi)^k \gamma_j^*$ converges to a limit that is proportional to the eigenvector a as $k \to \infty$. By the Yule-Walker equation, we have

$$\gamma_{j+1} = \varphi_1 \gamma_j + \cdots + \varphi_p \gamma_{j-p+1},$$

which implies that $\gamma_{j+1}^* = \Phi \gamma_j^*$. Thus if we define the vector $v_j = \pi^{-j}\gamma_j^* \in \mathbb{R}^p$, then $v_{j+1} = (\pi^{-1}\Phi)v_j$. This shows that v_j, as $j \to \infty$, approaches the left eigenvector associated with π, which implies that $v_{j+1} - v_j \to 0$ as $j \to \infty$. By considering the first elements of the vectors v_{j+1} and v_j, (which are nonzero because a is strictly positive), it now follows that $\pi^{-j+1}\gamma_{j+1} - \pi^{-j}\gamma_j \to 0$, so that $\gamma_{j+1}/\gamma_j \to \pi$. ■

https://doi.org/10.1017/S0266466613000121 Published online by Cambridge University Press