2002s-88

Semi-Parametric Weak Instrument Regressions with an Application to the Risk-Return Trade-off

Benoit Perron

Série Scientifique
Scientific Series

Montréal
Novembre 2002

CIRANO
Centre interuniversitaire de recherche
en analyse des organisations

CIRANO

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du ministère de la Recherche, de la Science et de la Technologie, de même que des subventions et mandats obtenus par ses équipes de recherche.

CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de la Recherche, de la Science et de la Technologie, and grants and research mandates obtained by its research teams.

Les organisations-partenaires / The Partner Organizations

- École des Hautes Études Commerciales
- École Polytechnique de Montréal
- Université Concordia
- Université de Montréal
- Université du Québec à Montréal
- Université Laval
- Université McGill
- Ministère des Finances du Québec
- MRST
- Alcan inc.
- AXA Canada
- Banque du Canada
- Banque Laurentienne du Canada
- Banque Nationale du Canada
- Banque Royale du Canada
- Bell Canada
- Bombardier
- Bourse de Montréal
- Développement des ressources humaines Canada (DRHC)
- Fédération des caisses Desjardins du Québec
- Hydro-Québec
- Industrie Canada
- Pratt & Whitney Canada Inc.
- Raymond Chabot Grant Thornton
- Ville de Montréal

ISSN 1198-8177

Semi-Parametric Weak Instrument Regressions with an Application to the Risk-Return Trade-off*

Benoit Perron[†]

Résumé / *Abstract*

Des recherches récentes démontrent qu'une corrélation faible entre les instruments et les variables explicatives peut mener à de sérieux problèmes d'inférence dans les régressions avec variables instrumentales. Nous étendons l'analyse locale à zéro des modèles avec instruments faibles aux modèles avec des instruments et régresseurs estimés et avec de la dépendance dans les moments supérieurs. Ainsi, cet environnement devient applicable aux modèles linéaires avec des variables anticipatoires qui sont estimées de façon non paramétrique. Deux exemples de tels modèles sont la relation entre le risque et les rendements en finance et l'impact de l'incertitude de l'inflation sur l'activité économique réelle. Nos résultats démontrent que l'inférence basée sur les tests du multiplicateur de Lagrange (LM) est plus robuste à la présence d'instruments faibles que l'inférence basée sur les tests de Wald. En utilisant des intervalles de confiance construits selon les tests de LM, nous concluons qu'il n'y a pas de prime de risque significative dans les rendements de l'indice S&P 500, les rendements excédentaires entre les Bons du Trésor de 6 mois et de 3 mois et les rendements du taux de change spot entre le yen japonais et le dollar américain.

Recent work shows that a low correlation between the instruments and the included variables leads to serious inference problems. We extend the local-to-zero analysis of models with weak instruments to models with estimated instruments and regressors and with higher-order dependence between instruments and disturbances. This framework is applicable to linear models with expectation variables that are estimated non-parametrically such as the risk-return trade-off in finance and the impact of inflation uncertainty on real economic activity. Our simulation evidence suggests that Lagrange Multiplier (LM) confidence intervals have better coverage in these models. We apply these methods to excess returns on the S&P 500 index, yen-dollar spot returns, and excess holding yields between 6-month and 3-month Treasury bills.

Mots clés: Variables instrumentales, instruments faibles, analyse locale à zéro, tests du multiplicateur de Lagrange, tests de Wald, prime de risque, anticipations, modèles semi-paramétriques, noyau.

Keywords: Instrumental Variables, Weak Instruments, Local-to-Zero Analysis, LM Tests, Wald Tests, Risk Premium, Expectations, Semi-Parametric Models, Kernel.

---

* Thanks go to two referees, the Editor, Peter Phillips, Oliver Linton, Donald Andrews, Hyungsik Moon, John Galbraith and seminar participants at Yale, Toronto, Montréal, Concordia, Rutgers, Windsor, Florida International University, and the 2000 World Congress of the Econometric Society for helpful comments and suggestions. The usual disclaimer necessarily applies. Financial assistance from the Alfred P. Sloan Foundation, the Social Science Research Council of Canada, the Fonds pour la Formation des chercheurs et l'aide à la recherche (FCAR), and the Mathematics of Information Technology and Complex Systems (MITACS) network is gratefully acknowledged.

[†] Département de sciences économiques, CIREQ, and CIRANO, Université de Montréal, C.P. 6 28, Succursale Centre-ville, Montréal (Québec), H3C 3J7. E-mail: benoit.perron@umontreal.ca.

1. Introduction

Recently, the problem of weak correlation between instruments and regressors in instrumental variable (IV) regressions has become a focal point of much research. Staiger and Stock (1997) developed an asymptotic theory for this type of problem using a local-to-zero framework. They show that standard asymptotics for IV estimators can be highly misleading when this correlation is low. Following this methodology, Zivot, Startz, and Nelson (1998), Wang and Zivot (1998), and Startz, Nelson, and Zivot (2001) show that usual testing procedures are unreliable in such situations, while Chao and Swanson (2000) provide expressions for the bias and MSE of the IV estimator based on higher-order asymptotic approximations. Extensions of this approach to nonlinear models have been developed in Stock and Wright (2000) . Earlier analyses of models under partial identification conditions are given in Phillips (1989) and Choi and Phillips (1992), and Dufour(1997).

This paper extends the weak instrument literature using the Staiger and Stock framework in two ways: first, we analyze a restricted class of semi-parametric models in which both regressors and instruments are estimated, and second, we allow for higher-order dependence between the instruments and the disturbances. These extensions are meant to make the analysis applicable to the many theoretical models in finance and macroeconomics that suggest a linear relationship between a random variable and an expectation term of the general form,

$$y_t = \gamma' x_t + \delta' Z_t + e_t \tag{1.1}$$

where y_t is a scalar, x_t is a vector of exogenous and predetermined variables, and Z_t is a vector of unobservable expectation variables.

The estimation of these models has proven difficult because a proxy has to be constructed for the unobservable expectation term. A complete parametric approach would assume functional forms for the expectation processes of agents which can then be estimated along with (1.1) by, for example, maximum likelihood. A semi-parametric approach, which is of interest in this paper, leaves the functional form of the expectation terms unspecified but uses the linear structure in (1.1) to estimate the parameters of interest once estimates of the expectation terms are obtained.

Of particular interest is the case where Z_t is a conditional variance term, and in this framework, interest centers on the parameter δ as it measures the response of y_t to increased risk. One such example includes the risk-return trade-off in finance where agents have to be compensated with higher expected returns for holding riskier assets. This trade-off has been examined by several authors, including French, Schwert, and Stambaugh (1987), Glosten, Jagannathan, and Runkle (1993), and Braun. Nelson, and Sunier (1995) , and a good survey can be found in Lettau and Ludvigson (2001). In this case, Z_t is the conditional variance of the asset, and x_t would generally include variables measuring the fundamental value of the asset. A second example is the analysis of the effect of inflation uncertainty on real economic activity where Z_t is the variance of the inflation rate conditional on past information, and y_t is some real aggregate variable such as real GDP or industrial production.

In the case where Z_t is a variance term, Engle, Lilien, and Robins (1987) have introduced the parametric AutoRegressive Conditional Heteroskedasticity-in-Mean (ARCH-M) model which postulates that $Z_t = \sigma_t^2$, the variance of returns, follows an ARCH(p) model. A popular generalization is the Generalized ARCH-M (GARCH-M) model with σ_t^2 of the form:

$$\sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \ldots + \alpha_p e_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \ldots + \beta_q \sigma_{t-q}^2 \tag{1.2}$$

with (1.1) and (1.2) estimated jointly by maximum likelihood. Two problems surface when using such models. First, global maximization of the likelihood function can be difficult unless p and q are kept small. Second, estimates in the mean equation will be inconsistent if the variance equation is misspecified because the information matrix is not block diagonal. Given the lack of restrictions on the behavior of the conditional variance provided by economic theory, this seems quite problematic.

An alternative approach that is robust to specification was suggested by Pagan and Ullah (1988) and Pagan and Hong (1991). Their suggestion is to first replace Z_t by its realized values, say Y_t, estimating this quantity non-parametrically, and using a non-parametric estimate of Z_t as an instrument. This approach is itself problematic since it does not solve the necessity to keep the number of conditioning variables low due to the curse of dimensionality. Moreover, a common problem when using such a semi-parametric approach is that the estimated conditional variance is poorly correlated with $\widehat{Y}_t$, the estimated realized values. This paper will focus on addressing this second problem. The first problem is addressed by using a semi-parametric estimator suggested by Engle and Ng (1993).

It will turn out that weak instrument asymptotics are useful in improving the quality of inference in this class of models. In particular, the use of confidence intervals based on the Lagrange multiplier principle provide much better coverage than more standard Wald confidence intervals.

The rest of the paper is divided as follows: section 2 presents the instrumental variable procedure described above in detail under the standard assumptions. In section 3, we present evidence on the presence of weak instruments in the risk-return trade-off. Next, in section 4, we develop asymptotic theory for the instrumental variable estimator described above under the weak instrument assumption. In section 5, results from a simulation experiment are presented to outline the difficulties involved in carrying out analysis in this type of models. Section 6 contains the results from applying the techniques developed in previous sections to three financial data sets: excess returns on the Standard and Poor's 500 index, yen-dollar spot returns, and excess holding yields on Treasury bills. Finally, section 7 provides some concluding comments.

2. Semi-parametric models with conditional expectations

As discussed above, we consider linear models such as,

$$y_t = \gamma' x_t + \delta' Z_t + e_t \tag{2.1}$$

where y_t is a scalar, x_t is a $k_1 \times 1$ vector of exogenous and predetermined variables, and Z_t is a $k_2 \times 1$ vector of unobservable expectation variables. One example of particular interest is where Z_t is a vector of variances and covariances of a vector ψ_t of the form $vech\left(E\left[Y_t|\mathcal{F}_t\right]\right)$, with $Y_t = \left(\psi_t - E\left[\psi_t|\mathcal{F}_t\right]\right)\left(\psi_t - E\left[\psi_t|\mathcal{F}_t\right]\right)'$ and where $\mathcal{F}_t$ is the information set available to agents in the economy at the beginning of period t. In this framework, interest centers on the parameter δ as it measures the response of y_t to an increase in the measure of risk. Such models were first investigated along the lines followed here by Pagan and Ullah (1988). In addition to that paper, the proposed IV estimator has been applied in Pagan and Hong (1991), Bottazzi and Corradi (1991), and Sentana and Wadhwani (1991). Except for Pagan and Ullah, all these papers analyze the trade-off between financial returns and risk as postulated by mean-variance analysis. Pagan and Ullah look at the forward premium in the foreign exchange market and the real effects of inflation uncertainty.

The first step in tackling this problem is to replace the conditional expectation Z_t by its realized value Y_t. In the following, we assume that Y_t is not observable as is the case in the variance example since Y_t is itself a function of an expectation. Thus, an extra step is required in replacing Y_t by an estimate, $\widehat{Y}_t$. The model to be estimated is then:

$$\begin{aligned} y_t &= \gamma' x_t + \delta' \widehat{Y}_t + e_t + \delta' \left(Y_t - \widehat{Y}_t \right) + \delta' \left(Z_t - Y_t \right) \\ &= \gamma' x_t + \delta' \widehat{Y}_t + u_t \end{aligned}$$

In general, an ordinary least squares regression of y_t on x_t and $\widehat{Y}_t$ will lead to inconsistent estimates of γ and δ due to the correlation between $\widehat{Y}_t$ and $(Z_t - Y_t)$. The solution suggested by Pagan (1984) and by Pagan and Ullah (1988) is to use an instrumental variable estimator with $\widehat{Z}_t$ used as instruments for $\widehat{Y}_t$. In fact, to obtain consistent estimates, any variable in $\mathcal{F}_t$ could be used as instrument. We could consider finding an optimal instrument as $E\left[\widehat{Y}_t | \mathcal{F}_t \right]$ which in general will be different from $\widehat{Z}_t$ because of the bias arising from the estimation of Y_t. The steps used to construct the estimator are illustrated as follows:

$$Z_t \xrightarrow[\text{replace with}]{} Y_t \xrightarrow[\text{estimate with}]{} \widehat{Y}_t \xrightarrow[\text{instrument with}]{} \widehat{Z}_t$$

This problem will be semi-parametric when Y_t and Z_t are estimated non-parametrically. As in many semi-parametric models, despite the lower rate of convergence of the non-parametric estimators, the estimates of γ and δ will converge at the usual $\sqrt{n}$ rate under certain conditions where n is the sample size.

Define $\overline{Z}_t = (x_t, Z_t)$, $\overline{Y}_t = (x_t, Y_t)$, $\overline{Z} = \left(\overline{Z}_1, \ldots, \overline{Z}_n\right)'$, $\overline{Y} = \left(\overline{Y}_1, \ldots, \overline{Y}_n\right)'$ with $\widetilde{Z}$ and $\widetilde{Y}$ similarly defined but with $\widehat{Z}_t$ and $\widehat{Y}_t$ replacing Z_t and Y_t. Further let $\overline{u}_t = e_t + \delta' (Z_t - Y_t)$ and $\theta = (\gamma, \delta)'$. Consider the IV estimator for this model:

$$\widehat{\theta} = \left(\widetilde{Z}'\widetilde{Y}\right)^{-1} \widetilde{Z}'y$$

Andrews (1994) proved the asymptotic normality of this estimator. There are two conditions of interest here: the first one is that $\widehat{Y}$ be $\sqrt{n}$-consistent. This ensures that the asymptotic distribution of the IV estimator of $\widehat{\theta}$ is not affected by replacing Y_t and Z_t by $\widehat{Y}_t$ and $\widehat{Z}_t$ respectively. This will generally not be the case when $\widehat{Y}$ is estimated non-parametrically. However, it will hold in the special case where Z_t is a variance term as long as the mean of $E\left[\psi_t | \mathcal{F}_t\right]$ is estimated at rate $n^{1/4}$. Conditions under which this holds can be found in Andrews (1995).

The second key assumption is that the matrix $n^{-1}\widetilde{Z}'\widetilde{Y}$ converge to a nonsingular limit. It is a key assumption because the quality of the instrument $\widehat{Z}_t$ will determine the quality of the asymptotic approximation obtained by Andrews (1994). This assumption is nearly violated in many practical situations, and this is the motivation for the development of the weak instrument literature. The next section will document this phenomenon for financial data.

3. Evidence of weak instruments

In the case of interest in which $Y_t = e_t^2$ and $Z_t = \sigma_t^2$, it will generally be the case that the correlation between the two estimates, $\widehat{e}_t^2$ and $\widehat{\sigma}_t^2$, is very low, suggesting a weak instrument problem. Table 1 shows the value of R^2 for the regression of $\widehat{e}_t^2$ on a constant and $\widehat{\sigma}_t^2$ for three financial data sets using two different non-parametric estimators.

The first data set analyzed represents monthly excess returns on the Standard and Poor's 500 between January 1965 and December 1997 measured at the end of each month. The data is taken from CRSP, and the risk-free rate is the return on three-month Treasury bills. The second data set is made of monthly returns on the yen-dollar spot rate obtained from International Financial Statistics between September 1978 and June 1998. Finally, the last data series consists of quarterly excess holding yields on 6-month versus 3-month Treasury bills between 1959:1 and 2000:2. A similar, but shorter, data set has already been analyzed by Engle, Lilien, and Robins (1987) using their ARCH-M methodology and Pagan and Hong (1991) using the above instrumental variable estimator. The three data sets are plotted in figure 1.

**** Insert figure 1 here ****

The first nonparametric estimator is based on multivariate leave-one out kernel. First, we estimate the mean of y_t and y_t^2,denoted τ_{1t} and τ_{2t} respectively, as:

$$\hat{\tau}_{jt} = \frac{\sum_{i \neq t} y_i^j K\left(\frac{w_i - w_t}{b_j}\right)}{\sum_{i \neq t} K\left(\frac{w_i - w_t}{b_j}\right)}$$

for $j = 1, 2$ with the kernel function $K(w)$ taken to be the multivariate standard normal. The bandwidth b_j and the number of lags of y_t in the conditioning set p_j are selected using a modified version of the criterion suggested in Tjostheim and Auestad (1994) that penalizes small bandwidths and large lag lengths. Accordingly, we choose the bandwidth (b_j) and lag length (p_j) so as to minimize

$$\ln\left[\frac{1}{n}\sum_{t=1}\left(y_t^j - \hat{\tau}_{jt}\right)^2\right] + \frac{\ln n}{n}\left(\frac{K(0)}{b_j}\right)^{p_j} \frac{\sum_{t=1}^{n} \frac{\left(y_t^j - \hat{\tau}_{jt}\right)^2}{f(w_t)}}{\sum_{t=1}^{n}\left(y_t^j - \hat{\tau}_{jt}\right)^2}$$

where $K(0)$ is the kernel evaluated at 0 and $f(w_t)$ is the density of the conditioning variables. The bandwidth takes the form:

$$b_j = c_j s n^{-\frac{1}{4+p_j}}$$

where s is the standard deviation of y_t and c_j is a constant to be selected. We then define $\widehat{e}_t^2 = (y_t - \widehat{m}_{1t})^2$ and obtain an estimate of σ_t^2 as:

$$\widehat{\sigma}_t^2 = \widehat{m}_{2t} - (\widehat{m}_{1t})^2 .$$

A theoretical analysis of this non-parametric estimator of the conditional variance can be found in Masry and Tjostheim (1995). In order to avoid unbelievably small bandwidth choices for all three series, we left out outliers in the bandwidth selection process. The extreme 25% of the data was not used in the computation of the information criteria.

The second estimator was first proposed by Engle and Ng (1993). It provides more structure to the conditional variance and will approximate the conditional variance function much better than the kernel when the variance is persistent (see Perron (1999) for simulation evidence). The estimator is implemented by first estimating the mean by a kernel estimate as above and then fitting an additive function for σ_t^2 as follows:

$$\sigma_t^2 = \omega + f_1\left(\widehat{e}_{t-1}\right) + \ldots + f_p\left(\widehat{e}_{t-p}\right) + \beta\sigma_{t-1}^2$$

where the $f_j(\cdot)$ are estimated as splines with knots using a Gaussian likelihood function. This allows for a flexible effect of recent information on the conditional variance while allowing for persistence. This framework includes most parametric models suggested in the literature such as the GARCH class. The number of segments in the spline functions acts as a smoothing parameter and is selected using BIC. The knots in the spline were selected using the order statistics such that each bin has roughly the same number of observation subject to the constraint of an equal number of bins in the positive and negative regions.

**** Insert table 1 here ****

A quick look at table 1 reveals that only the excess holding yield data has R^2 greater than 5.5%. The reason for this low correlation is that e_t^2 and σ_t^2 have very different volatility. Even if $E\left[e_t^2|\mathcal{F}_t\right] = \sigma_t^2$, financial returns are extremely volatile and therefore, the difference between e_t^2 and σ_t^2 can be quite large. This is true even if we did not have to estimate these two quantities; having to estimate them complicates matters further. We can illustrate by looking at the GARCH(1,1) model:

$$\begin{aligned} y_t &= \mu + \sigma_t \varepsilon_t = \mu + e_t \\ \sigma_t^2 &= \omega + \alpha e_{t-1}^2 + \beta \sigma_{t-1}^2. \end{aligned}$$

Andersen and Bollerslev (1998) show that the population R^2 in the regression

$$(y_t - \mu)^2 = a_0 + a_1 \hat{\sigma}_t^2 + v_t$$

where $\hat{\sigma}_t^2$ is the one-period ahead forecast obtained from the GARCH model is

$$R^2 = \frac{\alpha^2}{1 - \beta^2 - 2\alpha\beta}$$

which will in general be very small even though $E\left[(y_t - \mu)^2 |\mathcal{F}_t\right] = \sigma_t^2$. Figure 2 plots the value of R^2 for different values of α and β for thsi GARCH(1,1) example. The value of R^2 is highly sensitive to the value of α, and this reflects that $\alpha = 0$ makes the model unidentified. It is usual in the literature to find point estimates of GARCH(1,1) models in the neighborhood of $\alpha = 0.05$ and $\beta = 0.9$. The figure clearly shows that for such values, the correlation between e_t^2 and σ_t^2 will typically be quite low. The problem in this case is that σ_t^2 has very low variance relative to that of y_t^2; a low value of α means that σ_t^2 is nearly constant locally.

**** Insert figure 2 here ****

We can expect that table 1 does not even provide an accurate picture of the problem of weak instruments. Using data sampled at higher frequency (e.g. daily or even intra-day) would result in even lower correlation. The lower frequency allows some averaging which reduces the variance of e_t^2. Potentially a better solution is to use "model-free" measures of volatility such as those proposed by Andersen, Bollerslev, Diebold, and Labys (2001) which are obtained by summing squared returns from higher frequency data. We do not pursue this possibility here, but note that its variance-reducing property could be helpful in this context.

4. Asymptotics with weak instruments

Staiger and Stock (1997) have recently shown, in the framework of a linear simultaneous equation system, that having instruments that are weakly correlated with the explanatory variables makes the usual asymptotic theory work poorly. Their assumed model is:

$$y = Y\delta + X\gamma + u \tag{4.1}$$

$$Y = Z\Pi + X\Gamma + V \tag{4.2}$$

where Y is the matrix of included endogenous variables that are to be replaced by at least k_2 instruments. Since in our case, it will always be true that the model is exactly identified (that is, there will be as many regressors as instruments since the instruments are estimates of the expected value of the regressors), we will concentrate on the case where Z is a $n \times k_2$ matrix. The weak instrument assumption is imposed by assuming that:

$$\Pi = \frac{G}{\sqrt{n}} \tag{4.3}$$

for some fixed $k_2 \times k_2$ matrix $G \neq 0$ This assumption implies that in the limit, Y and Z are uncorrelated.

We extend the analysis of weak instruments in Staiger and Stock (1997) to our case of interest by allowing Y and Z to be unobserved and estimated by $\hat{Y}$ and $\hat{Z}$ respectively. Moreover, we allow for the possibility of higher-order dependence between the instruments and the disturbances. Simple algebra leads to:

$$\begin{aligned} \hat{Y} &= \hat{Z}\Pi + \left(Z - \hat{Z}\right)\Pi + \left(\hat{Y} - Y\right) + X\Gamma + V \\ &= \hat{Z}\Pi + X\Gamma + \zeta \end{aligned}$$

so that the correlation between $\hat{Y}_t$ and $\hat{Z}_t$ is also low.

There might be two reasons for a low correlation between the estimated instrument and explanatory variable in a given data sample. The first may be that the estimators used in constructing $\hat{Z}$ and $\hat{Y}$ are poor and will not approach their true value in small samples. On the other hand, the estimators may not be poor in any sense, but Y and its expected value may be weakly correlated such as in the GARCH(1,1) example above.

Recall that the IV estimator of δ is:

$$\begin{aligned}\hat{\delta} &= \left(\hat{Z}' M_x \hat{Y}\right)^{-1} \hat{Z} M_X y \\ &= \delta + \left(\hat{Z}' M_X \hat{Y}\right)^{-1} \hat{Z}' M_X u\end{aligned}$$

where $M_X = I - X\left(X'X\right)^{-1} X'$. In order to derive the distribution of $\hat{\delta}$, we need to make an extra assumption on the reduced-form coefficients of X. We will also assume that they are local to zero:

$$\Gamma = \frac{H}{\sqrt{n}} \tag{4.4}$$

for some $k_1 \times k_2$ matrix $H \neq 0$. This assumption is made because if Γ were fixed, X and Y would be collinear in the limit and the moment matrices would be singular, and it plays no role in the analysis of the behavior of $\hat{\delta}$.

The distribution of the estimators is given in the following theorem. All proofs are relegated to the appendix.

Theorem 4.1. In the model (4.1) – (4.3), assume the following:

1. $\sqrt{n}\left(\hat{Y} - Y\right) \xrightarrow{p} 0$;
2. $\hat{Z} = Z + o_p(1), \quad Z < \infty$ a.s.;
3. θ_0 is the interior of $\Theta \subset R^{k+1}$;
4. $(n^{-1}X'X, n^{-1}X'Z, n^{-1}Z'M_X Z) \xrightarrow{p} (\Sigma_{XX}, \Sigma_{XZ}, \Sigma_{ZZ})$;
5. $\left(n^{-\frac{1}{2}}X'u, n^{-\frac{1}{2}}Z'M_X u, n^{-\frac{1}{2}}X'V, n^{-\frac{1}{2}}Z'M_X V\right) \Rightarrow (\Psi_{Xu}, \Psi_{Zu}, \Psi_{XV}, \Psi_{ZV})$.

Define:

$$\sigma_{Zu} = \lim_{n\to\infty} \frac{1}{n} \sum_s \sum_t Z_s^{\perp} u_t' u_s Z_t^{\perp\prime}$$

$$\sigma_{ZV} = \lim_{n\to\infty} \frac{1}{n} \sum_s \sum_t Z_s^{\perp} V_s' V_t Z_t^{\perp\prime}$$

$$\sigma_{Xu} = \lim_{n\to\infty} \frac{1}{n} \sum_s \sum_t X_s u_s' u_t X_t'$$

$$\sigma_{XV} = \lim_{n\to\infty} \frac{1}{n} \sum_s \sum_t X_s V_s' V_t X_t'$$

$$\rho_Z = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} Z_t^{\perp} V_t' \sigma_{ZV}^{-\frac{1}{2}\prime} \sigma_{Zu}^{-\frac{1}{2}} u_s Z_s^{\perp\prime}$$

$$\rho_X = \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{s=1}^{n} X_t V_t' \sigma_{Zv}^{-\frac{1}{2}\prime} \sigma_{Zu}^{-\frac{1}{2}} u_s X_s$$

where $Z_t^{\perp}$ is the residuals from the projection of Z_t onto X, *i.e.* it is the transpose of the t^{th} row of $Z^{\perp} = M_X Z$.

Then,

1. $\hat{\delta} - \delta \xrightarrow{d} \Xi = \sigma_{ZV}^{-\frac{1}{2}} (\lambda + z_v)^{-1} \sigma_{Zu}^{\frac{1}{2}} z_u$ with $\lambda = \sigma_{ZV}^{-\frac{1}{2}} \sum_{ZZ} G$, where $z_u = z_v \rho_Z + (1 - \rho_Z \rho_Z')^{\frac{1}{2}} \xi$, and $(vec(z_v), \xi) \sim N\left(0, I_{k_2(k_2+1)}\right)$,
2. In addition, with (4.4), $\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} \sum_{XX}^{-1} \left[\sigma_{Xu}^{\frac{1}{2}} x_u + \left(\sum_{XZ} G + \sum_{XX} H + \sigma_{XV}^{\frac{1}{2}} x_v \right) \Xi \right]$, where $x_u = x_v \rho_X + \left(I_{k_1} - \rho_X \rho_X' \right)^{\frac{1}{2}} \zeta$, and $(vec(x_v), \zeta) \sim N\left(0, I_{k_1(k_2+1)}\right)$.

Assumptions 1-3 of the theorem are the same as used by Andrews (1994) to derive the asymptotic distribution of $\hat{\theta}$, while assumptions 4 and 5 are similar to those of Staiger and Stock (1997). Several aspects of this result can be pointed out, all the outcome of the poor identification of δ. First, the IV estimator of δ does not converge to the true population value, but rather to a random variable as in Phillips (1989). Second, the limit distribution is the ratio of correlated normal random variables. This suggests that the distribution will, in some cases, have thick tails and be bimodal. Moreover, the distribution depends on nuisance parameters λ, and ρ_Z, making inference difficult. If $\lambda \to \infty$ at a rate of $\sqrt{n}$, Ξ will approach the usual normal distribution.

In addition, the distribution of the coefficients on the exogenous variables x_t is contaminated by the poor identification of δ. Specifically, we expect that the usual standard errors will understate the true uncertainty as these are based on the first term of the limiting distribution only. This will lead to over-rejection of hypotheses of the type $H_0 : \gamma = \gamma_0$.

The basic distribution theory described above is very closely related to that derived by Staiger and Stock. The form of the covariance matrix is different because we do not assume that the instruments, Z_t, are independent of the error terms u_t and v_t; we only assume that they are uncorrelated. This adjustment allows for higher-order dependence between Z_t on the one hand and u_t and v_t on the other. In cases where there is no higher dependence between the instruments and the error terms, this distribution coincides with the one derived by Staiger and Stock.

The assumptions on the properties of the data are given in terms of high-level conditions, a joint weak law of large numbers and a weak convergence result. This is done to make the conditions similar to those used by Staiger and Stock. Many sets of primitive conditions can lead to these two results. For example, sufficient conditions are that the vector (u_t, V_t) be a martingale difference sequence with respect to the filtration $\left\{(u_{t-j-1}, V_{t-j-1}, Z_{t-j}, X_{t-j}), j \geq 0\right\}$ with uniform finite $(2+\eta)$ moments for some $\eta > 0$ and the vector (Z_t, X_t) be α-mixing with mixing numbers of size $-\kappa/(\kappa-1)$ and $(r+\kappa)$ finite moments for some $r \geq 2$. These conditions imply that in the variance case, $Z_t = \sigma_t^2$, we need σ_t^8 to be finite for all t. This is a difficult requirement for financial data as there is some evidence that many financial series do not even have four finite moments. For this reason, we will use highly aggregated data (for example monthly and quarterly data) for applications. However, our simulation results will show that reliable inference can still be done even under moment condition failure.

Use of the asymptotic theory developed above is hampered by the presence of the nuisance parameters, λ and ρ, which cannot be consistently estimated. Suppose we want to test the null hypothesis $H_0 : R\delta = r$ using the usual Wald statistic:

$$W = \left(R\hat{\delta} - r\right)' \left[R\left(\hat{Z}' M_X \hat{Z}\right)^{-1} var\left(\hat{Z}' M_X \hat{u}\right)\left(\hat{Z}' M_X \hat{Z}\right)^{-1} R'\right]^{-1} \left(R\hat{\delta} - r\right).$$

The following proposition gives the asymptotic theory of Wald statistics in the above model:

Proposition 4.2. Under the null hypothesis $H_0 : R\delta = r$,

$$W \xrightarrow{d} \Xi' R' \left[R\sigma_{ZV}^{-\frac{1}{2}} (\lambda + z_v)^{-1} \sigma_{Zu}^{*} (\lambda + z_v)^{-1} \sigma_{ZV}^{-\frac{1}{2}} R'\right]^{-1} R\Xi$$

where

$$\begin{aligned}
\sigma^*_{Zu} &= \sigma_{Zu} - \sigma^*_{uv} - \sigma^{*\prime}_{uv} + \sigma^*_{ZV} \\
\sigma^*_{uv} &= \lim_{n\to\infty} \frac{1}{n} \sum_s \sum_t Z_t^{\perp} u_t' \Xi' V_s Z_s^{\perp\prime} \\
\sigma^*_{ZV} &= \lim_{n\to\infty} \frac{1}{n} \sum_s \sum_t Z_t^{\perp} V_t' \Xi \Xi' V_s Z_s^{\perp\prime}.
\end{aligned}$$

As Wang and Zivot (1998) have noticed, in the case of just-identified models as is the case here, if we use the restricted estimate of σ_{Zu}, test statistics will have a limiting χ^2 distribution. In over-identified models, these test statistics will be bounded from above by a $\chi^2(K)$ distribution where $K > k_2$ is the number of instruments. Thus LM statistics will be appropriate if our concern is to control test size and construct asymptotically valid confidence intervals. These LM confidence intervals can be obtained as the set of δ such that the LM test statistic does not reject the null hypothesis which involves, in the case with no higher-order dependence, solving a quadratic equation as shown by Zivot, Startz, and Nelson (1998). The resulting confidence intervals could thus be a bounded set, the union of two unbounded intervals, or the entire real line. The possibility that confidence intervals be unbounded reflects the great uncertainty about the parameter of interest. Dufour (1997) has shown that a valid $(1-\alpha)$ confidence interval for a locally unidentified parameter will be unbounded with probability $(1-\alpha)$. Since Wald intervals are always bounded (being constructed by adding and subtracting two standard errors to the point estimate), they cannot have infinite volume and hence cannot provide valid inference in this type of model in the sense that they must have zero coverage asymptotically. Unfortunately, these Wald intervals are almost always used in practice.

In our case here, we need to adjust the LM statistic for the higher order dependence. This is done in the following proposition for our just-identified case:

Proposition 4.3. Let $g = n^{-1}\hat{Z}' M_X \left(y - \hat{Y}\delta\right)$. Then under the null hypothesis, $H_0 : \delta = \delta_0$, $LM = n g' \hat{\sigma}^{-1}_{Zu,0} g \xrightarrow{d} \chi^2(k_2)$ where $\hat{\sigma}_{Zu,0} = n^{-1} \sum_s \sum_t Z_s^{\perp} u_{t,0}' u_{s,0} Z_t^{\perp\prime}$ is an estimator of σ_{Zu} computed under the null hypothesis.

Unfortunately, in this case, there is no easy way to write the inequality that defines the confidence intervals as a quadratic equation in δ. Confidence intervals must be computed numerically by defining a grid of δ and verifying for each point on the grid whether the LM statistic defined in the above proposition is less than the appropriate critical value from the $\chi^2(k_2)$ distribution. This method is easily implemented in the scalar case, but could hardly be carried out in high dimensions.

Another approach to obtaining confidence intervals, suggested by Staiger and Stock (1997), is to use the Anderson-Rubin (1949) statistic. It is usually defined as the F-statistic for the significance of δ^* in the regression

$$y - \widehat{Y}\delta_0 = X\gamma^* + \widehat{Z}\delta^* + u^*$$

where $\gamma^* = \gamma + \Gamma(\delta - \delta_0)$, $\delta^* = \Pi(\delta - \delta_0)$, and $u^* = u + v(\delta - \delta_0)$. Since we have a case with heteroskedasticity, we need to use robust standard errors to compute the test statistic. It turns out that in the just-identified case, this statistic is identical to the above LM statistic. This fact is stated in the following proposition:

Proposition 4.4. Let $AR = n\widehat{\delta}^{*\prime}\widehat{V}^{-1}\widehat{\delta}^{*}$ where $\widehat{V} = \left(\frac{1}{n}\widehat{Z}'M_X\widehat{Z}\right)^{-1}\widehat{\sigma}_{Zu,0}\left(\frac{1}{n}\widehat{Z}'M_X\widehat{Z}\right)^{-1}$. Then, under the null hypothesis $H_0 : \delta = \delta_0$, $AR = LM$.

The above propositions thus give us two equivalent ways to construct asymptotically valid confidence intervals for the entire vector δ. The two methods are exactly the same as long as the same estimate of σ_{Zu} is used to construct either LM or AR. The performance of these intervals in a small sample situation will be analyzed in the simulation experiment in the next section. In the case where a confidence interval on a linear combination of a subvector of δ is desired, one can proceed by the projection method discussed in Dufour and Jasiak (2001) and further analyzed in Dufour and Taamouti (2000). Such an approach would be valid but conservative.

In a related paper, Dufour and Jasiak (2001) have obtained exact tests based on AR-type statistics in models with generated regressors and weak instruments. However, their results only apply to parametrically-estimated regressors that will converge at rate $\sqrt{n}$ and not to the non-parametric estimators analyzed here.

Startz, Nelson, and Zivot (2001) have developed an alternative set of statistics, which they call S statistics, that take into account the degree of identification. They show in the case of a single regressor and instrument ($k_1 = k_2 = 1$) that these are equivalent to the AR statistic. We suspect that this correspondence is more general and carries over to the exactly identified case that we treat here, but we have no proof for this conjecture.

5. Simulation Results

In this section, the behavior of the procedures described above will be analyzed through a small simulation experiment. Important issues to be analyzed include the choice of smoothing parameters, the appropriateness of the various confidence intervals, and the distribution of the resulting estimators.

Consider the GARCH-M$(1, 1)$ DGP:

$$\begin{aligned} y_t &= \gamma + \delta\sigma_t^2 + e_t = \gamma + \delta\sigma_t^2 + \sigma_t\varepsilon_t \\ \sigma_t^2 &= \omega + \alpha e_{t-1}^2 + \beta\sigma_{t-1}^2 \\ \varepsilon_t &\sim i.i.d.\,(0, 1) \end{aligned}$$

In terms of the above notation, we have $v_t = e_t^2 - \sigma_t^2,\ u_t = e_t - \delta v_t,\ \ Y_t = e_t^2$, and $Z_t = \sigma_t^2$. The distribution of ε_t is either normal or Student t. This allows us to check the robustness of the procedures to the restrictive moment assumptions required by the asymptotic theory developed above. We use six sets of parameters, all estimated from data, which are presented in table 2.

**** Insert table 2 here ****

The point estimates for the stock data are similar to those usually obtained in this context, for example by Glosten, Jagannathan, and Runkle (1993), and will lead to a rather persistent σ_t^2 and to a weak instrument. Sample sizes of 450, 300, and 150 are used for the experiments, with the first 50 observations deleted to remove the effect of the initial condition (taken as the mean of the unconditional distribution). The length of the samples nearly match those of the S&P, exchange rate, and excess holding yield data.

One disadvantage of the current setup is that the correlation between $\hat{\sigma}_t^2$ and $\hat{\varepsilon}_t^2$ cannot be controlled. We can control the correlation between the unobservable variables, but due to estimation, the correlation between observable variables will be different in general.

The values of the nuisance parameters in this setup can be obtained in terms of the moments of the conditional variance process as:

$$\begin{aligned}
\sigma_{ZV} &= (\kappa_4 - 1)\left[E\left(\sigma_t^8\right) - 2E\left(\sigma_t^2\right)E\left(\sigma_t^6\right) + E\left(\sigma_t^2\right)^2 E\left(\sigma_t^4\right)\right] \\
\sigma_{Zu} &= \delta^2\sigma_{Zv} + E\left(\sigma_t^6\right) - 2E\left(\sigma_t^2\right)E\left(\sigma_t^4\right) + E\left(\sigma_t^2\right)^3 \\
&\quad -2\delta\kappa_3\left[E\left(\sigma_t^7\right) - 2E\left(\sigma_t^5\right)E\left(\sigma_t^2\right) + E\left(\sigma_t^3\right)E\left(\sigma_t^2\right)^2\right] \\
\rho_Z &= \frac{-\delta\sigma_{Zv} + \kappa_3\left[E\left(\sigma_t^7\right) - 2E\left(\sigma_t^5\right)E\left(\sigma_t^2\right) + E\left(\sigma_t^3\right)E\left(\sigma_t^2\right)^2\right]}{\sigma_{Zu}^{\frac{1}{2}}\sigma_{ZV}^{\frac{1}{2}}} \\
\lambda &= \frac{\sqrt{n}\left[E\left(\sigma_t^4\right) - E\left(\sigma_t^2\right)^2\right]}{\sigma_{Zv}^{\frac{1}{2}}} \\
\sigma_{Zu}^* &= \sigma_{Zu} - 2\sigma_{uv}\Xi + \sigma_{ZV}\Xi^2
\end{aligned}$$

where $\kappa_j = E\left(\varepsilon_t^j\right)$ is the j^{th} moment of ε_t. The values of the first 4 even moments of σ_t^2 are derived recursively in Bollerslev (1986) as a function of ω, α, and β and the moments of ε_t. This allows for the easy computation of the nuisance parameters which are included in table 2. Note that the moment condition assumed in theorem 4.1 is only satisfied for the first two sets of parameters.

Figure 3 shows a plot of the asymptotic distribution of the usual t statistic (from proposition 4.2) using the above estimates of the nuisance parameters and the standard normal distribution obtained under the usual asymptotic theory for the first two sets of parameters. The figure is drawn with 500,000 draws taken from each distribution. For the other experiments since the higher-order moments necessary to obtain the limit distribution do not exist, we cannot use the weak instrument limiting distribution to describe the behavior of the estimator.

The top panel in figure 3 represents the distribution for the S&P 500 data. For those values of the parameters, the t statistic has a highly skewed distribution. On the other hand, the bottom panel reveals that for the second experiment, the t statistic is both highly skewed and has fat tails. In fact, a good part of the probability mass (about 7 %) lies outside of the $[-4, 4]$ interval. The shape of the distribution is controlled by 2 nuisance parameters, λ and ρ_Z. These experiments show that low λ and high $|\rho_Z|$ give distributions very far from normality. To measure the impact of these properties on coverage probabilities, note that only 77.7% of the mass is between -1.96 and 1.96 in the bottom panel, while the same figure is 96.5% in the top panel. We conclude that the first experiment will have usual (Wald-based) 95% confidence intervals with coverage rates higher than their nominal level, while those in the second experiment will exhibit low coverage.

**** Insert figure 3 here ****

To demonstrate convergence to normality, figure 4 shows the same picture for $n = 50,000$ for both experiments. Since the weak instrument approximation approaches the standard normal as $n \to \infty$ in this case because $\lambda \to \infty$ at rate $\sqrt{n}$, we see that both skewness and excess kurtosis are much reduced. In this case, 95.2% and 95.1% of the mass lies between -1.96 and 1.96 respectively. Because of the parameter values, the distribution for the second experiment requires a much larger sample size than the first one in order to have a reasonably normal distribution and accurate 95% Wald-based confidence intervals.

**** Insert figure 4 here ****

The simulation results are presented in figures 5 and 6 and tables 3 and 4. Figure 5 provides a plot of the density of the weak IV approximation and of the infeasible IV estimator that uses the actual values of σ_t^2 and e_t^2 generated; this estimator is infeasible since these values are unobservable in practice. Figure 6 provides the sane information for the two nonparametric estimators. In tables 3 and 4, the first column shows the median of the IV estimator (rather than the mean because of the heavy tails of the distributions). The next two columns indicate the coverage rate of the appropriate 95% confidence intervals. The fifth column contains the mean R^2 of a regression of $\hat{e}_t^2$ on a constant and $\hat{\sigma}_t^2$. The next two columns provide the Kolmogorov-Smirnov (KS) statistic as a measure of fit of the small-sample distribution to the two alternative asymptotic approximations (if applicable). Finally, the last two columns compare the fits of the nonparametric estimates of both the regressor and instrument by reporting the R^2 from a regression of the true values on a constant and the nonparametric estimates. The first line of each panel reports results of the infeasible estimator discussed above.

We first discuss the results for the infeasible estimator. All experiments with the infeasible estimator were repeated 10,000 times. The asymptotic approximation captures the finite-sample distribution of the t statistic well. It matches the skewness and kurtosis well and thus provides a much better description than the normal approximation.

**** Insert figure 5 here ****

In all six experiments, the infeasible IV estimator is biased upward. The Wald confidence intervals have a coverage rate that is higher than its nominal level for the S&P data and lower (and sometimes much lower) for the other two data sets, while the LM interval has coverage rate that is only slightly too low in all cases. Not surprisingly, the weak instrument approximation is more accurate according to the KS statistic in both cases where it can be computed. The improvement is much more dramatic in the very non-normal case of experiment 2. Note also that the overall results are not sensitive to conditional normality or the existence of moments.

**** Insert tables 3 and 4 here ****

We now turn our attention to the semi-parametric estimators. Estimates of e_t^2 and σ_t^2 are obtained using the same two nonparametric methods as above, either a kernel or the semi-parametric Engle-Ng estimator using data-based selection for all smoothing parameters. Each experiment with the non-parametric estimators was repeated 5000 times.

The need to estimate σ_t^2 and e_t^2 changes the result quite dramatically relative to the infeasible estimator. The results using the kernel estimates are presented in the second row of each panel of tables 3 and 4 and as the dashed line in figure 6, while those for the Engle-Ng are presented in the third row of each panel and as the dotted line in figure 6. Overall, the Engle-Ng procedure leads to an IV estimator that much more closely matches the infeasible one. In particular, its distribution has a similar shape to that of the infeasible IV (and that of the weak IV approximation), and the coverage rate of the confidence intervals based on it are much closer to those of the infeasible estimator. The reason for this is clear: it provides a better approximation to the instrument (σ_t^2) than does the kernel as evidence by the higher R^2 in the regression of true conditional variance on a constant and its estimate which is consistent with the simulation evidence in Perron (1999) .The regressor (e_t^2) is well approximated by any method. Note also that once we estimate the regressor and instrument, the IV estimator of δ is strongly biased towards zero (with the exception of experiment 3).

**** Insert figure 6 here ****

An important practical result is that LM-based confidence intervals are more robust (in terms of having correct coverage) to both the presence of weak instruments and to the estimation of regressors and instruments. In all cases, the coverage rate of LM-based confidence intervals is closer to 95% than Wald–based intervals. If, in addition, the Engle-Ng estimator is used, coverage is almost exact. These should therefore be preferred in empirical work.

Table 5 provides details on the nonparametric estimators used in the simulation. We report the mean bandwidth constant, lag length selected, sum of the first 10 squared autocorrelation coefficients of the variance residuals, as well as the median constant and slope coefficient from the regression of the true instrument and regressor on a constant and the non-parametric estimates. The R^2 from these regressions has already been reported in tables 3 and 4.

**** Insert table 5 here ****

The BIC-type criterion seems to overpenalize the number of lags as it always chooses a single lag for all kernel estimates. However, it does suggest that some oversmoothing relative to the *i.i.d.* normal case is typically warranted (since in that case, the optimal bandwidth constant is 1.06). This is not surprising and is usually the case for dependent data. The criterion also seems to penalize heavily the number of bins in the Engle-Ng estimator as the mean number of bins is not much above 2. However, it frequently chooses more than one lag.

The main feature of table 5 however is the tight relation between the bias of the instrument estimates and the behavior of the resulting IV estimator relative to the infeasible estimator. In cases where the IV estimator with estimated regressor and instrument performs poorly (experiments 2, 5, and 6 for both estimators and experiment 4 for the kernel only), the median slope parameter from the instrument regression is always less than 0.5, suggesting a severe bias of the nonparametric estimator. This result is akin to the typical result in semiparametric estimation that it is preferable to undersmooth the nonparametric component so as to reduce bias. The averaging in the second step mitigates the higher variance that this undersmoothing typically entails, while it does not eliminate bias.

6. Empirical results

In this section, we analyze our three financial data sets to seek evidence of a risk-return trade-off. To reiterate, the series are monthly returns on the S&P 500 index, monthly returns on the yen-dollar spot rate, and quarterly excess holding yield between 6-month and 3-month Treasury bills. For each series, we postulate a model of the form

$$y_t = \gamma + \delta\sigma_t^2 + e_t$$

with $\sigma_t^2 = E\left[\{y_t - E[y_t|\mathcal{F}_{t-1}]\}^2 |\mathcal{F}_{t-1}\right]$ where $\mathcal{F}_{t-1}$ are lagged values of y_t. For all three series, the conditional variance was estimated using either the kernel or Engle-Ng estimator described above with the data-based selection of the tuning parameters. For comparison, we also report the results from a GARCH-M$(1,1)$ model estimated using Gaussian quasi-maximum likelihood.

The convergence to normality shown in the simulation might suggest that the use of higher frequency data is greatly desirable as it would increase sample size, but higher frequency would also lead to a more persistent conditional variance and hence a weaker instrument. The impact of this choice on the behavior of the IV estimator and its related statistics is therefore ambiguous. As discussed already, another potential use of high-frequency data (not pursued here) is to get better estimates of low-frequency volatility.

The estimation results are presented in table 6. In addition to the point estimates and their robust (White) standard errors, we present Wald-based and LM-based 95% confidence intervals for the coefficient on the risk variable, δ, the R^2 in a regression of $\widehat{\sigma}_t^2$ on $\widehat{e}_t^2$ and a constant, and the values of the tuning parameters used to construct the nonparametric estimates. The LM confidence intervals were computed by numerically inverting the LM statistic using a grid of 20,000 equi-spaced points between -1000 and 1000. For this reason, the infinite or very large confidence intervals are truncated at these two endpoints.

The trade-off between risk and return has been extensively studied for stocks with conflicting results. For example, French, Schwert, and Stambaugh (1987) find a positive relation between returns and the conditional variance, while Glosten, Jagannathan, and Runkle

(1993) find a negative relationship using a modified GARCH-M methodology. This conflicting evidence is not surprising in light of the results obtained by Backus, Gregory, and Zin (1989) and Backus and Gregory (1993). Using a general equilibrium setting, they provide simulation evidence that the relationship between expected returns and the variance of returns can go in either direction, depending on specification. Further doubt on the validity of the linearity assumption is provided in Linton and Perron (2000) using non-parametric methods.

Our results suggest that no significant risk premium exists in stock returns using any of the three methods. However, the main feature of the results is the wider confidence intervals obtained using the LM principle. Wald confidence intervals understate the uncertainty of the estimated parameters; the differences are not dramatic however. The results are also similar to those obtained from the GARCH-M(1,1) model.

**** Insert table 6 here ****

The results for the yen-dollar returns are presented next with all point estimates negative. In the case of the kernel estimator, this finding is actually significantly different from 0. The relationship for this series appears to be the least identified as all estimators have large standard errors (and the first-stage R^2 is very low). Both the Wald and LM confidence intervals are quite wide, reflecting poor identification of the model. The LM interval with the Engle-Ng estimator is even unbounded in this case.

Finally, the results of the estimation for excess holding yields present a similar picture. All point estimates are positive, with the GARCH-M result being significantly different from 0. This conclusion is the same as Engle, Lillien, and Robbins (who used a restricted ARCH-M(4) structure). For the kernel estimator, the effect is almost significant at the 5% level. Once again, the LM intervals are much wider than their Wald counterparts.

Figure 7 presents a time plot of the estimated conditional variance for all three series. Except for the excess holding yield, the Engle-Ng and GARCH-M models offer a very similar picture. On the other hand, the kernel estimates are much more volatile (not surprisingly given than they do not have an autoregressive structure) over time. The results for the excess holding yield might seem strange at first sight since the GARCH-M gives such a different picture (especially around the Volker experiment of 1979-82). The reason lies in the bandwidth choice for the estimation of the conditional mean of this series. The mean is estimated with a very small bandwidth (constant is 0.28) thus implying little smoothing of neighboring observations, and as a result, the residuals are much smaller than with GARCH-M (and hence have smaller variance).

**** Insert figure 7 here ****

7. Conclusion

This paper follows several others in showing that inference using instrumental variables is greatly affected by a low correlation between the instruments and the explanatory variables. It extends the current literature to linear semi-parametric models with non-parametrically estimated regressors and instruments and to cases with higher-order dependence. The analysis shows that the limit theory is similar to that currently available in the literature.

Simulation evidence reveals that the additional step of estimating both the regressors and the instruments may lead to a loss in the quality of asymptotic approximations. Using a semi-parametric estimator proposed by Engle and Ng (1993) and carrying out inference using Lagrange Multiplier procedures allows for inference that is more robust than the alternatives considered here.

Empirical application to three financial series suggests that conclusions may hinge on the use of appropriate confidence intervals. Using the appropriate LM confidence intervals and the semi-parametric estimator of the conditional variance leads us to conclude that none of the series considered includes a statistically significant risk premium. This differs in some cases from inference based on the usual Wald confidence intervals and on a parametric

GARCH-M model. However, because of the wide confidence intervals, the results are also consistent with the presence of large risk premia. The data is simply not informative enough to precisely estimate the relationship between risk and returns.

Further work on this problem is clearly warranted. In particular, other more commonly used estimators such as maximum likelihood are likely to face similar problems as the IV estimator analyzed here. This analysis could follow the methodology developed in Stock and Wright (2000) for GMM estimators. Finally, a critical avenue for future research is the development of techniques to diagnose cases where weak identification hinders inference using usual methods. Recent testing procedures along these lines have been suggested by Arellano, Hansen, and Sentana (1999) ,Wright (2000) , and Hahn and Hausman (2002) .

REFERENCES

Andersen, Torben G. and Tim Bollerslev, “Answering the Critics: Yes, ARCH Models Do Provide Good Volatility Forecasts”, International Economic Review, 39, November 1998, 885-905.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, “The Distribution of Realized Exchange Rate Volatility”, Journal of the American Statistical Association, 96, March 2001, 42-55.

Anderson, T. W. and Herman Rubin, “Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations”, Annals of Mathematical Statistics, 20, March 1949, 46-63.

Andrews, Donald W. K., “Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity”, Econometrica, 62, January 1994, 43-72.

Andrews, Donald W. K., “Examples of MINPIN Estimators: Supplement to Asymptotics for Semiparametric Models via Stochastic Equicontinuity”, manuscript, Yale University, 1992.

Andrews, Donald W. K., “Nonparametric Kernel Estimation for Semiparametric Models”, Econometric Theory, 11, August 1995, 560-596.

Arellano, Manuel, Lars P. Hansen, and Enrique Sentana, “Underidentification?”, manuscript, CEMFI, July 1999.

Backus, David K., Allan W. Gregory, and Stanley E. Zin, “Risk Premiums in the Term Structure: Evidence from Artificial Economies”, Journal of Monetary Economics, 24, November 1989, 371-399.

Backus, David K. and Allan W. Gregory , “Theoretical Relations Between Risk Premiums and Conditional Variances,” Journal of Business and Economic Statistics, 11, April 1993, 177-185.

Bollerslev, Tim, “Generalized Conditional Heteroskedasticity”, Journal of Econometrics, 31, April 1986, 307-327.

Bottazzi, Laura and Valentina Corradi, “Analysing the Risk Premium in the Italian Stock Market: ARCH-M Models versus Non-parametric Models”, Applied Economics, 23, March 1991, 535-542.

Braun, Phillip A., Daniel B. Nelson, and Alain M. Sunier, “Good News, Bad News, Volatility and Betas”, Journal of Finance, 50, december 1995, 1575-1604.

Chao, John. and Norman. R. Swanson, “Bias and MSE of the IV Estimator Under Weak Identification”, manuscript, University of Maryland and Purdue University, August 2001.

Choi, In and Peter C. B. Phillips, “Asymptotic and Finite Sample Distribution Theory for the IV Estimators and Tests in Partially Identified Structural Relations”, Journal of Econometrics, 51, February 1992, 113-150.

Dufour, Jean-Marie, “Some Impossibility Theorems in Econometrics with Applications to Instrumental Variables, Dynamic Models, and Cointegration”, Econometrica, 65, November 1997, 1365-1387.

Dufour, Jean-Marie and Joann Jasiak, “Finite Sample Inference Methods for Simultaneous Equations and Models with Unobserved and Generated Regressors”, International Economic Review, 42, August 2001, 815-843.

Dufour, Jean-Marie and Mohamed Taamouti, “Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instru-ments”, manuscript, Université de Montréal, 2001.

Engle, Robert F., and Victor K. Ng, “Measuring and Testing the Impact of News on Volatility”, Journal of Finance, 48, December 1993, 1749-1778.

Engle, Robert F., Lilien, David M., and Russell P. Robins, "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model", Econometrica, 55, March 1987, 391-407.

French, Kenneth, G. William Schwert, and Robert F. Stambaugh, "Expected Stock Returns and Volatility", Journal of Financial Economics, 19, September 1987, 3-29.

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle, "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", Journal of Finance, 48, December 1993, 1779-1801.

Hahn, Jinyong and Jerry Hausman, "A New Specification Test for the Validity of Instrumental Variables", Econometrica, 70, January 2002, 163-189.

Hall, Alastair R., Glenn D. Rudebusch, and David W. Wilcox, "Judging Instrument Relevance in Instrumental Variables Estimation", International Economic Review, 37, May 1996, 283-298.

Lettau, Martin and Sydney Ludvigson, "Measuring and Modeling Variation in the Risk-Return Tradeoff", manuscript, New York University and Federal Reserve Bank of New York, October 2001.

Linton, Oliver and Benoit Perron, "The Shape of the Risk Premium: Evidence From a Semiparametric GARCH Model", CRDE working paper 0899, Université de Montréal, 1999 (revised February 2002).

Masry, Elias and Dag Tjostheim, "Nonparametric Estimation and Identification of Nonlinear Time Series: Strong Convergence and Asymptotic Normality", Econometric Theory, 11, June 1995, 258-289.

Nelson, Charles R., Startz, Richard, "The Distribution of the Instrumental Variables Estimator and its t-ratio when the Instrument is a Poor One", Journal of Business, 63, January 1990, S125-S140.

Pagan, Adrian, "Econometric Issues in the Analysis of Regressions with Generated Regressors", International Economic Review, 25, February 1984, 221-247.

Pagan, Adrian R. and Y. S. Hong, "Nonparametric Estimation and the Risk Premium" in Barnett, William A., James Powell, and George E. Tauchen eds., Nonparametric and Semiparametric Methods in Econometrics and Statistics: Proceedings of the Fifth International Symposium in Economic Theory and Econometrics, Cambridge: Cambridge University Press, 1991, 51-75.

Pagan, Adrian and Aman Ullah. "The Econometric Analysis of Models with Risk Terms", Journal of Applied Econometrics, 3, April 1988, 87-105.

Perron, Benoit, "A Monte Carlo Comparison of Non-parametric Estimators of the Conditional Variance", manuscript, Université de Montréal, 1999.

Phillips, Peter C. B., "Partially Identified Econometric Models", Econometric Theory, 5, August 1989, 181-240.

Sentana, Enrique and Sushil Wadhwani, "Semi-parametric Estimation and the Predictability of Stock Market Returns: Some Lessons from Japan", Review of Economics Studies, 58, May 1991, 547-563.

Staiger, Douglas and James H. Stock, "Instrumental Variables Regression with Weak Instruments", Econometrica, 65, May 1997, 557-586.

Startz, Richard, Charles R. Nelson, and Eric Zivot, "Improved Inference for the Instrumental Variable Estimator", manuscript, Univesity of Washington, February 2001.

Stock, James and Jonathan Wright, "GMM with Weak Identification", Econometrica, 68, September 2000, 1055-1096.

Tjostheim, Dag and Bjorn H. Auestad, "Nonparametric Identification of Nonlinear Time Series: Selecting Significant Lags", Journal of the American Statistical Association, 89, december 1994, 1410-1419.

Wang, Jiahui and Eric Zivot, “Inference on a Structural Parameter in Instrumental Regression with Weak Instruments”, Econometrica, 66, November 1998, 1389-1404.

Wright, Jonathan, “Detecting Lack of Identification in GMM”, International Finance Discussion Paper 674, Board of Governors of the Federal Reserve System, July 2000.

Zivot, Eric, Richard Startz and Charles R. Nelson “Valid Confidence Intervals and Inference in the Presence of Weak Instruments”, International Economic Review, 39, November 1998, 1119-1144.

8. Appendix

A. Proofs

A.1. Preliminary results

Before proving the results in the paper, we will collect the required preliminaries in the following lemma.

Lemma A.1. Suppose the conditions of theorem (4.1) are satisfied. Then, the following hold:

1. $\frac{1}{\sqrt{n}}\left(\hat{Z}'M_X\hat{Y}\right) = \frac{1}{\sqrt{n}}\left(Z'M_XY\right) + o_p(1)$
2. $\frac{1}{\sqrt{n}}\left[\hat{Z}'M_X(Z-Y)\delta\right] = \frac{1}{\sqrt{n}}\left[Z'M_X(Z-Y)\delta\right] + o_p(1)$
3. $\frac{1}{\sqrt{n}}\left(\hat{Z}'M_Xe\right) = \frac{1}{\sqrt{n}}\left(Z'M_Xe\right) + o_p(1)$
4. $\frac{1}{n}\left(\hat{Z}'M_X\hat{Z}\right) = \frac{1}{n}\left(Z'M_XZ\right) + o_p(1)$
5. $\frac{1}{\sqrt{n}}X'\hat{Y} = \frac{1}{\sqrt{n}}X'Y + o_p(1)$
6. $\frac{1}{\sqrt{n}}\left[\hat{Z}'M_X\left(Y-\hat{Y}\right)\delta\right] \xrightarrow{p} 0$
7. $\frac{1}{\sqrt{n}}\left(\hat{Z}'M_Xu\right) = \Psi_{Zu} + o_p(1)\,.$

Proof. To prove the first result, note that

$$\begin{aligned}
\frac{1}{\sqrt{n}}\widehat{Z}'M_X\widehat{Y} &= \frac{1}{\sqrt{n}}\left(\widehat{Z}-Z\right)'M_X\left(\widehat{Y}-Y\right)+\frac{1}{\sqrt{n}}\left(\widehat{Z}-Z\right)'M_XY \\
&\quad +\frac{1}{\sqrt{n}}Z'M_X\left(\widehat{Y}-Y\right)+\frac{1}{\sqrt{n}}Z'M_XY \\
&= \frac{1}{\sqrt{n}}\left(\widehat{Z}-Z\right)'M_X\left(\widehat{Y}-Y\right)+\frac{1}{\sqrt{n}}\left(\widehat{Z}-Z\right)'M_XY \\
&\quad +\frac{1}{\sqrt{n}}Z'M_X\left(\widehat{Y}-Y\right)+\frac{1}{\sqrt{n}}Z'M_XY \\
&= \frac{1}{\sqrt{n}}Z'M_XY+\frac{1}{\sqrt{n}}\left(\widehat{Z}-Z\right)'M_X\left(\widehat{Y}-Y\right)+\frac{1}{\sqrt{n}}\left(\widehat{Z}-Z\right)'M_X\left(Y-Z\right) \\
&\quad +\frac{1}{\sqrt{n}}\left(\widehat{Z}-Z\right)'M_X\left[Z-E\left(Z\right)\right]+\frac{1}{\sqrt{n}}\left(\widehat{Z}-Z\right)'M_XE\left(Z\right) \\
&= \frac{1}{\sqrt{n}}Z'M_XY+A_1+A_2+A_3+A_4
\end{aligned}$$

We will next bound each of the A_i, $i=1,\ldots,4$. Let $|A|$ be the matrix norm of A. First,

$$\begin{aligned}
|A_1| &= \left\|\frac{1}{\sqrt{n}}\left(\widehat{Z}-Z\right)'M_X\left(\widehat{Y}-Y\right)\right\| \\
&\leq \left\|\widehat{Z}-Z\right\|\left\|\frac{1}{\sqrt{n}}M_X\left(\widehat{Y}-Y\right)\right\| \\
&= o_p\left(1\right)
\end{aligned}$$

by assumptions 1 and 2. Next,

$$\begin{aligned}
|A_2| &= \left\|\frac{1}{\sqrt{n}}\left(\widehat{Z}-Z\right)'M_X\left(Y-Z\right)\right\| \\
&\leq \left\|\widehat{Z}-Z\right\|\left\|\frac{1}{\sqrt{n}}M_X\left(Y-Z\right)\right\| \\
&= o_p\left(1\right)
\end{aligned}$$

by assumption 2 and since the quantity inside the second norm will be $O_p\left(1\right)$. The third term is:

$$\begin{aligned}
|A_3| &= \left\|\frac{1}{\sqrt{n}}\left(\widehat{Z}-Z\right)'M_X\left[Z-E\left(Z\right)\right]\right\| \\
&\leq \left\|\widehat{Z}-Z\right\|\left\|\frac{1}{\sqrt{n}}M_X\left[Z-E\left(Z\right)\right]\right\| \\
&= o_p\left(1\right)
\end{aligned}$$

again by assumption 2 and since the term inside the second norm is $O_p(1)$. Finally, the fourth term can be bounded as:

$$\begin{aligned} |A_4| &= \left| \frac{1}{\sqrt{n}} \left(\hat{Z} - Z\right)' M_X E(Z) \right| \\ &\leq \left| \frac{1}{\sqrt{n}} \left(\hat{Z} - Z\right)' M_X \right| |E(Z)| \\ &= o_p(1) \end{aligned}$$

as $\hat{Z} \xrightarrow{p} Z$ and $|E(Z)| < \infty$ with probability one since $Z_t < \infty$ for all t. Thus,

$$\frac{1}{\sqrt{n}} \hat{Z}' M_X \hat{Y} = \frac{1}{\sqrt{n}} Z' M_X Y + o_p(1)$$

as required.

The second result is obtained as:

$$\begin{aligned} \frac{1}{\sqrt{n}} \left[\hat{Z}' M_X (Z - Y)\delta\right] &= \frac{1}{\sqrt{n}} \left[\left(\hat{Z} - Z\right)' M_X (Z - Y)\delta\right] + \frac{1}{\sqrt{n}} \left[Z' M_X (Z - Y)\delta\right] \\ &= \frac{1}{\sqrt{n}} \left[Z' M_X (Z - Y)\delta\right] + o_p(1) \end{aligned}$$

where the last line follows from:

$$\begin{aligned} \left| \frac{1}{\sqrt{n}} \left(\hat{Z} - Z\right)' M_X (Z - Y)\delta \right| &\leq \left|\hat{Z} - Z\right| \left| \frac{1}{\sqrt{n}} M_X (Z - Y)\delta \right| \\ &= o_p(1) \cdot O_p(1) \\ &= o_p(1) \end{aligned}$$

The third result follows from:

$$\frac{1}{\sqrt{n}} \hat{Z}' M_X e = \frac{1}{\sqrt{n}} \left[\left(\hat{Z} - Z\right)\right]' M_X e + \frac{1}{\sqrt{n}} Z' M_X e$$

and noting that the first term can be bounded by:

$$\begin{aligned} \left| \frac{1}{\sqrt{n}} \left[\left(\hat{Z} - Z\right)\right]' M_X e \right| &\leq \left|\hat{Z} - Z\right| \left| \frac{1}{\sqrt{n}} M_X e \right| \\ &= o_p(1) \cdot O_p(1) \\ &= o_p(1) \end{aligned}$$

by assumptions 2 and 4.

The fourth result is proven by rewriting the left hand side as:

$$\begin{aligned}\frac{1}{n}\hat{Z}'M_X\hat{Z} &= \frac{1}{n}\left(\hat{Z}-Z\right)'M_X\hat{Z}+\frac{1}{n}Z'M_X\hat{Z}\\ &= \frac{1}{n}Z'M_XZ+\frac{1}{n}\left(\hat{Z}-Z\right)'M_X\left(\hat{Z}-Z\right)+\frac{1}{n}\left(\hat{Z}-Z\right)'M_XZ+\frac{1}{n}Z'M_X\left(\hat{Z}-Z\right)\\ &= \frac{1}{n}Z'M_XZ+B_1+B_2+B_2'\end{aligned}$$

where B_j, $j=1,2$, is bounded in turn by an $o_p(1)$ term. For B_1, we do so as:

$$\begin{aligned}|B_1| &= \left|\frac{1}{n}\left(\hat{Z}-Z\right)'M_X\left(\hat{Z}-Z\right)\right|\\ &\leq \left|\frac{1}{\sqrt{n}}\left(\hat{Z}-Z\right)'\iota\right||M_X|\left|\frac{1}{\sqrt{n}}\iota'\left(\hat{Z}-Z\right)\right|\\ &= o_p(1)\end{aligned}$$

by assumption 2 where ι is a vector of ones. The second term is bounded as:

$$\begin{aligned}|B_2| &= \left|\frac{1}{n}\left(\hat{Z}-Z\right)'M_XZ\right|\\ &\leq \left|\hat{Z}-Z\right|\left|\frac{1}{n}M_XZ\right|\\ &= o_p(1)\cdot O_p(1)\\ &= o_p(1)\end{aligned}$$

by assumption 2. The fourth result follows.

The fifth result is obtained as:

$$\begin{aligned}\frac{1}{\sqrt{n}}X'\hat{Y} &= \frac{1}{\sqrt{n}}X'Y+\frac{1}{\sqrt{n}}X'\left(\hat{Y}-Y\right)\\ &= \frac{1}{\sqrt{n}}X'Y+o_p(1)\end{aligned}$$

by assumption 1.

The sixth result is obtained from the decomposition:

$$\begin{aligned}
\left|\frac{1}{\sqrt{n}}\left[\widehat{Z}'M_X\left(Y-\widehat{Y}\right)\delta\right]\right| &= \left|\frac{1}{\sqrt{n}}\left(\left(\widehat{Z}-Z\right)'M_X\left(Y-\widehat{Y}\right)\delta\right)+\frac{1}{\sqrt{n}}\left[Z'M_X\left(Y-\widehat{Y}\right)\delta\right]\right| \\
&\leq \left|\frac{1}{\sqrt{n}}\left(\left(\widehat{Z}-Z\right)'\left(Y-\widehat{Y}\right)\delta\right)\right| \\
&\quad +\left|\frac{1}{\sqrt{n}}\left[Z-E\left(Z\right)\right]'M_X\left[\left(Y-\widehat{Y}\right)\delta\right]\right|+\left|\frac{1}{\sqrt{n}}E\left(Z\right)'M_X\left(Y-\widehat{Y}\right)\delta\right| \\
&\leq \left|\widehat{Z}-Z\right|\left|\frac{1}{\sqrt{n}}\iota'\left(Y-\widehat{Y}\right)\delta\right| \\
&\quad +\left|Z-E\left(Z\right)\right|\left|\frac{1}{\sqrt{n}}\iota'\left(Y-\widehat{Y}\right)\delta\right|+\left|E\left(Z\right)\right|\left|\frac{1}{\sqrt{n}}\iota'\left(Y-\widehat{Y}\right)\delta\right| \\
&= o_p\left(1\right)
\end{aligned}$$

where the last line follows from assumption 1 and $E\left(Z\right)<\infty$.

Finally, the last result is obtained by rewriting the left hand side as:

$$\frac{1}{\sqrt{n}}\widehat{Z}'M_Xu=\frac{1}{\sqrt{n}}\widehat{Z}'M_Xe+\frac{1}{\sqrt{n}}\widehat{Z}'M_X\left(Y-\widehat{Y}\right)\delta+\frac{1}{\sqrt{n}}\widehat{Z}'M_X\left(Z-Y\right)\delta$$

and using results 2, 3, and 6 of the lemma. ■

A.2. Proof of theorem 4.1

The instrumental variable estimator of δ is

$$\widehat{\delta}-\delta=\left(\widehat{Z}'M_X\widehat{Y}\right)^{-1}\widehat{Z}'M_Xu$$

To derive the asymptotic distribution, we use the first result of the lemma to obtain:

$$\begin{aligned}
\frac{1}{\sqrt{n}}\widehat{Z}'M_X\widehat{Y} &= \frac{1}{\sqrt{n}}Z'M_XY+o_p\left(1\right) \\
&= \frac{1}{\sqrt{n}}\left[Z'M_X\left(Z\Pi+V\right)\right] \\
&= \frac{1}{n}Z'M_XZG+\frac{1}{\sqrt{n}}Z'M_XV \\
&\xrightarrow{d} \left(\Sigma_{ZZ}G+\Psi_{ZV}\right) \\
&= \sigma_{ZV}^{\frac{1}{2}}\left(\sigma_{ZV}^{-\frac{1}{2}}\Sigma_{zz}G+z_v\right) \\
&= \sigma_{ZV}^{\frac{1}{2}}\left(\lambda+z_v\right)
\end{aligned}$$

while $\frac{1}{\sqrt{n}}(Z'M_Xu) \xrightarrow{d} \Psi_{Zu} = \sigma_{Zu}^{1/2} z_u$ by assumption. Putting these pieces together gives us the desired result for the distribution of $\hat{\theta}$, $\hat{\theta} - \delta \xrightarrow{d} \Xi$.

To derive the distribution of $\hat{\gamma}$, note that:

$$\begin{aligned}\hat{\gamma} &= (X'X)^{-1} X' \left(y - \hat{Y}\hat{\theta}\right) \\ &= \gamma + (X'X)^{-1} X'\hat{Y}\left(\delta - \hat{\theta}\right) + (X'X)^{-1} X'u\end{aligned}$$

so that

$$\begin{aligned}\sqrt{n}\,(\hat{\gamma} - \gamma) &= \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'\hat{Y}}{\sqrt{n}}\right)\left(\delta - \hat{\theta}\right) + \left(\frac{X'X}{n}\right)^{-1}\frac{X'u}{\sqrt{n}} \\ &= \left(\frac{X'X}{n}\right)^{-1}\left(\frac{X'Y}{\sqrt{n}}\right)\left(\delta - \hat{\theta}\right) + \left(\frac{X'X}{n}\right)^{-1}\frac{X'u}{\sqrt{n}} + o_p(1) \\ &\xrightarrow{d} -\Sigma_{XX}^{-1}\left(\Sigma_{XZ}G + \Sigma_{XX}H + \Psi_{XV}\right)\Xi + \Sigma_{XX}^{-1}\psi_{Xu}\end{aligned}$$

where the term in parentheses is derived from:

$$\begin{aligned}\frac{1}{\sqrt{n}}X'Y &= \frac{1}{\sqrt{n}}X'(Z\Pi + X\Gamma + V) \\ &= \frac{1}{n}X'ZG + \frac{1}{n}X'XH + \frac{1}{\sqrt{n}}X'V \\ &\xrightarrow{d} \Sigma_{XZ}G + \Sigma_{XX}H + \Psi_{XV}\end{aligned}$$

by assumption. ■

A.3. Proof of Proposition 4.2

From the proof of theorem 4.1, $n^{-\frac{1}{2}}\hat{Z}'M_X\hat{Y} \xrightarrow{d} \sigma_{ZV}^{\frac{1}{2}}(\lambda + z_v)$. The only part that remains to derive is the limiting behavior of $\lim_{n\to\infty} var\left(\frac{1}{\sqrt{n}}\hat{Z}'M_X\hat{u}\right)$. The residual orthogonal to X, $M_X\hat{u}$ can be written as $M_X\left(y - \hat{Y}\hat{\theta}\right) = M_Xu - M_X\hat{Y}\left(\hat{\theta} - \delta\right)$ and the term of interest is therefore:

$$
\begin{aligned}
\lim_{n\to\infty} var\left(\frac{1}{\sqrt{n}}\hat{Z}'M_X\hat{u}\right) &= \lim_{n\to\infty} var\left(\frac{1}{\sqrt{n}}\hat{Z}'M_X'\left[M_Xu - M_X\hat{Y}\left(\hat{\theta}-\delta\right)\right]\right)\\
&= \lim_{n\to\infty} var\left(\frac{1}{\sqrt{n}}\left[\hat{Z}'M_X'u - \hat{Z}'M_X'\hat{Y}\left(\hat{\theta}-\delta\right)\right]\right)\\
&= \lim_{n\to\infty} var\left(\frac{1}{\sqrt{n}}\left[Z'M_X'u - Z'M_X'\left(Z\Pi+V\right)\left(\hat{\theta}-\delta\right)\right]+o_p\left(1\right)\right)\\
&= \lim_{n\to\infty} var\left(\frac{1}{\sqrt{n}}\sum\left[Z_t^{\perp}u_t' - Z_t^{\perp}Z_t^{\perp\prime}\Pi\left(\hat{\theta}-\delta\right) - Z_t^{\perp}V_t'\left(\hat{\theta}-\delta\right)\right]+o_p\left(1\right)\right)\\
&= \lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}u_t'u_sZ_s^{\perp}\\
&\quad+\lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}Z_t^{\perp\prime}\Pi\left(\hat{\theta}-\delta\right)\left(\hat{\theta}-\delta\right)'\Pi'Z_s^{\perp}Z_s^{\perp\prime}\\
&\quad+\lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}V_t'\left(\hat{\theta}-\delta\right)\left(\hat{\theta}-\delta\right)'V_sZ_s^{\perp\prime}\\
&\quad-\lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}u_t'\left(\hat{\theta}-\delta\right)'\Pi'Z_s^{\perp}Z_s^{\perp\prime}\\
&\quad-\left(\lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}u_t'\left(\hat{\theta}-\delta\right)'\Pi'Z_s^{\perp}Z_s^{\perp\prime}\right)'\\
&\quad-\lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}u_t'\left(\hat{\theta}-\delta\right)'V_sZ_s^{\perp\prime}\\
&\quad-\left(\lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}u_t'\left(\hat{\theta}-\delta\right)'V_sZ_s^{\perp\prime}\right)'\\
&\quad+\lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}Z_t^{\perp\prime}\Pi\left(\hat{\theta}-\delta\right)\left(\hat{\theta}-\delta\right)'V_sZ_s^{\perp\prime}\\
&\quad+\left(\lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}Z_t^{\perp\prime}\Pi\left(\hat{\theta}-\delta\right)\left(\hat{\theta}-\delta\right)'V_sZ_s^{\perp\prime}\right)'\\
&= \sigma_{Zu}+C_1+C_2-C_3-C_3'-C_4-C_4'+C_5+C_5'
\end{aligned}
$$

The second term is

$$
\begin{aligned}
C_1 &= +\lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}Z_t^{\perp\prime}\Pi\left(\hat{\theta}-\delta\right)\left(\hat{\theta}-\delta\right)'\Pi'Z_s^{\perp}Z_s^{\perp\prime}\\
&= \lim_{n\to\infty}\frac{1}{n^2}\sum_s\sum_t Z_t^{\perp}Z_t^{\perp\prime}G\Xi\Xi'G'Z_s^{\perp}Z_s^{\perp\prime}\\
&= o_p\left(1\right)
\end{aligned}
$$

while the third one is:

$$\begin{aligned} C_2 &= \lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}V_t'\left(\hat{\theta}-\delta\right)\left(\hat{\theta}-\delta\right)'V_sZ_s^{\perp\prime} \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}V_t'\Xi\Xi'V_sZ_s^{\perp\prime} \\ &= \sigma_{ZV}^{*}. \end{aligned}$$

The next term is:

$$\begin{aligned} C_3 &= \lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}u_t'\left(\hat{\theta}-\delta\right)'\Pi'Z_s^{\perp}Z_s^{\perp\prime} \\ &= \lim_{n\to\infty}\frac{1}{n^{\frac{3}{2}}}\sum_s\sum_t Z_t^{\perp}u_t'\Xi'G'Z_s^{\perp}Z_s^{\perp\prime} \\ &= o_p(1) \end{aligned}$$

while the fifth term in the sum is:

$$\begin{aligned} C_4 &= \lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}u_t'\left(\hat{\theta}-\delta\right)'V_sZ_s^{\perp\prime} \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}u_t'\Xi'V_sZ_s^{\perp\prime} \\ &= \sigma_{uv}^{*} \end{aligned}$$

and finally

$$\begin{aligned} C_5 &= \lim_{n\to\infty}\frac{1}{n}\sum_s\sum_t Z_t^{\perp}Z_t^{\perp\prime}\Pi\left(\hat{\theta}-\delta\right)\left(\hat{\theta}-\delta\right)'V_sZ_s^{\perp\prime} \\ &= \lim_{n\to\infty}\frac{1}{n^{\frac{3}{2}}}\sum_s\sum_t Z_t^{\perp}Z_t^{\perp\prime}G\Xi\Xi'V_sZ_s^{\perp\prime} \\ &= o_p(1) \qquad \blacksquare \end{aligned}$$

A.4. Proof of Proposition 4.3

By result 7 of the lemma, $\sqrt{n}g \xrightarrow{d} \Psi_{Zu} \overset{d}{=} N(0,\sigma_{Zu})$ under the null hypothesis, while $\hat{\sigma}_{Zu,0} \xrightarrow{p}, \sigma_{Zu}$. Standard arguments show the desired result, $ng'\hat{\sigma}_{Zu,0}^{-1}g \xrightarrow{d} \chi^2(k_2)$. ■

A.5. Proof of Proposition 4.4

The estimator of δ^* is defined as:

$$\begin{aligned}
\widehat{\delta}^* &= \left(\widehat{Z}'M_X\widehat{Z}\right)^{-1}\widehat{Z}'M_X\left(y-\widehat{Y}\delta_0\right) \\
&= \left(\widehat{Z}'M_X\widehat{Z}\right)^{-1}\widehat{Z}'M_X\left(X\gamma^*+\widehat{Z}\delta^*+u+v\left(\delta-\delta_0\right)\right) \\
&= \delta^*+\left(\widehat{Z}'M_X\widehat{Z}\right)^{-1}\widehat{Z}'M_Xu+\left(\widehat{Z}'M_X\widehat{Z}\right)^{-1}\widehat{Z}'M_XV\left(\delta-\delta_0\right)
\end{aligned}$$

so that

$$\begin{aligned}
\sqrt{n}\left(\widehat{\delta}^*-\delta^*\right) &= \left(\frac{\widehat{Z}'M_X\widehat{Z}}{n}\right)^{-1}\frac{\widehat{Z}'M_Xu}{\sqrt{n}}+\left(\frac{\widehat{Z}'M_X\widehat{Z}}{n}\right)^{-1}\frac{\widehat{Z}'M_XV\left(\delta-\delta_0\right)}{\sqrt{n}} \\
&= \left(\frac{\widehat{Z}'M_X\widehat{Z}}{n}\right)^{-1}\frac{\widehat{Z}'M_Xu}{\sqrt{n}}
\end{aligned}$$

under the null hypothesis. By results 4 and 7 of the lemma, $\sqrt{n}\left(\widehat{\delta}^*-\delta^*\right)\rightarrow N\left(0,\Sigma_{ZZ}^{-1}\sigma_{Zu}\Sigma_{ZZ}^{-1}\right)$.

Define $\Omega=\left[\left(\widehat{Z}'M_X\widehat{Z}\right)^{-1}\widehat{\sigma}_{Zu,0}\left(\widehat{Z}'M_X\widehat{Z}\right)^{-1}\right]$. The robust AR statistic is:

$$\begin{aligned}
AR &= n\left(y-\widehat{Y}\delta_0\right)'M_X\widehat{Z}\left(\widehat{Z}'M_X\widehat{Z}\right)^{-1}\Omega^{-1}\left(\widehat{Z}'M_X\widehat{Z}\right)^{-1}\widehat{Z}'M_X\left(y-\widehat{Y}\delta\right) \\
&= n\left(y-\widehat{Y}\delta_0\right)'M_X\widehat{Z}\widehat{\sigma}_{Zu,0}^{-1}\widehat{Z}'M_X\left(y-\widehat{Y}\delta\right) \\
&= LM
\end{aligned}$$

after simplification. ■

Table 1. R^2 from regression of $\hat{e}_t^2$ on $\hat{\sigma}_t^2$ and constant (%)

	Period	Kernel	Engle-Ng
S&P 500 returns	1965:1-1997:12	5.26	2.02
Yen-dollar returns	1978:10-1998:12	5.07	0.38
Excess holding yield	1959:1-2000:2	16.35	21.56

Table 2. Parameter values for simulation experiments

$$\text{DGP: } y_t = \gamma + \delta\sigma_t^2 + \sigma_t\varepsilon_t$$

$$\sigma_t^2 = \omega + \alpha e_{t-1}^2 + \beta\sigma_{t-1}^2$$

$$\varepsilon_t \sim i.i.d.N\,(0,1) \text{ for experiments I-III}$$

$$\varepsilon_t \sim i.i.d.t\,(\nu) \text{ for experiments IV-VI}$$

Parameter	I (S&P 500)	II (yen-dollar)	III (holding yields)	IV (S&P 500 - t)	V (yen-dollar - t)	VI (holding yields - t)
γ	-0.009	0.059	0.001	-0.012	0.109	0.0005
ω	1.44×10^{-4}	8.42×10^{-4}	1.70×10^{-7}	2.03×10^{-4}	8.91×10^{-4}	2.03×10^{-7}
α	0.066	0.061	0.312	0.064	0.043	0.330
β	0.855	0	0.680	0.821	0	0.651
δ	6.676	-65.661	48.722	8.444	-115.349	36.265
ν	-	-	-	7.425	5.570	4.051
n	400	250	150	400	250	150
λ	2.145	0.757	-	-	-	-
ρ_Z	-0.472	0.953	-	-	-	-
σ_{Zu}	6.819×10^{-10}	7.304×10^{-11}	-	-	-	-
σ_{Zv}	3.404×10^{-12}	1.540×10^{-14}	-	-	-	-
R^2 (%)	2.77	0.37	-	-	-	-

Table 3. Simulation results

GARCH-M $(1,1)$ parameters - conditional normality

Method	Median	Coverage of 95% CI		First-stage	K-S statistic		Fit of	Fit of
		Wald	LM	R^2 (%)	Weak IV	Normal	instrument (%)	regressor (%)
Experiment I - S&P data (true slope parameter = 6.676)								
Infeasible	7.90	97.4	93.7	2.19	0.087	0.107	100.0	100.0
Kernel	2.79	92.3	94.1	2.33	0.204	0.216	10.7	97.5
Engle-Ng	6.31	98.2	95.6	1.99	0.048	0.094	46.4	90.5
Experiment II- Yen-dollar data (true slope parameter = -65.661)								
Infeasible	-41.73	81.8	92.6	0.76	0.095	0.328	100.0	100.0
Kernel	-13.72	19.3	73.7	2.73	0.583	0.797	28.7	96.0
Engle-Ng	-28.87	66.6	97.8	2.28	0.111	0.349	22.8	88.9
Experiment III- Excess holding yield data (true slope parameter = 48.736)								
Infeasible	76.68	89.8	92.8	14.82	-	-	100.0	100.0
Kernel	72.75	88.2	93.2	14.75	-	-	30.7	85.3
Engle-Ng	79.71	96.0	94.0	12.23	-	-	66.5	77.1

Table 4. Simulation results

GARCH-M $(1,1)$ parameters - conditional t

Method	Median	Coverage of 95% CI		First-stage	Fit of	Fit of
		Wald	LM	R^2 (%)	instrument (%)	regressor (%)
Experiment IV - S&P data (true slope parameter = 8.444)						
Infeasible	9.60	97.1	94.3	1.66	100.0	100.0
Kernel	3.00	68.6	90.3	3.61	10.7	94.9
Engle-Ng	7.26	92.1	95.7	2.27	49.3	87.8
Experiment V - Yen-dollar data (true slope parameter = -115.349)						
Infeasible	-46.16	72.2	92.3	0.62	100.0	100.0
Kernel	-10.81	7.8	71.4	3.70	23.9	92.4
Engle-Ng	-32.31	48.8	98.5	1.32	20.1	89.1
Experiment VI - Excess holding yield data (true slope parameter = 36.265)						
Infeasible	54.95	95.3	91.2	17.11	100.0	100.0
Kernel	0.89	28.4	92.4	26.21	22.3	26.0
Engle-Ng	3.59	35.4	80.9	10.81	51.3	26.8

Table 5. Simulation results

Details of nonparametric estimators

Method	Mean estimation		Variance estimation		Sum of 10 squared	Instrument regression		Regressor regression	
	bandwidth	lag length	bandwidth/bins	lag length	autocorrelation	Constant	Slope	Constant	Slope
Experiment I - S&P data - conditional normality									
Kernel	1.99	1.00	1.84	1.00	2.61	0.001	0.532	0.000	1.000
Engle-Ng	1.99	1.00	2.01	1.47	2.73	0.000	0.991	0.000	1.023
Experiment II- Yen-dollar data - conditional normality									
Kernel	1.97	1.00	1.80	1.00	0.05	0.001	0.252	0.000	0.990
Engle-Ng	1.97	1.00	2.00	1.61	0.14	0.001	0.447	0.000	1.040
Experiment III- Excess holding yield data - conditional normality									
Kernel	1.82	1.00	1.32	1.00	1.81	0.000	1.141	0.000	1.013
Engle-Ng	1.82	1.00	2.40	1.09	2.03	0.000	1.899	0.000	1.041
Experiment IV - S&P data - conditional t									
Kernel	1.95	1.00	1.58	1.00	1.28	0.001	0.300	0.000	0.997
Engle-Ng	1.95	1.00	2.06	1.42	0.85	-0.000	0.879	0.000	1.027
Experiment V- Yen-dollar data - conditional t									
Kernel	1.66	1.00	1.40	1.00	0.06	0.001	0.065	0.000	0.986
Engle-Ng	1.66	1.00	2.04	1.66	0.15	0.001	0.248	0.000	1.057
Experiment VI- Excess holding yield data - conditional t									
Kernel	1.05	1.86	0.69	1.82	0.51	0.005	0.015	0.007	0.015
Engle-Ng	1.05	1.86	3.58	1.39	1.53	0.001	0.078	0.007	0.018

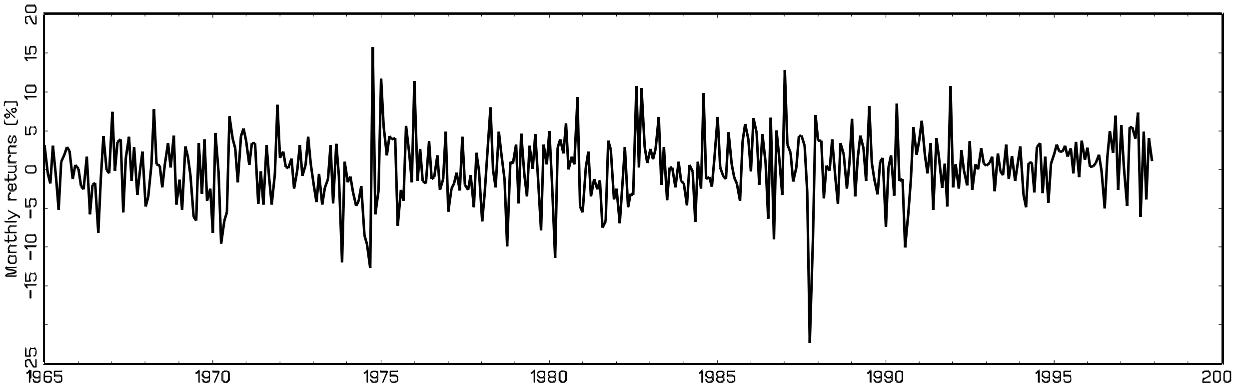
Table 6. Estimation results

Robust standard errors in parentheses

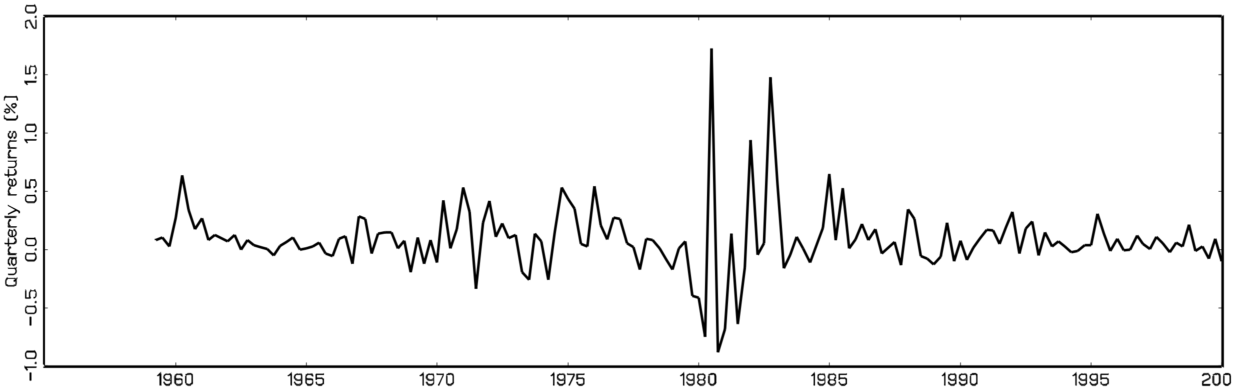
Estimator		Kernel	Engle-Ng	GARCH-M
	constant	-0.002 (0.007)	-0.006 (0.010)	-0.009 (0.010)
S&P 500 returns	$\hat{\sigma}_t^2$	2.112 (4.155)	4.585 (5.916)	6.676 (5.852)
1965:1-1997:12	Wald 95% CI	$[-6.0.10.2]$	$[-7.0, 16.1]$	$[-4.8, 18.1]$
	LM 95% CI	$[-9.4, 11.3]$	$[-12.7, 20.7]$	
	1st stage R^2 (%)	5.26	2.02	
	Mean estimation: bandwidth	1.09	1.09	
	Mean esimation: lag length	1	1	
	Variance estimation: bandwidth/bins	1.01	2	
	Variance estimation: lag length	1	1	
	constant	0.014 (0.005)	0.158 (0.280)	0.059 (0.087)
Yen-dollar returns	$\hat{\sigma}_t^2$	-20.041 (6.537)	-200.103 (349.85)	-65.661 (98.284)
1978:10-1998:12	Wald 95% CI	$[-32.8, -7.3]$	$[-885.8, 485.6]$	$[-240.2, 145.1]$
	LM 95% CI	$[-34.3, 53.8]$	$[-\infty, -43.2] \cup [51.2, \infty]$	
	1st stage R^2 (%)	5.07	0.38	
	Mean estimation: bandwidth	1.09	1.09	
	Mean estimation: lag length	1	1	
	Variance estimation: Bandwith/bins	0.52	2	
	Variance estimation: lag length	1	1	
	constant $\left(\times 10^{-2}\right)$	0.007 (0.030)	0.059 (0.017)	0.001 (0.001)
Excess holding yield	$\hat{\sigma}_t^2$	184.271 (94.918)	41.903 (35.754)	48.722 (16.508)
1959:1-2000:2	Wald 95% CI	$[-1.7, 370.3]$	$[-28.1, 111.9]$	$[16.4, 81.1]$
	LM 95% CI	$[-1000, 1000]$	$[-235.4, 110.3]$	
	1st stage R^2 (%)	16.35	21.56	
	Mean estimation: bandwidth	0.28	0.28	
	Mean estimation: lag length	1	1	
	Variance estimation: bandwidth/bins	0.06	4	
	Variance estimation: lag length	1	1	

Figure 1. Data

Returns on S&P 500 index
1965-1997



Excess holding yield
1959:1-2000:2



Yen-dollar spot returns
1978-1998

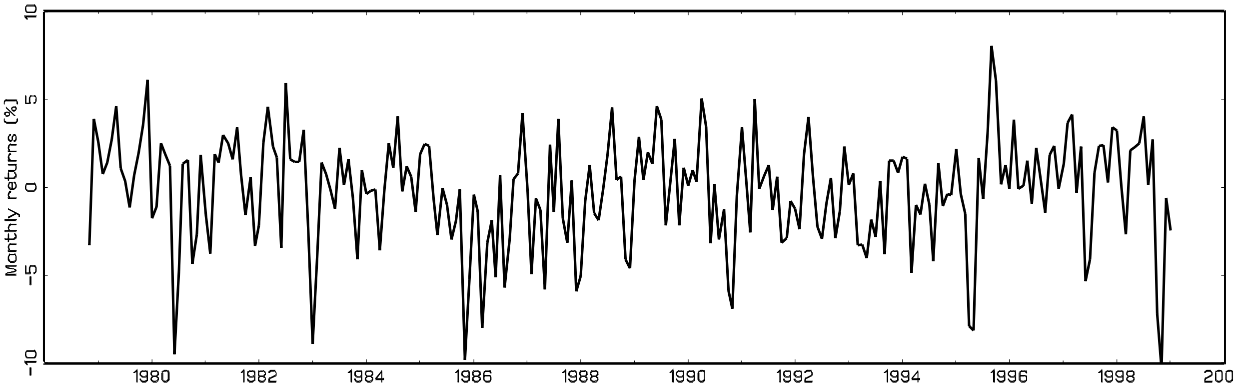


Fig. 2. Theoretical R-squared in GARCH(1,1) model

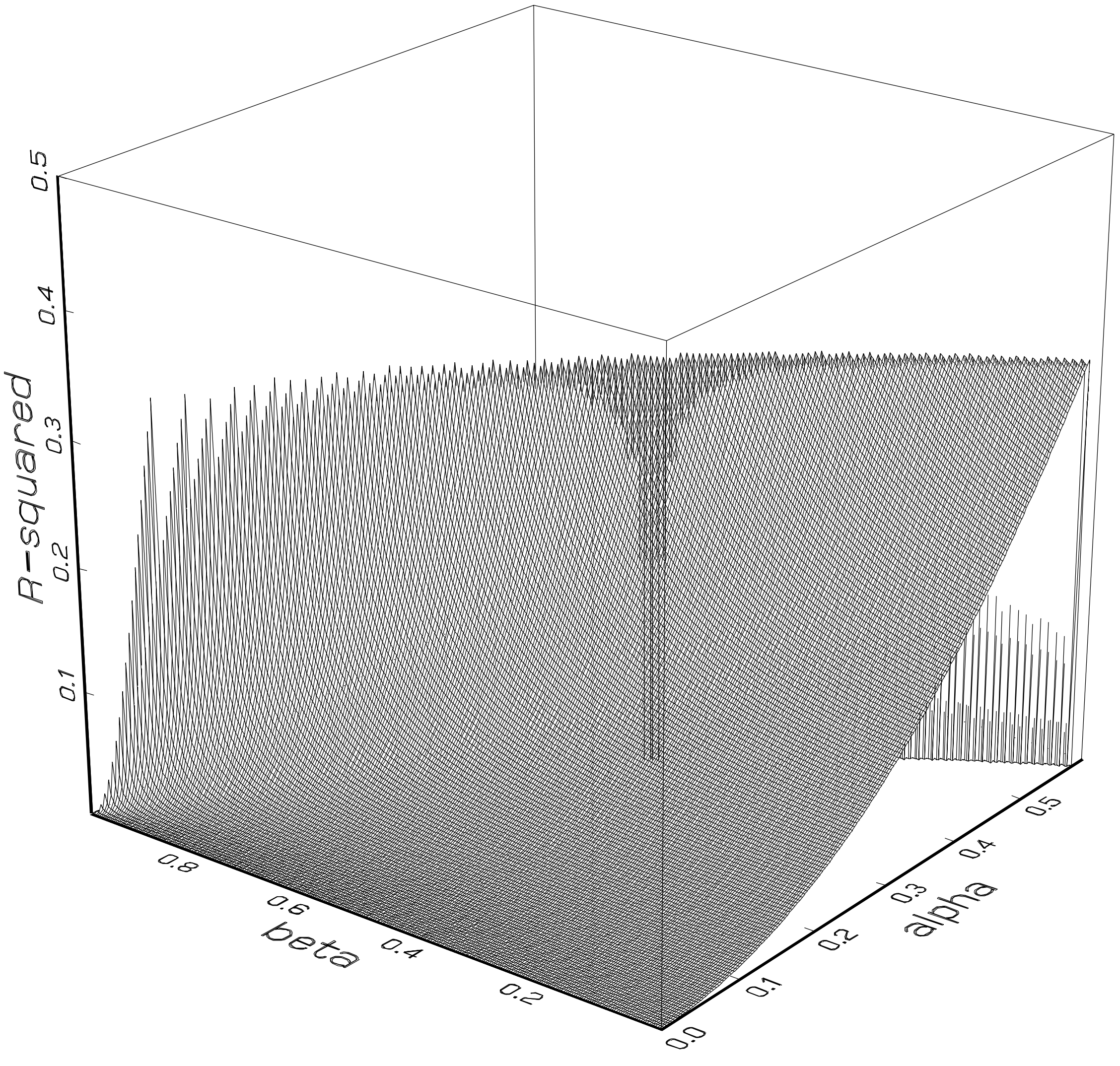
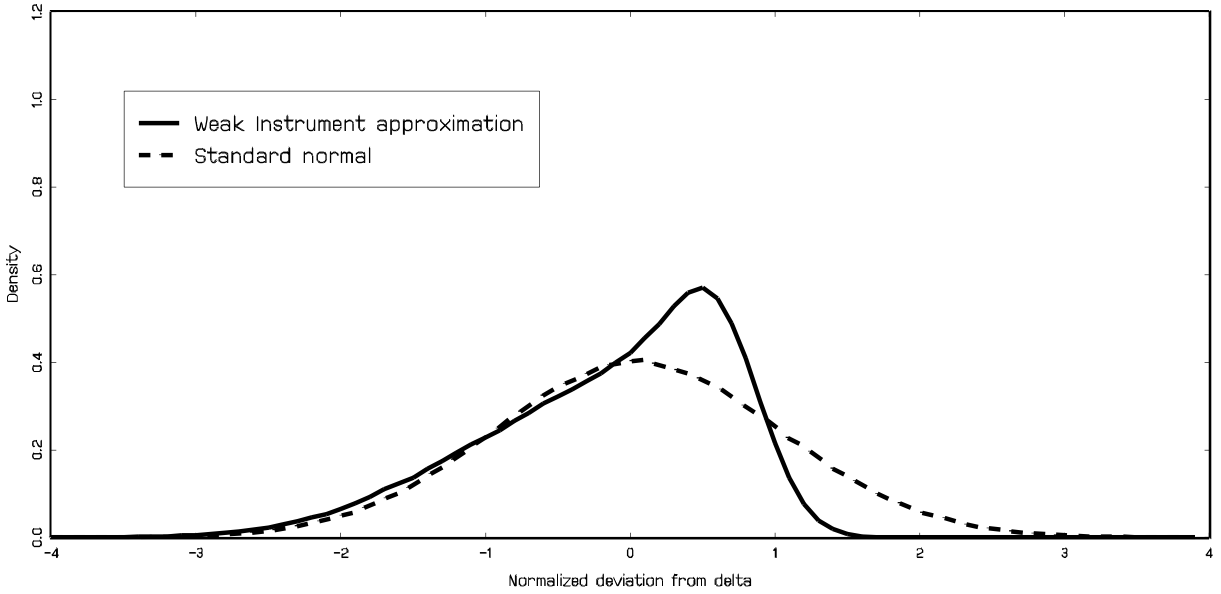


Fig 3. Distribution of t-statistic - GARCH-M(1,1) model

Experiment 1 - Estimates from S&P 500 data



Experiment 2 - Estimates from yen-dollar data

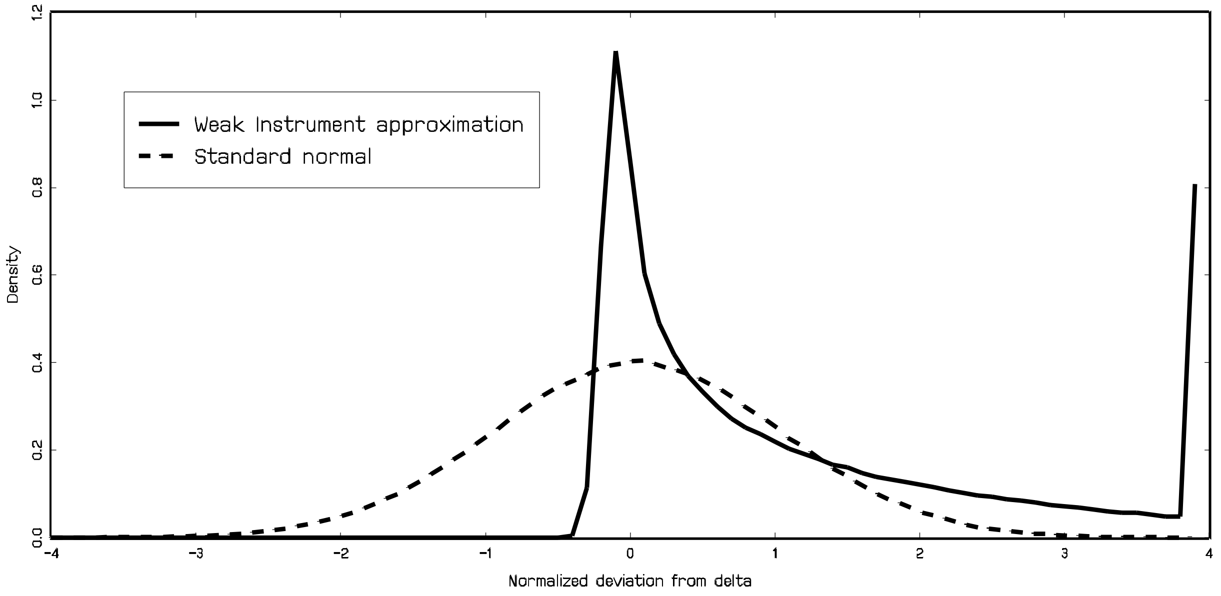
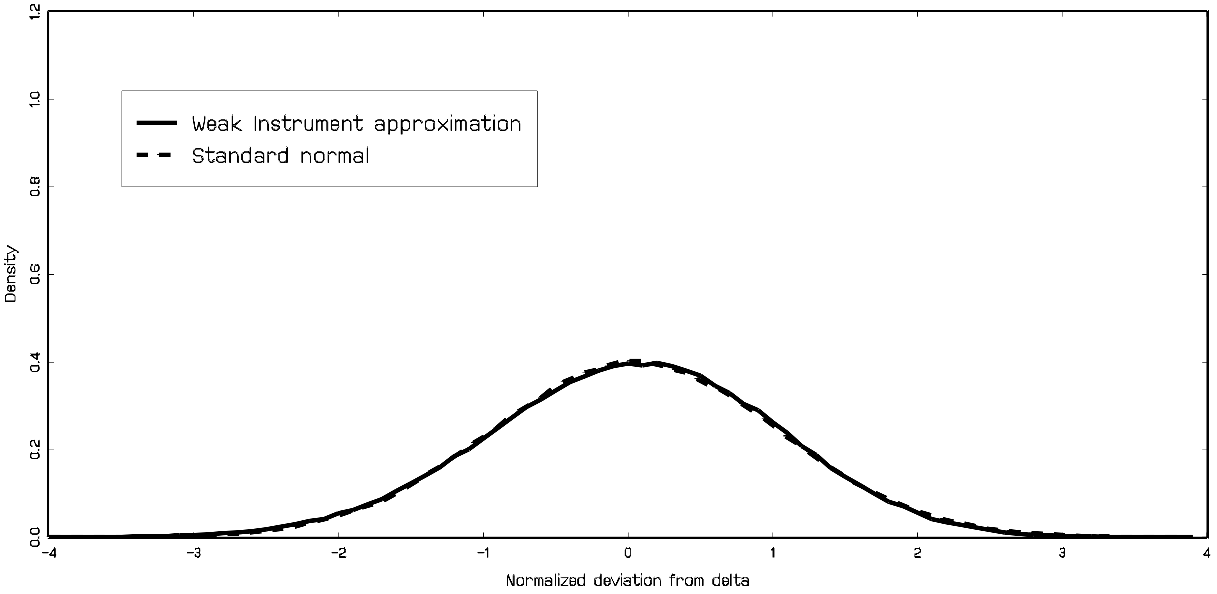


Fig 4. Distribution of t-statistic - GARCH-M(1,1) model

Experiment 1 - Estimates from S&P 500 data



Experiment 2 - Estimates from yen-dollar data

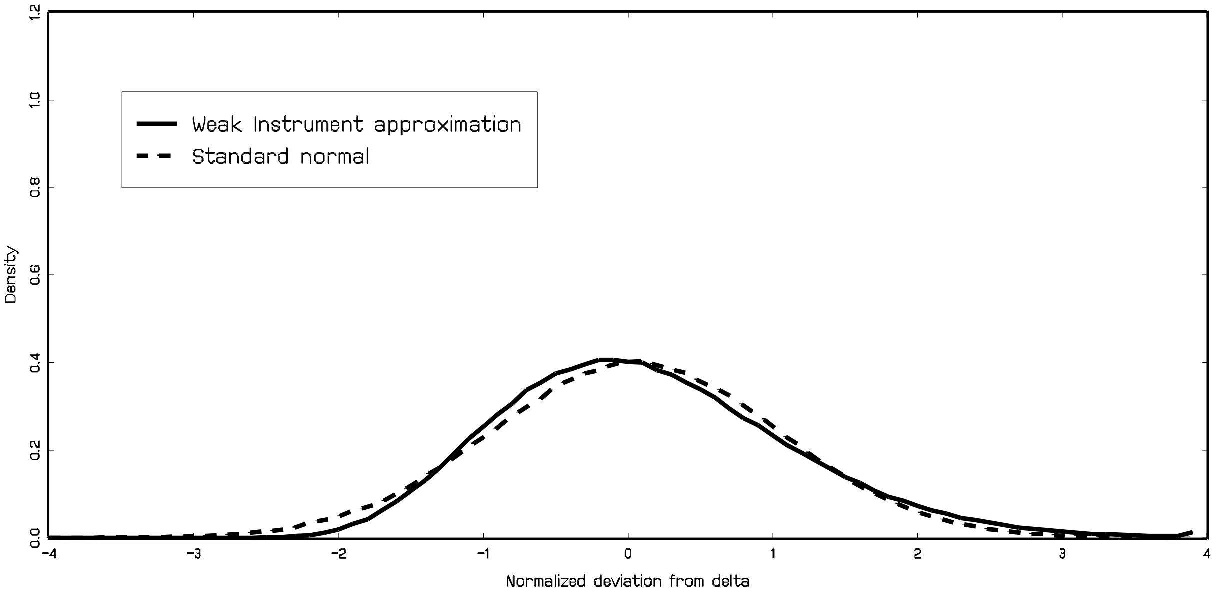
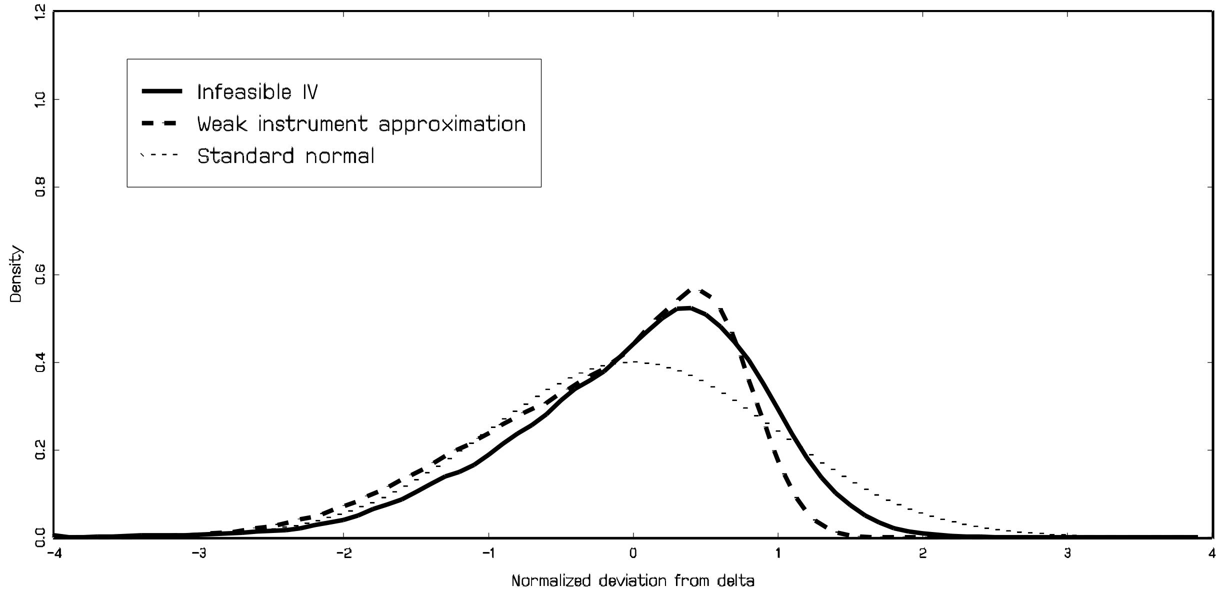


Fig 5. Distribution of infeasible IV estimator

Experiment 1 - Estimates from S&P 500 data



Experiment 2 - Estimates from yen-dollar data

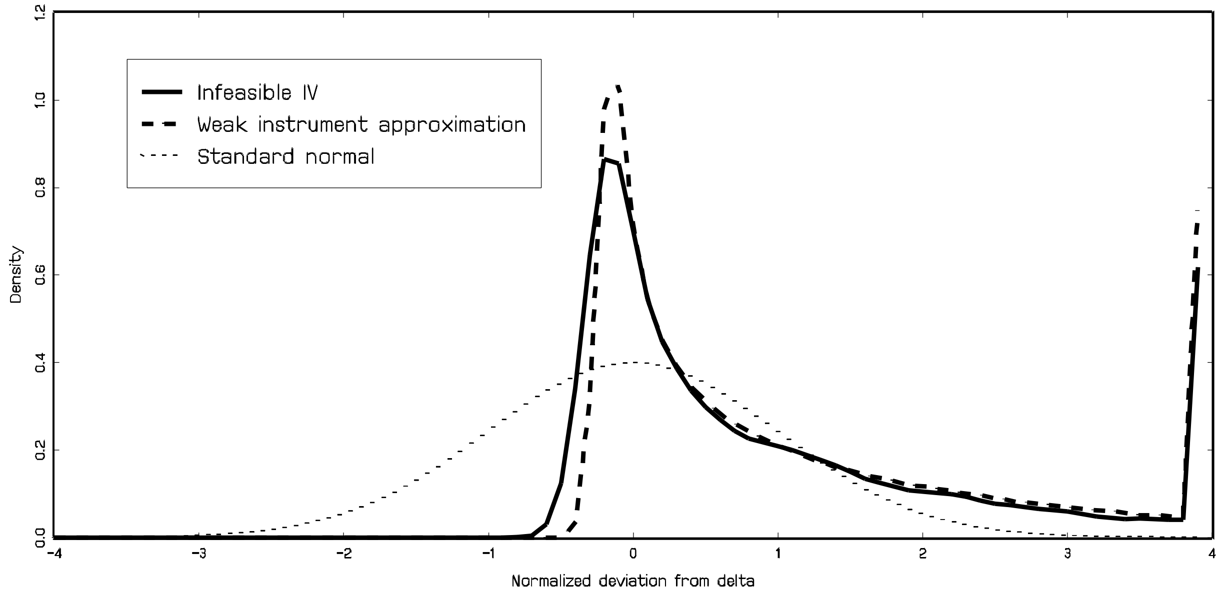
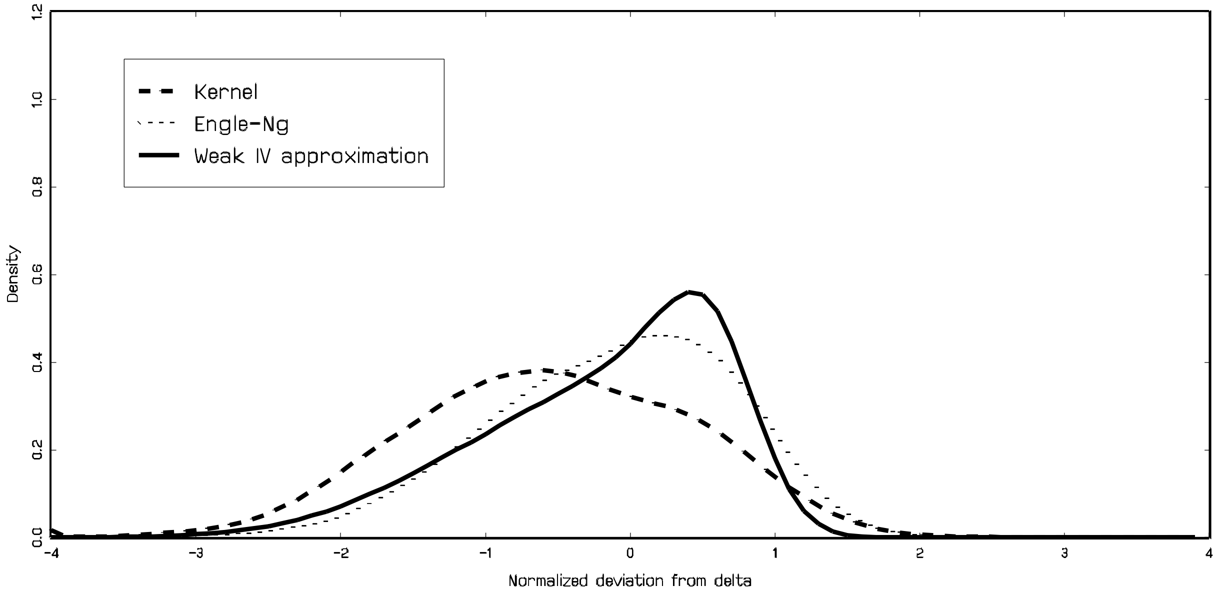


Fig 6. Distribution of IV estimator

Experiment 1 - Estimates from S&P 500 data



Experiment 2 - Estimates from yen-dollar data

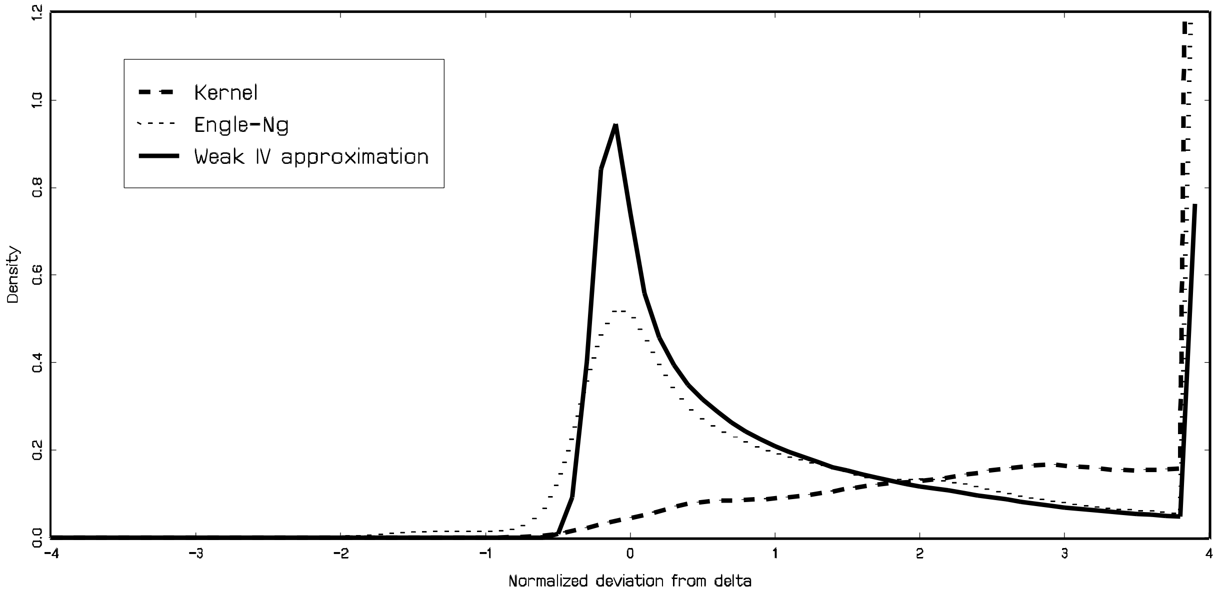
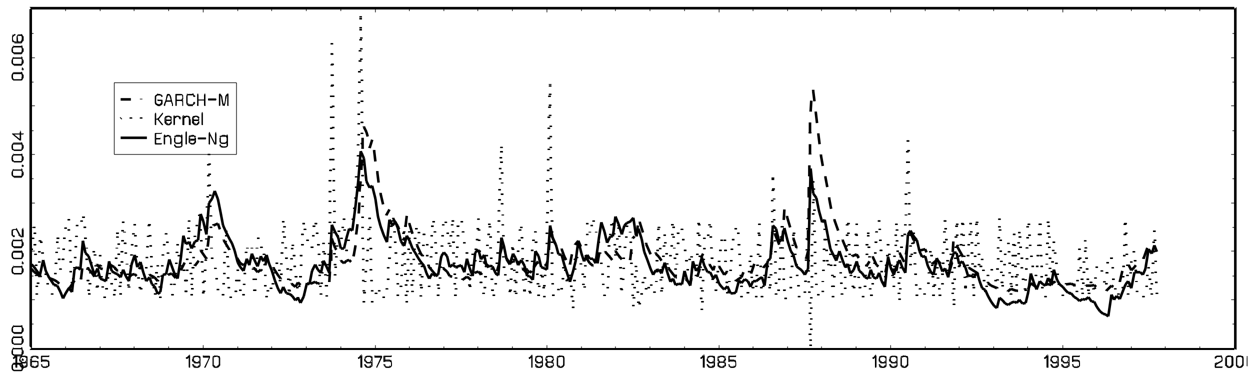
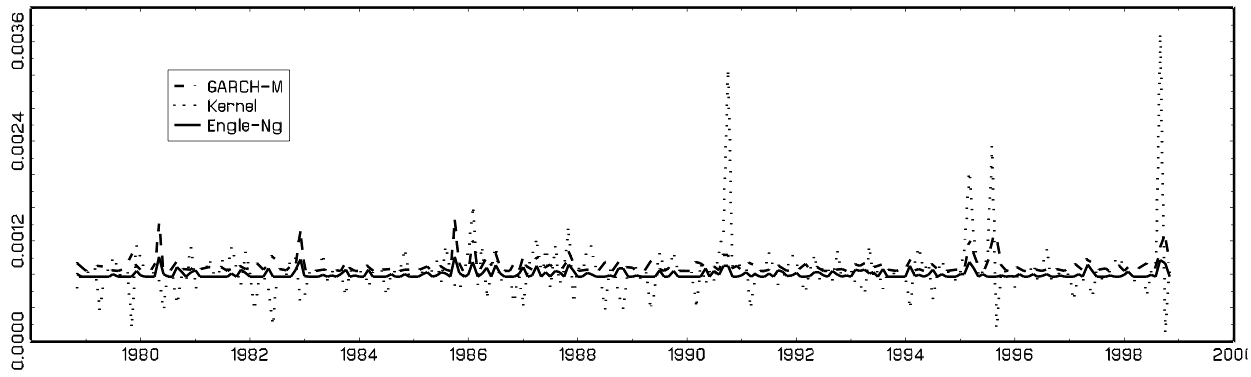


Figure 7. Time plots of volatility estimates

Returns on S&P 500 index
1965-1997



Yen-dollar spot returns
1978-1998



Excess holding yield
1959:1-2000:2