Combined Estimator of Time Series
Conditional Heteroskedasticity∗

Santosh Mishra† and Aman Ullah‡
Department of Economics, University of California, Riverside
August 2003

Abstract

We propose a new combined estimator, called semiparametric estimator, which incorporates the parametric and nonparametric estimators of the conditional variance in a multiplicative way. We derive bias, variance, and asymptotic normality of the combined estimator. Semiparametric estimators are found to be superior to parametric and nonparametric estimators, both in simulation and empirical analysis (S&P500 index return). We find that semiparametric models capture residual asymmetry in the conditional variance ignored by the corresponding symmetric and asymmetric parametric models. Semiparametric model is found to be superior to other models in a forecast evaluation based on VaR related loss function.

Key Words: Semiparametric Models, Nonparametric Estimator, and Conditional Variance

JEL Classifications: C3, C5, G0.

*The authors are grateful to John Galbraith, Gloria González-Rivera, Qi Li, Essie Maasoumi, Nour Meddahi, and Victoria Zinde-Walsh for their discussion and comments on the subject matter of this paper. They are also thankful for the comments by the participants of the seminars at Southern Methodist University, Texas A&M University, McGill University and Université de Montréal. Second author acknowledges the support from the academic senate, UCR.

†Department of Economics, University of California, Riverside, CA 92521-0427, U.S.A., e-mail: santoshucr@yahoo.com

‡Corresponding Author. Department of Economics, University of California, Riverside, CA 92521-0427, U.S.A., Tel: (909) 827-1591, Fax: (909) 787-5685, e-mail: aman.ullah@ucr.edu
1 Introduction

Volatility is a very important issue in many branches of economics including macroeconomics and finance. The main approach to volatility has been to model for the conditional variance ($\sigma^2_t$ at time $t$) in various ways. One method of the modelling assumes the unobservable $\sigma^2_t$ to be deterministic and mixture of past square returns. This method (called the GARCH family of models) was introduced in Engle (1982) and accounted for clusters of activity and fat-tail behavior of the data. It soon spawned a plethora of complicated models to capture empirically stylized facts of different fields. Surveys as in Bollerslev, Chou, and Kroner (1992), Bera and Higgins (1993), Bollerslev, Engle, and Nelson (1994) go into the details of related issues. Another method (called the Stochastic Volatility (SV) models) treats $\sigma^2_t$ as a latent variable and is expressed as a mixture of predictable and noise component. This class of models was more tractable for application in continuous time finance literature. Ghysels, Harvey, and Renault (1996) provides a good survey in this field.

Both these methods provide the parametric models of conditional variance $\sigma^2_t$ capturing a particular type of non-linearity in the data. It is well known that if the parametric model is correctly specified then it gives a consistent estimation of $\sigma^2_t$. In general, if the parametric model is incorrect, then its estimator may not be a consistent estimator of $\sigma^2_t$. However, one can still consistently estimate the unknown conditional heteroskedastic model of $\sigma^2_t$ by nonparametric estimation techniques. This approach has been explored recently in Fan and Yao (1998) and Ziegelman (2002). For a comprehensive exposition on nonparametric approach to modeling we refer to Hardle (1992) and Pagan and Ullah (1999). In general, the misspecified parametric fit will be poor (high bias) but it will be smooth (low variance). On the other hand
the data based nonparametric technique may trace the irregular pattern in
the data (less bias) but may be more variable (high variance).

In this paper we propose a new multiplicative combined estimator of
\( \sigma_t^2 \) which we refer to as a semiparametric estimator. In this case we first
model for the parametric part of the conditional variance and then the con-
ditional variance of the standardized residual factor (correction factor) is
modeled nonparametrically to capture the residual nonlinearity. Thus the
combined heteroskedasticity model is a multiplicative combination of a para-
metric model and the nonparametric model for correction factor. The global
parametric estimate of \( \sigma_t^2 \) can be obtained by using any parametric model
and the estimate of the conditional variance of the correction function can
be obtained by nonparametric methods. Here we use the kernel based lo-
cal nonparametric estimation. The idea behind the combined estimation is
that if the parametric estimator of \( \sigma_t^2 \) captures some information about the
true shape of \( \sigma_t^2 \), the correction factor will be less variable than \( \sigma_t^2 \) itself and
therefore easier to estimate nonparametrically. Another issue of interest is the
asymmetric impact of innovation on conditional variance which is frequently
encountered in the financial markets. It is well known that symmetric mod-
eels like GARCH (1,1) and ARCH (1) cannot capture the asymmetric part of
the conditional variance. This, we believe, can be traced by the combined
models since the correction factor captures the nonlinearity in the residuals.

We note that the combined estimation of the density function was in-
troduced by Olkin and Speigelman (1987). Glad (1998) and Fan and Ullah
(1999) developed the multiplicative and additive combined estimators, re-
spectively, in the context of regression (conditional mean) functions, also see
Gozalo and Linton (2000) for the semiparametric estimations of regression
functions. Engle and Gonzalez-Rivera (1991) provided a semiparametric like-
likelihood estimation of the ARCH model with a nonparametric estimation of the innovation density.

We calculate the bias, variance for the combined estimator and also indicate the asymptotic normality of the same. We also verify the relative performance of our combined (semiparametric) model with respect to pure local nonparametric exponential model (Ziegelmann (2002)) and global parametric exponential as well as simple GARCH models. We consider 11 data generating processes (DGP) and fit seven models to it. We compare them by the Mean Square Error (MSE) criteria and investigate the news impact curve to see whether they capture the pattern of true volatility. We get several interesting findings from this simulation. They are listed as follows.

1. The semiparametric models perform better than their parametric counterpart in case of misspecification. For those DGPs that are related to the GARCH family of models but with asymmetric conditional variance structure the semiparametric models corresponding to GARCH and ARCH (called $SP_2$ and $SP_3$ respectively) perform the best. In case of the exponential DGPs with data dependent time varying structure the semiparametric model (called $SP_1$) with parametric exponential part prevail. (2) Semiparametric models also capture the asymmetric conditional variance structure. We have included DGPs that have opposite asymmetric structure of conditional variance. The threshold GARCH models of Glosten, Jagannathan, and Runkle (GJR) (1993) have higher conditional variance for negative shocks (compared to positive shock) and global exponential models have the opposite conditional variance structure. We find that all the semiparametric models capture the asymmetry structure well. (3) The effect of residual nonlinearity modeling is more pronounced in $SP_2$ and $SP_3$ models. While the parametric part is symmetric for these models the asymmetry is captured purely by the
residual nonparametric modeling.

In the empirical section we fit a host of standard models along with the semiparametric models to S&P500 daily returns data. We find that both the semiparametric and nonparametric parameter estimates vary considerably with respect to past innovation, thus justifying against the global parametric modeling of conditional variance. Further the estimates for the pure nonparametric modeling and the semiparametric modeling are quite different in the exponential models. This attests to the fact that there is residual nonlinearity to be captured by the semiparametric model. We also see the asymmetric news impact curve for all the semiparametric models.

Time series literature has considered many asymmetric parametric specification for conditional variance in the past. It is useful to check whether the semiparametric model considered here can capture the residual asymmetry even when the parametric model is an asymmetric model. When the parametric model itself is asymmetric, we find that a corresponding semiparametric model captures residual asymmetry.

Out of sample evaluation is an important aspect of any modeling exercise. We consider a Value at Risk (VaR) based loss function to evaluate the out of sample performance of our semiparametric models. We find that both in terms of loss and reality check p-values our semiparametric model performs well.

The structure of the paper is as follows. Section 2 elaborates on the semiparametric estimator and provides the asymptotic property of the same. Section 3 provides simulations to compare the models by MSE criteria. Section 4 is related to modelling and forecast evaluation relating to S&P500 data by the semiparametric model. Finally section 5 provides the conclusion.
2 Semiparametric Model

Let \( \{y_t, x_{1t}, x_{2t}\}_{t=1}^{T} \) be a multi-dimensional stochastic process having the same marginal distribution (there may be overlap in the \( x_{1t} \) and \( x_{2t} \)) and both \( x_{1t}, x_{2t} \) are previsible processes. In this paper we assume \( x_{2t} \equiv y_{t-1} \). We write our model as follows

\[
y_t = x_{1t}' \phi + \varepsilon_t
\]

where \( E(\varepsilon_t|F_{t-1}) = 0 \) and \( E(\varepsilon_t^2|F_{t-1}) = \sigma_t^2 \) given the information set \( F_{t-1} \) (\( \sigma \)-field) at time \( t-1 \), \( x_{1t} \) is a \( k \times 1 \) vector of observations, and \( \phi \) is a \( (k \times 1) \) vector of parameters. The parametric estimation of \( \sigma_t^2 \) is often carried out by specifying the ARCH and GARCH family of models and the estimation method is usually maximum likelihood, see Engle (1982) and Bollerslav, Engle and Nelson (1994) for details. For asymptotic theory on the same see Lumsdaine (1996) and Comte and Lieberman (2001). There are some existing techniques to model for the conditional variance in the nonparametric framework. These are obtained by a local minimization exercise.

\[
(\hat{\alpha}, \hat{\beta}) = \min_{\alpha, \beta} \sum_{t=1}^{T} [\hat{\varepsilon}_t^2 - \Psi(\alpha + \beta(x_{2t} - x_2))]^2 K_{h_1}(x_{3t} - x_3)
\]

where \( \hat{\varepsilon}_t = (y_t - x_{1t}'\hat{\phi}) \), \( x_{3t} \) is the state variable for the kernel, \( h_1 \) is the bandwidth responsible for the extent of smoothing in the estimator. \( \hat{\phi} \) can be estimated by simple least square estimation. \( x_{3t} \) can be identical to \( x_{2t} \) but it can also be other exogenous variables that affect conditional variance. \( K_{h_1}(\xi) \equiv \frac{1}{h_1} K(\frac{\xi}{h_1}) \) is the kernel function and \( \Psi \) determines the specification of the model. If \( \Psi(x) \equiv x \) we get the local linear estimator by Fan and Yao (1998) and if \( \Psi(x) \equiv e^x \) we get the local exponential model by Ziegelmann (2002). The authors also provide the results on consistency and asymptotic normality of their nonparametric local estimators.
Here we propose a semiparametric estimator for the conditional variance based on a multiplicative combined model. The point of divergence from the existing works come in modelling for the residual nonlinearity in the variance equation. The total conditional variance is represented by the multiplication of the two parts; parametric and nonparametric. The nonparametric modeling is done on the standardized residuals obtained from the data and the estimated parametric model of variance. Thus the variance modelling is given as

\[
\varepsilon_t = \sigma_{spt} \nu_t = \sigma_{pt} \sigma_{npt} \nu_t
\]  

(3)

where \( \nu_t \) is i.i.d standard normal distribution and \( \sigma_{pt}^2 \) and \( \sigma_{npt}^2 \) are parametric and nonparametric (both are previsible processes) parts of the conditional variance. \( \sigma_{spt}^2 \) is the consolidated semiparametric conditional variance and it is given by

\[
\sigma_{spt}^2 = E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma_{pt}^2 E\{ (\frac{\varepsilon_t}{\sigma_{pt}})^2 | \mathcal{F}_{t-1} \} = \sigma_{pt}^2 E\{ z_t^2 | \mathcal{F}_{t-1} \} = \sigma_{pt}^2 \sigma_{npt}^2
\]  

(4)

where \( z_t \) is the standardized residual. Essentially \( \sigma_{pt}^2 \) in equation (3) is any parametric model of the conditional variance which may not have the correct specification and thus only roughly approximates \( \sigma_{spt}^2 \). The nonparametric component \( \sigma_{npt}^2 = E\{ z_t^2 | \mathcal{F}_{t-1} \} \) is a correction function and it can be nonparametrically estimated by the local exponential function as

\[
(\hat{\alpha}, \hat{\beta}) = \min_{\alpha, \beta} \sum_{t=2}^{T} [\hat{z}_t^2 - \Psi \{ \alpha + \beta (x_{t-1} - \hat{z}_t) \}]^2 K_{h_2}(x_3 - x_3)
\]  

(5)

where \( h_2 \) is the new appropriate bandwidth and \( \hat{z}_t = \frac{\hat{\varepsilon}_t}{\hat{\sigma}_{pt}} \) is an estimated standardized residual. Thus the estimation process is carried out in two steps.
Step 1: Model for the mean and variance parametrically and obtain \( \hat{\varepsilon}_t \) and the parametric conditional variances by \( \hat{\sigma}_{pt} \). In this paper we have considered three type of parametrization for the conditional variance

\[
GARCH(1,1) : \sigma^2_{pt} = \omega + \gamma \sigma^2_{t-1} + \delta \varepsilon^2_{t-1}.
\]

\[
ARCH(1) : \sigma^2_{pt} = \omega + \delta \varepsilon^2_{t-1}
\]

\[
GE : \sigma^2_{pt} = \exp(a + b \varepsilon_{t-1})
\]

where GE stands for global exponential. The standardized residual is then estimated as \( \hat{z}_t \).

The last model is a variation of a class of GARCH model. In this model \( \ln(\sigma^2_t) \) is modeled linearly without any parameter constraints in the variance equation. Here we directly model for the \( \sigma^2_t \). The EGARCH model by Nelson (1993) is also in the similar vein but there determinants of volatility are past i.i.d white noise rather than the past innovations.

Step 2: Once the standardized residuals \( \hat{z}_t \) are obtained we model the square of residuals by a local exponential model as in equation (4). We get the estimate of the semiparametric conditional variance by multiplying the appropriate parametric and nonparametric conditional variances obtained earlier. We get

\[
\hat{\sigma}^2_{spt} = \hat{\sigma}^2_{pt} \hat{\sigma}^2_{npt}
\]  

Now we present the asymptotic normality result of \( \hat{\sigma}^2_{spt} \). For this we first introduce \( \lambda^2(z) = E[(\nu_t^2 - 1)^2|z_{t-1} = z] \), \( \hat{\sigma}^2_{npt}(z) \) and \( \hat{\sigma}^2_{npt}(z) \) as first and second order Partial derivatives of \( \sigma^2_{npt}(z) \) with respect to \( z \) respectively, and \( \sigma^2_K = \int \psi^2 K(\psi)d\psi \). Then

\[
(Th)^{0.5}\{\hat{\sigma}^2_{spt}(y) - \sigma^2_{spt}(y) - \partial_t\} - \mathbf{N}(0, f(z)^{-1}\sigma^4_{spt}(y)\lambda^2(z) \int K^2(\psi)d\psi)
\]  

7
where \( \vartheta_t = 0.5\sigma_{pt}^2(y)h^2(\hat{\sigma}_{npt}^2(z) - \beta^2 e^\alpha)\sigma_K^2 \).

The proof for this result is given in Appendix B.

We see here that the bias is of order \( O_p(h^2) \) and as expected it depends on the parametric estimates of the conditional variance and also the estimated parameters of the second stage \( (\hat{\sigma}_{npt}^2(z)) \). If we assume \( (Th)^{0.5}h^2 \) to be 0 as \( T \to \infty \) then the bias vanishes asymptotically. We also note that for the local linear estimator \( \vartheta_t = 0.5\sigma_{pt}^2(y)h^2\sigma_{npt}^2(z)\sigma_K^2 \), which is larger than the bias in the exponential case as long as \( \sigma_{npt}^2(z) > 0.5\beta^2 e^\alpha \).

3 Simulation

3.1 Simulation Design

We provide a set of simulations to check the relative performance of the semiparametric estimators with respect to parametric and non parametric estimators. We consider 11 data generating processes and we model the conditional variance by ARCH (1), GARCH (1,1), Parametric Exponential (GE), Nonparametric Exponential (NP) and semiparametric Exponential. There are three version of semiparametric Exponential estimator, \( SP_1 \), \( SP_2 \) and \( SP_3 \) which corresponds to parametric exponential, GARCH (1,1) and ARCH (1) parametric models respectively. We consider \( y_t = \varepsilon_t \), where \( E(\varepsilon_t|\mathcal{F}_{t-1}) = 0 \) and \( E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_t^2 \).

The dynamics considered for \( \sigma_t^2 \) are

- **DGP 1**: The ARCH\((p)\) model of Engle (1982):

\[
\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i\varepsilon_{t-i}^2
\]

where we consider \( p = 1, \alpha = 0.5 \) and \( \omega = 0.2 \).
• DGP 2: Nonlinear ARCH(1) model \((NARCH)\)

\[
\sigma_t^2 = \omega + \{\alpha_1 + \alpha_2(1(\varepsilon_{t-1} \geq 0))\varepsilon_{t-1}^2
\]

where \(\omega = 0.2, \alpha_1 = 0.2, \alpha_2 = 0.3\) and \(1(\cdot)\) is an indicator function.

• DGP 3: The GARCH(1, 1) model of Bollerslev (1986):

\[
\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \varepsilon_{t-1}^2
\]

where \(\omega = 0.3, \alpha = 0.1, \) and \(\beta = 0.7\)

• DGP 4-6: Three threshold GARCH models with increasing nonlinearity by GJR that allows for possible asymmetric effects of positive and negative innovations \((GJR1, GJR2, \) and \(GJR3)\)

\[
\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \varepsilon_{t-1}^2 + \gamma \varepsilon_{t-1}^2 1(\varepsilon_{t-1} \leq 0)
\]

where \(1(\cdot)\) is an indicator function, \(\omega = 0.3, \alpha = 0.08, \beta = 0.6,\) and three values of \(\gamma\) are given by \(\gamma = 0.1, 0.2, \) and \(0.3\) for DGP 4, 5, and 6 respectively.

• DGP 7: Global Exponential Model \((GE)\)

\[
\sigma_t^2 = \exp(a + b\varepsilon_{t-1})
\]

where \(a = 0.1\) and \(b = 0.1\)

• DGP 8: Local (data dependent) Exponential model Type 1 \((LE1)\)

\[
\sigma_t^2 = \exp(a + b_t\varepsilon_{t-1})
\]

where \(a = 0.1, b_t = 0.3\frac{\exp(\varepsilon_{t-1})}{(1+\exp(\varepsilon_{t-1}))},\) and the nonlinear parameter is the coefficient of \(\varepsilon_{t-1}\).
DGP 9: Local (data dependent) Exponential model Type 2 (LE2)

\[ \sigma_t^2 = \exp(a_t + b\varepsilon_{t-1}) \]

where \( a_t = 1.5\frac{\exp(\varepsilon_{t-1})}{1+\exp(\varepsilon_{t-1})} \) and \( b = 0.1 \), and here the nonlinear parameter is the constant.

DGP 10: Mixed Local Exponential Model Type 1 (MLE1)

\[ \sigma_t^2 = \exp[a + (b + c_t)\varepsilon_{t-1}] \]

where \( a = 0.1 \), \( b = 0.1 \), and \( c_t = 0.3\frac{\exp(\varepsilon_{t-1})}{1+\exp(\varepsilon_{t-1})} \)

DGP 11: Mixed Local Exponential Model Type 2 (MLE2)

\[ \sigma_t^2 = \exp[(a_t) + b\varepsilon_{t-1}] \]

where \( a_t = a + c_t \), and \( a = 0.1 \), \( b = 0.1 \), and \( c_t = 1.5\frac{\exp(\varepsilon_{t-1})}{1+\exp(\varepsilon_{t-1})} \)

We draw replications of \( \nu_t \) from a standard normal distribution. So we generate \( \nu_{it} (i = 1, ..., M) \) and use the same to generate \( \varepsilon_{it} \) through the specified conditional variance DGPs. The simulation was carried out for sample size \( T = 100, 200 \) and \( 300 \). The number of replications is \( M = 200 \). For brevity we only present the tables and graphs for \( T = 200 \). The results are qualitatively similar for other sample sizes too. Let

\[
B^j_t = \frac{1}{M} \sum_{i=1}^{M} [(\sigma^j_t)^2 - (\hat{\sigma}^j_{it})^2], \quad t = 1, ..., T \text{ and } j = 1, ..., 11
\]

\[
S^j_t = \frac{1}{M} \sum_{i=1}^{M} [(\hat{\sigma}^j_{it})^2 - \frac{1}{M} \sum_{i=1}^{M} (\hat{\sigma}^j_{it})^2]^2, \quad t = 1, ..., T \text{ and } j = 1, ..., 11
\]

\[
B^j = [B^j_1, ..., B^j_T]' \text{ and } S^j = [S^j_1, ..., S^j_T]'
\]
where \(i\) and \(j\) stands for the index of replication and model respectively. \((\sigma^j_t)^2\) is the true conditional variance for model \(j\) at time \(t\), \((\hat{\sigma}^j_t)^2\) is the estimated conditional variance for \(i^{th}\) replication and \(j^{th}\) model at time \(t\). \(B\) and \(S\) stand for bias and variance. \(M\) is the total number of replication (in this case \(M = 200\)). Let \(MSE^j_t = (B^j_t)^2 + S^j_t\) be the mean square error of estimates of \(\sigma^2_t\). Thus we calculate the average MSE for model \(j\) by the following equation.

\[
MSE^j = \frac{1}{T} \sum_{t=1}^{T} MSE^j_t
\]

We compare \(MSE^j\) for all the models for 11 DGPs considered in this simulation. The lowest value suggests the best fit of the model.

### 3.2 Simulation Findings

Table 1 provides the comparisons of MSEs. The MSE values for each DGP is presented across the rows (for example the first row corresponds to DGP ARCH(1) ). The model corresponding to the minimum MSE is the best model. Following points of the simulation results are worth noting.

(i) When the true DGP is same as the model fitted then the fitted model, expectedly, performs the best. For example ARCH (1) and GARCH (1,1) are the best models when the DGP happens to be ARCH (1) and GARCH (1,1).

(ii) We find that compared to misspecified fitted parametric model the semiparametric models always perform better. For example when the DGP is GJR1 the MSE for \(SP\) is 0.0559 and it is 0.0720 for GARCH (1,1) model.

(iii) The GJR class of DGPs (GJR1, GJR2, and GJR3) are more related to the GARCH family of models. So naturally \(SP\) 2 and \(SP\) 3 perform the
best for those models. For example in case of DGPs $GJR1$ and $GJR3$ $SP3$ performs the best and $SP2$ is the best model for DGP $GJR2$.

(iv) When the DGPs belong to the exponential family (say $MLE1$ and $MLE2$) the semiparametric model with exponential parametric part is found to be superior. $SP1$ has the minimum MSE for these models.

(v) We also find that $SP1$ is superior model to $NPE$ except for the case when the DGPs are driven by parameters that are local ($LE1$ and $LE2$). This shows that the semiparametric models capture nonlinearities missed by pure nonparametric modeling.

Another issue in question is the shape of the news impact curve for different volatility models. This is important because asymmetric volatility pattern is often observed in financial markets. So our second purpose is to investigate whether semiparametric models can capture the asymmetric volatility dynamics. Our DGPs have both type of asymmetric dynamics. In GJR class of DGPs the conditional variance is more sensitive to negative rather than positive innovation (this DGP mimics the leverage effect seen in stock market). The impact of innovation on conditional variance is exactly opposite in $GE$ model. We look at the average of conditional variance at a given time $t$ over replications.

$$AV_{jt} = \frac{1}{M} \sum_{i=1}^{M} (\hat{\sigma}_{it}^2)^2, i = 1, ..., M, t = 1, ..., T$$

where $i$ and $j$ stand for the index of the simulation and model respectively, $(\hat{\sigma}_{it}^2)^2$ is the estimated conditional variance at time $t$ for $i^{th}$ simulation from $j^{th}$ model. Let us define $AV^j = [AV_{1}^j, ..., AV_T^j]$. We plot the true conditional variance (with increasing innovation value) from the DGPs and the
conditional variances estimated from seven separate models considered in this paper. For brevity we only present selected graphs to convey the main summaries of the result. It is observed that when the model is properly specified (say the fitting of GARCH model when the DGP is GARCH) the parametric model performs well, but in the presence of certain type of misspecification the behavior of parametric model deteriorates quickly. For example the Global exponential model performs well when the DGPs are GE, LE1, and MLE2 but it performs badly for all GJR family of models and also for GARCH and NARCH. This is true of other parametric specification. From that perspective all the semiparametric models are robust to many types of misspecification. graph 1 attests to this point. It is clear that all the semiparametric models capture the structure of conditional variance adequately. The wiggly fit of the SP2 can be attributed to the relatively more noisy GARCH component in the estimation. Another important issue is that it captures the asymmetric effect in case of the GJR, NARCH, and the DGPs in the exponential family. We clearly see this effect in case of SP2 and SP3 where the parametric component is symmetric and the capturing of the asymmetry is left to the nonparametric modeling in the second stage. As graph 2 shows the semiparametric models capture this asymmetry very well.

4 Empirical Data Analysis

In this section we fit the semiparametric model to the real data and do some diagnostic checking to see whether the model is able to capture the changing conditional variance and leverage effect. We consider the S&P500 daily returns from January 2nd 1995 through August 31 2001, a total of 1430 observations (data collected from Datastream.). As is well known the return
series has fat tails (kurtosis=8.0). To establish notation, suppose the return series \( \{y_t\}_{t=1}^T \) of a financial asset follows the stochastic process \( y_t = c + \varepsilon_t \), where \( E(\varepsilon_t|\mathcal{F}_{t-1}) = 0 \) and \( E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_t^2 \) given the information set \( \mathcal{F}_{t-1} \) (\( \sigma \)-field) at time \( t-1 \). Let \( \nu_t \equiv \varepsilon_t / \sigma_t \) have the conditional distribution \( \Phi \) with zero conditional mean and unit conditional variance, i.e., \( \nu_t \sim \Phi(0, 1) \). We fit the three type of semiparametric models and other volatility models to the data. The state variable for the kernel is given as the past value of returns \( (y_{t-1}) \). The Box-Pierce statistics indicate that the standardized residuals obtained from the semiparametric conditional variance are uncorrelated for long lags. We also tried with higher level of lags in the specification of conditional variance. The improvement in the likelihood function was marginal. So we settled with the parametric model with one lag of the return series. Graph 3 presents news impact curves associated with the real data. We see that the news impact curve for semiparametric class of models is quite different from that of nonparametric model. For the GARCH family of models we plot the parametric and the corresponding semiparametric model (ARCH vs SP3 and GARCH vs SP2). To clarify the role of asymmetry we have also plotted the corresponding nonparametric estimates of the squared standardized residuals from the parametric models (represented by legend nonlinearity). First two plots of graph 4 present these results. As can be seen from the graphs the fitted value of squared standardized residuals have a asymmetric relation to the innovation. This means that multiplicative factor to the corresponding parametric model is higher for negative shocks than that for the positive shocks, which renders the final fitted model asymmetric. We see that models render substantially different values of conditional variance for extreme observations. It is the tail behavior that is of interest in the financial market. We address this issue in the out of sample comparison of parametric,
nonparametric, and semiparametric volatility models in VaR framework.

The empirical literature is aware of many asymmetric parametric models and it can be said that the semiparametric model considered here only picks up the asymmetry which can also be captured by standard asymmetric models. So we wanted to check whether our semiparametric model can capture the residual asymmetric effect even after our parametric model captures some amount of asymmetry. The universe of asymmetric parametric model is quite large and here in this paper we only consider GJR model. So we fit a parametric GJR model and nonparametrically model squared standardized residuals. The semiparametric model for this case is called SP4. As seen in the last plot of graph 4, even after fitting for the GJR type model we still find unaccounted for residual nonlinearity.

We know that volatility is affected by many other exogenous variables. This is usually accounted for by adding these variables in the GARCH equation. But this suffers from the standard linearity constraint. Although not attempted here, our semiparametric model can handle this issue more flexibly by using these variables as state variables within the kernel function, thus incorporating the nonlinear impact of exogenous variables on conditional variance.

4.1 Out of Sample Evaluation

In the last subsection we observed that our model is suitable to capture the standard features observed in financial data (thick tail and asymmetry). It is often found that while nonparametric models perform well in the insample analysis the same cannot said about the out of sample performance. That is why we carry out a comparison of the volatility models considered in this paper. To take care of the data snooping we apply White’s reality check
framework. Appendix A provides a small note on White’s reality check. The models considered here are ARCH, GARCH, GJR, GE, SP1, SP2, SP3 and SP4. In the following subsection we briefly talk about the loss function considered in this paper.

4.1.1 Loss function and result

The conditional Value-at-Risk, denoted as $VaR_{t+1}(\alpha)$, can be defined as the conditional quantile

$$\Pr(\varepsilon_{t+1} \leq VaR_{t+1}(\alpha)|\mathcal{F}_t) = \alpha. \quad (8)$$

If the density of $\varepsilon$ belongs to the location-scale family (e.g., Lehmann 1983, p. 20), it may be estimated from

$$VaR_{t+1}(\alpha) = \Phi^{-1}_{t+1}(\alpha)\sigma_{t+1}(\hat{\theta}_t), \quad (9)$$

where $\Phi_{t+1}(\cdot)$ is the forecast cumulative distribution of the standardized return and $\sigma^2_{t+1}(\theta) = E(\varepsilon^2_{t+1}|\mathcal{F}_t)$ the conditional variance forecast based on the volatility models considered above, and $\hat{\theta}_t$ is the parameter vector estimated by using the information up to time $t$. We assume conditional normality of the standardized return. We consider the quantile $\alpha = 0.01$ and thus $\Phi^{-1}_{t+1}(0.05) = -2.12$ for all $t$.

Our statistical loss function $V$ is the loss function used in the quantile estimation (see e.g., Koenker and Bassett 1978), that is

$$V \equiv P^{-1} \sum_{t=R}^{T} |\varepsilon_{t+1} - VaR_{t+1}(\alpha)| \times [(1 - \alpha)\mathbf{1}(\varepsilon_{t+1} < VaR_{t+1}(\alpha)) + \alpha\mathbf{1}(\varepsilon_{t+1} \geq VaR_{t+1}(\alpha))], \quad (10)$$

where $P$ and $R$ are the out of sample and in sample length respectively. For this paper we consider $P = 530$ and $R = 900$. This is an asymmetric loss.
function that penalizes more heavily the observations for which \( \varepsilon - VaR(\alpha) < 0 \). Smaller \( V \) indicates a better goodness of fit. In this case we consider two state variables used in the kernel, past return and absolute value of past return. Table 2 provides result for the out of sample comparison of volatility models. The numbers in the table provide the loss values associated with different models and the reality check p-values (for the given model as the benchmark model). The parametric model has the same loss values for two different state variables. From the loss values we find that the semiparametric GARCH and semiparametric GJR model (SP2 and SP4 respectively) have smaller loss values compared to their parametric counterparts. We also see that for 1% significance level the SP2 model has uniformly small losses for both of the state variables. The numbers in the bracket (below the loss values) give the reality check p-values for the given model as the benchmark model. So when SP2 is the benchmark model (with past return as the state variable) the reality check p-value is given as 0.695. From the p-values we find that the best model is SP2 with absolute value of past return as the state variable. SP4 model also performs well. But each of the remaining parametric, nonparametric, and semiparametric models (SP1 and SP3) are individually dominated in the sense that reality check p-values are small and loss values tend to be in the high end.

| Table 2 Somewhere Here |

5 Conclusion

This paper proposes a new semiparametric estimator for time varying conditional variance. The purpose is to model for nonlinearity in the conditional variance dynamics. This is accomplished by combining parametric and nonparametric estimators in a multiplicative way. The estimation is a two step
procedure. (i) In the first stage we model for the conditional variance parametrically to retrieve the standardized residuals which contains nonlinearity. (ii) In the second stage we model for the standardized residual obtained in the first stage by local exponential estimation method. We consider three type of parametric specification in the first stage, namely GE, GARCH (1,1), and ARCH(1). The other purpose of the paper is to investigate whether the semiparametric estimator captures the asymmetric impact of innovation on the conditional variance. This is important because leverage effect is often found in financial market and an useful volatility model should be able to capture this. We carry out a simulation exercise that includes DGPs of all stripes (symmetric, asymmetric, and DGPs where the parameters are data dependent). We compare the relative performance of all three semiparametric models $SP_1$, $SP_2$ and $SP_3$ to parametric exponential, nonparametric exponential, $GARCH (1,1)$ and $ARCH (1)$ models. We find that (i) When the true DGP is same as the model fitted then the fitted model, expectedly, performs the best, (ii) Compared to misspecified fitted parametric model the semiparametric models always perform better, (iii) The GJR class of DGPs ($GJR_1$, $GJR_2$, and $GJR_3$) are more related to the GARCH family of models. So naturally $SP_2$ and $SP_3$ perform the best for those models, (iv) When the DGPs belong to the exponential family (say $MLE_1$ and $MLE_2$) the semiparametric model with exponential parametric part is found to be superior, and (v) $SP_1$ is superior model to $NPE$ except for the case when the DGPs are driven by parameters that are purely nonparametric($LE_1$ and $LE_2$). We plot the news impact curves to check whether semiparametric model captures asymmetry or not, especially in case of $SP_2$ and $SP_3$ where the parametric parts are symmetric. We find that both of these semiparametric model captures residual asymmetry in the conditional variance. We
fit the above models to S&P500 returns. The semiparametric models capture the leverage effect in the data. It is in the tails that the response of conditional variance to innovation differs substantially across models. To check whether the semiparametric model can capture residual nonlinearity when the parametric model itself is asymmetric, we consider a new semiparametric model for which the parametric part is the GJR model. We find that our model still captures residual asymmetry. Out of sample evaluation is an important aspect of any modeling exercise. We consider a VaR based loss function to evaluate the out of sample performance of our semiparametric models. We find that both in terms of loss and reality check p-values our model performs well.
References


Lumsdaine, R. L. (1996), “Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models”, *Econometrica* 64, 575-596.


Table 1: Mean Square Error (MSE) for Simulation

<table>
<thead>
<tr>
<th>MODEL</th>
<th>ARCH(1)</th>
<th>GARCH</th>
<th>NPE</th>
<th>PE</th>
<th>SP1</th>
<th>SP2</th>
<th>SP3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARCH(1)</td>
<td>0.0002</td>
<td>0.0003</td>
<td>0.0417</td>
<td>0.0139</td>
<td>0.0106</td>
<td>0.0005</td>
<td>0.0003</td>
</tr>
<tr>
<td>NARCH</td>
<td>0.0003</td>
<td>0.0038</td>
<td>0.0445</td>
<td>0.0139</td>
<td>0.0200</td>
<td>0.0036</td>
<td>0.0001</td>
</tr>
<tr>
<td>GARCH</td>
<td>0.0125</td>
<td>0.0062</td>
<td>0.0361</td>
<td>0.1796</td>
<td>0.0167</td>
<td>0.0110</td>
<td>0.0110</td>
</tr>
<tr>
<td>GJR1</td>
<td>0.0171</td>
<td>0.0720</td>
<td>0.0684</td>
<td>0.0282</td>
<td>0.0175</td>
<td>0.0559</td>
<td>0.0092</td>
</tr>
<tr>
<td>GJR2</td>
<td>0.0581</td>
<td>0.1705</td>
<td>0.1262</td>
<td>0.1018</td>
<td>0.0661</td>
<td>0.0100</td>
<td>0.0501</td>
</tr>
<tr>
<td>GJR3</td>
<td>0.0340</td>
<td>0.1125</td>
<td>0.0304</td>
<td>0.0385</td>
<td>0.0268</td>
<td>0.0248</td>
<td>0.0148</td>
</tr>
<tr>
<td>GE</td>
<td>0.0384</td>
<td>0.0389</td>
<td>0.0117</td>
<td>0.0057</td>
<td>0.0068</td>
<td>0.0087</td>
<td>0.0061</td>
</tr>
<tr>
<td>LE1</td>
<td>0.0347</td>
<td>0.0165</td>
<td>0.0089</td>
<td>0.0118</td>
<td>0.0101</td>
<td>0.0157</td>
<td>0.0139</td>
</tr>
<tr>
<td>LE2</td>
<td>0.2013</td>
<td>0.1891</td>
<td>0.1050</td>
<td>0.1517</td>
<td>0.1140</td>
<td>0.1730</td>
<td>0.1230</td>
</tr>
<tr>
<td>MLE1</td>
<td>0.0187</td>
<td>0.0320</td>
<td>0.0066</td>
<td>0.0030</td>
<td>0.0017</td>
<td>0.0250</td>
<td>0.0157</td>
</tr>
<tr>
<td>MLE2</td>
<td>0.1531</td>
<td>0.1121</td>
<td>0.1387</td>
<td>0.1312</td>
<td>0.0953</td>
<td>0.1091</td>
<td>0.1231</td>
</tr>
</tbody>
</table>

Notes: Table for MSEs from simulations for 11 DGPs and 7 models fitted to it. The sample size $T = 200$ and the number of simulations is $M = 200$. We present the mean square error (MSE) associated with the different DGPs. The comparison for a particular DGP is done across the row and the best model is the model with the lowest MSE value.
Table 2: Table for loss and Reality check p-values in forecast evaluation

<table>
<thead>
<tr>
<th>Model</th>
<th>ARCH</th>
<th>GARCH</th>
<th>GJR</th>
<th>GE</th>
<th>NP</th>
<th>SP1</th>
<th>SP2</th>
<th>SP3</th>
<th>SP4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\varepsilon_{t-1}$</td>
<td>0.0763 (0.008)</td>
<td>0.0663 (0.014)</td>
<td>0.0692 (0.010)</td>
<td>0.0857 (0.001)</td>
<td>0.0815 (0.002)</td>
<td>0.0789 (0.009)</td>
<td>0.0506 (0.695)</td>
<td>0.0815 (0.003)</td>
<td>0.0534 (0.634)</td>
</tr>
<tr>
<td>$</td>
<td>\varepsilon_{t-1}</td>
<td>$</td>
<td>0.0763 (0.008)</td>
<td>0.0663 (0.014)</td>
<td>0.0692 (0.010)</td>
<td>0.0857 (0.001)</td>
<td>0.0858 (0.001)</td>
<td>0.0858 (0.001)</td>
<td>0.0487 (0.949)</td>
</tr>
</tbody>
</table>

Note: This table presents the VaR loss values and reality check p-values (given within the brackets) for the given model being the benchmark model. The loss function is given as: $V = P^{-1} \sum_{t=R}^{T} |\varepsilon_{t+1} - VaR_{t+1}(\alpha)| \times (1 - \alpha) 1(\varepsilon_{t+1} < VaR_{t+1}(\alpha)) + \alpha 1(\varepsilon_{t+1} \geq VaR_{t+1}(\alpha))$. $R=900$ and $P=530$. The lower is the loss value the better is the model. The loss is estimated for $\alpha = 0.01$. The bootstrap reality check $p$-values are computed with 1000 bootstrap resamples and smoothing parameter $q = 0.25$. See Politis and Romano (1994) or White (2000) for the details. The $p$-values for $q = 0.75$ and 0.50 are similar and are not reported.
Appendix

A: Note on White’s reality check

Consider various volatility models and choose one as a benchmark. For each model, we are interested in the out-of-sample one-step ahead forecast. This forecast will be fed into an objective function, for instance, a utility function or a loss function. Our interest is to compare the utility (loss) of each model to that of the benchmark model. We formulate a null hypothesis where the model with the largest utility (smallest loss) is not any better than the benchmark model. If we reject the null hypothesis, there is at least a model that produces more utility (less loss) than the benchmark.

Formally, the testing proceeds as follows. Let $l$ be the number of competing volatility models ($k = 1, \ldots, l$) to compare with the benchmark volatility model (indexed as $k = 0$). For each volatility model $k$, one-step predictions are to be made for $P$ periods from $R$ through $T$, so that $T = R + P - 1$. As the sample size $T$ increases, $P$ and $R$ may increase. For a given volatility model $k$ and observations 1 to $R$, we estimate the parameters of the model $\hat{\theta}_R$ and compute the one-step volatility forecast $\sigma^2_{k,R+1}(\hat{\theta}_R)$. Next, using observations 2 to $R + 1$, we estimate the model to obtain $\hat{\theta}_{R+1}$ and calculate the one-step volatility forecast $\sigma^2_{k,R+2}(\hat{\theta}_{R+1})$. We keep rolling our sample one observation at a time until we reach $T$, to obtain $\hat{\theta}_T$ and the last one-step volatility forecast $\sigma^2_{k,T+1}(\hat{\theta}_T)$. Consider an objective function that depends on volatility, for instance, a loss function $L(Z, \sigma^2(\theta))$ where $Z$ typically will consist of dependent variables and predictor variables. The best forecasting model is the one that minimizes the expected loss. Our interest is to compare the model that produces the minimum loss to that of the benchmark model. We test a hypothesis about an $l \times 1$ vector of moments, $E(f^*)$, where $f^* \equiv f(Z, \theta^*)$ is an $l \times 1$ vector with the $k$th element $f^*_k = L(Z, \sigma^2_0(\theta^*)) - L(Z, \sigma^2_k(\theta^*))$, for
\( \theta^* = \lim \hat{\theta}_T \). A test for a hypothesis on \( E(f^*) \) may be based on the \( l \times 1 \) statistic

\[
\bar{f} \equiv P^{-1} \sum_{t=R}^{T} \hat{f}_{t+1},
\]

where \( \hat{f}_{t+1} \equiv f(Z_{t+1}, \hat{\theta}_t) \). West (1996) showed that

\[
\sqrt{P} (\bar{f} - E(f^*)) \rightarrow N(0, \Omega) \text{ in distribution}
\]

as \( P \rightarrow \infty \) when \( T \rightarrow \infty \), where \( \Omega \equiv \lim_{T \rightarrow \infty} \text{var}[\sqrt{P} (\bar{f} - E(f^*))] \) is an \( l \times l \) matrix.

When we compare a single model \( (l = 1) \) with a benchmark, we can use the above asymptotic distribution as in Diebold and Mariano (1995) and West (1996). However, in our empirical section, the p-values will not be computed from the asymptotic distribution but from the bootstrap reality check method à la White (2000) as we discuss below. However, when we compare multiple forecasting models \( (l > 1) \) against a given benchmark, the sequential use of Diebold and Mariano (1995) and West (1996) tests may result in a data-snooping bias because the test statistics are mutually dependent (i.e., \( \Omega \) is not diagonal).

An appropriate null hypothesis is “the best model is no better than a benchmark”, expressed formally as

\[
H_0 : \max_{1 \leq k \leq l} E(f^*_k) \leq 0.
\]

This is a multiple hypothesis, the intersection of the one-sided individual hypotheses \( E(f^*_k) \leq 0, \ k = 1, \ldots, l \). The alternative is that \( H_0 \) is false, that is, the best model is superior to the benchmark. If the null hypothesis is rejected, there must be at least a model for which \( E(f^*_k) \) is positive.
White’s (2000) test statistic for $H_0$ is formed as

$$
\bar{V} \equiv \max_{1 \leq k \leq l} \sqrt{P} \bar{f}_k,
$$

which converges in distribution to $\max_{1 \leq k \leq l} G_k$ under $H_0$, where the limit random vector $G = (G_1, \ldots, G_l)'$ is $N(0, \Omega)$. However, the null limiting distribution of $\max_{1 \leq k \leq l} G_k$ is unknown and it is not feasible to derive it, even asymptotically. White (2000) suggested to use the “stationary bootstrap” of Politis and Romano (1994) to obtain the null distribution of $\bar{V}$. Under appropriate conditions and under the null hypothesis, the distribution of $\sqrt{P}(\bar{f}^* - \bar{f})$ converges to that of $\sqrt{P}(\bar{f} - E(f^*))$, where $\bar{f}^*$ is obtained from the stationary bootstrap. We obtain the empirical quantiles of the statistic

$$
\bar{V}^* = \max_{1 \leq k \leq l} \sqrt{P}(\bar{f}_k^* - \bar{f}_k)
$$

and, accordingly, the p-value for testing the null hypothesis that the best model has no superior predictive ability relative to the benchmark (White, 2000, Corollary 2.4). This p-value is called the “Reality Check p-value” for data snooping.
B: Asymptotic Normality of $\hat{\sigma}_{\text{spt}}$

Let us consider the combined estimator

$$\hat{\sigma}_{\text{spt}}^2(y) = \hat{\sigma}_{\text{pt}}^2(y)\hat{\sigma}_{\text{npt}}^2(z)$$

where $\hat{\sigma}_{\text{pt}}^2(y_t) = g(y_t, \hat{\theta})$ is any parametric estimator, $\hat{\sigma}_{\text{npt}}^2(z)$ is a nonparametric estimator given by (5) with $z = \frac{y}{\sigma_{\text{pt}}}$ and $\hat{\theta}$ is the estimator of the set of parameters in a parametric model. For simplicity we assume that the variable of interest $y_t$ is a martingale difference process, thus we don’t model for the mean. Further the regularity conditions for the asymptotic theory of nonparametric estimator are same as those stated in Ziegelmann (2002).

Using the Taylor series expansion we have

$$g(y_t, \hat{\theta}) = g(y_t, \theta^*) + \frac{d(g(m))}{dm}(\hat{\theta} - \theta)$$

where $\frac{d(g(m))}{dm} = O_p(T^{-0.5})$. So $\hat{\sigma}_{\text{pt}}^2(y) = \sigma_{\text{pt}}^2(y) + O_p(T^{-0.5})$

Therefore we can express $\hat{\sigma}_{\text{spt}}^2(y)$ as

$$\hat{\sigma}_{\text{spt}}^2(y) = [\sigma_{\text{pt}}^2(y) + O_p(T^{-0.5})][\hat{\sigma}_{\text{npt}}^2(z)]$$

Next following Ziegelmann’s (2002) expansion of the exponential $\hat{\sigma}_{\text{npt}}^2(z)$ we can write

$$\hat{\sigma}_{\text{npt}}^2(z) = \sigma_{\text{npt}}^2(z) - 0.5h^2\beta^2e^\alpha + o_p(h^2)$$

where $\beta^2e^\alpha$ is the second derivative of $e^{\alpha + \beta(z_{t-1} - z)}$ at $z_{t-1} = z$ and $\hat{\sigma}_{\text{npt}}^2$ is the local linear nonparametric estimator of $\sigma_{\text{npt}}^2$. Further from Lemma 2 of Yao and Tong (1996) we know that

$$\hat{\sigma}_{\text{npt}}^2(z) = \sigma_{\text{npt}}^2(z) + \frac{1}{Thf(z)}\sum_{t=1}^TK(\psi_t)\{z_t^2 - \sigma_{\text{npt}}^2(z) - \hat{\sigma}_{\text{npt}}^2(z)(z_{t-1} - z)\}[1 + o_p(1)]$$
Where \( \psi_t = \frac{(z_t-1)z_t}{\hat{h}} \) and \( z_t^2 = \frac{y_t^2}{\psi_{pt}(y_t)} = \frac{\sigma_{spt}^2(y_t)\nu_t^2}{\psi_{pt}(y_t)} = \sigma_{npt}^2(z_t)\nu_t^2 \) from (3).

Now using this \( z_t^2 \) and rearranging \( \hat{\sigma}_{npt}^2(z) \) we can obtain
\[
\hat{\sigma}_{npt}^2(z) = [\sigma_{npt}^2(z) - \frac{1}{2}\beta^2\epsilon^2\hat{h}^2 + I_1 + I_2 + o_p(h^2)]
\]
where
\[
I_1 = \frac{1}{Th_f(z)} \sum_{t=1}^{T} K(\psi_t)\{\sigma_{spt}^2(z_{t-1}) - \sigma_{npt}^2(z) - \hat{\sigma}_{npt}^2(z)(z_{t-1} - z)\}
\]
and
\[
I_2 = \frac{1}{Th_f(z)} \sum_{t=1}^{T} K(\psi_t)\{\sigma_{npt}^2(z_{t-1})(\nu_t^2 - 1)\}
\]

It can be easily shown that \( I_1 = 0.5h^2\hat{\sigma}_{npt}^2(z)\sigma_K^2 + o_p(h^2) \).

Further from Fan and Yao (1998) it directly follows that
\[
(Th)^{0.5}I_2 \xrightarrow{D} N(0, f(z)^{-1}\sigma_{spt}^4(z)\lambda_t^2(z) \int K^2(\psi)d\psi)
\]

Using the expression of \( \hat{\sigma}_{pt}^2(y) \) and \( \hat{\sigma}_{npt}^2(z) \) we can write
\[
\hat{\sigma}_{spt}^2(y) = \sigma_{spt}^2(y)\sigma_{npt}^2(z) - \sigma_{pt}^2(y)\beta^2\epsilon^2\hat{h}^2 + O_p(T^{-0.5})\sigma_{npt}^2(z)
\]
\[
+ \sigma_{pt}^2(y)[I_1 + I_2] + O_p(T^{-0.5})[I_1 + I_2] + [\sigma_{pt}^2(y) + O_p(T^{-0.5})o_p(h^2)]
\]

Thus \( A_2 \) will asymptotically dominate the \( (Th)^{0.5}A_3 \) and \( (Th)^{0.5}A_4 \). \( (Th)^{0.5}A_1 \) will vanish asymptotically. So we are interested in finding out the order of \( A_2 \). For this we write
\[
A_2 = \sigma_{pt}^2(y)I_1 + \sigma_{pt}^2(y)I_2
\]
where
\[
J_1 = \sigma_{pt}^2(y)[0.5h^2\sigma_{npt}^2(z)\sigma_K^2 + o_p(h^2)] - O_p(h^2)
\]
and \( (Th)^{0.5}J_2 = N(0, f(z)^{-1}\sigma_{spt}^4(y)\lambda_t^2(z) \int K^2(\psi)d\psi) \)

From this we conclude that
\[
(Th)^{0.5}\{\hat{\sigma}_{spt}^2(y) - \sigma_{spt}^2(y) - \hat{\psi}_t\} \xrightarrow{D} N(0, f(z)^{-1}\sigma_{spt}^4(y)\lambda_t^2(z) \int K^2(\psi)d\psi)
\]

where \( \hat{\psi}_t = 0.5\sigma_{pt}^2(y)h^2[\hat{\sigma}_{npt}^2(z) - \beta^2\epsilon^2\sigma_K^2] \).

Since \( \hat{z}_t = z_t + O_p(T^{-0.5}) \) we also note that the above result also holds for the case of \( \hat{\sigma}_{npt}(\hat{z}) \). Further in the case of using the state variables \( x_3 \) in (5) the above result again holds with \( z \) replaced by \( x_3 \).