In the classical linear model framework, it is typical to assume that the conditional variance $\text{var}(Y|X) = \sigma^2$ is constant. While this assumption simplifies development of testing procedures, this is rarely the case. Alternatives such as generalized least squares attempts to remedy this issue but requires previous knowledge of the covariance structure and is not locally adaptive. That is the covariance between (Y, X) must be constant across the sample space $X \in \mathcal{X}$. This restrictive condition typically leads to several model violations and is usually resolved by finding abstract transformation functions that hinder inferential power. In the nonparametric setting, however, we can remove this issue by considering the following problem; given the model $Y_i = m(X_i) + \sigma(X_i)\epsilon_i$ how can we estimate $\sigma(X_i)$?

In *Efficient estimation of conditional variance functions in stochastic regression*, authors Fan and Yao address this question. They point out that this class of models is quite broad and modeling the conditional variance has many far reaching implications. One they stress is the case where $X_i = Y_{i-1}$. In this case, the model is an autoregressive conditional heteroscedastic (ARCH) model and $\sigma(\cdot)$ is the volatility function. Typically in practice, this quantity is estimated by

$$\hat{\sigma}_d^2(x) = \hat{\nu}(x) - \{\hat{m}(x)\}^2$$

where we estimate $m(x) = E(Y|X = x)$ and $\nu = E(Y^2|X = x)$ by nonparametric regression techniques. The authors point out that this estimate is not ideal in that it is not always nonnegative and can produce very large bias if the kernels or bandwidths are chosen independently. Another popular method is the difference estimator as in ARIMA modeling. While simple to implement, the authors note that the estimator was inefficient even in the most simple cases.

Instead of these estimators, the authors propose a very simple technique based on the squared residuals of the nonparametric regression. They argue that modeling the squared residuals is a way to remove the $m(X_i)$ term from the give model. Specifically, for a known regression function $m(\cdot)$ and $r = (Y - m(X))^2$ we have $E(r|X = x) = \sigma^2$. Therefore, we see that this problem can be solved a regression analysis with X as the covariate and r as the

dependent variable. But notice that $m(\cdot)$ is unknown. Therefore, we must first estimate it via nonparametric regression methods. The authors suggest using local linear models for a few nice properties (all though I believe the technique will work for any regression method). From here we estimate the r values from the squared residuals values. Lastly, we fit the local linear regression model for the bivariate data $(\hat{r}, X)$. To review, the method can be outlined as

1. Let $\hat{m}(x) = \hat{a}$ where $(\hat{a}, \hat{b}) = \underset{(a,b)}{\arg\min} \sum_{i=1}^{n} (Y_i - a - b(X_i - x))^2 K\left(\frac{X_i - x}{h_1}\right)$ be the local linear estimate of the bivariate data (X, Y)

2. Define $\hat{r}_i = (Y_i - \hat{m}(X_i))^2$

3. Let $\hat{\nu}(x) = \hat{\alpha}$ where $(\hat{\alpha}, \hat{\beta}) = \underset{(\alpha,\beta)}{\arg\min} \sum_{i=1}^{n} (\hat{r}_i - \alpha - \beta(X_i - x))^2 W\left(\frac{X_i - x}{h_2}\right)$ be the local linear estimate of the bivariate data $(X, \hat{r})$

4. $\hat{\nu}(x)$ is the estimated conditional variance function

Here we see that both the kernel and bandwidth in (1) and (3) can differ. The authors suggest choosing an optimal bandwidth for step (1) then recalculating an optimal bandwidth given the $\hat{r}_i$ in step (3). In this way no new machinery is needed and instead we can use optimal results. While the authors do not mention the effect of density on estimates, I imagine that like most applications they have little effect. Under certain regularity conditions, the authors also show that the "benchmark model" which is the local linear regression function between the (X_i, r_i) is asymptotically normal with leading bias terms $\frac{h^2}{2}\sigma_K^2 \ddot{\sigma}^2(x)$. The key to this term is that it is independent of the regression function. That is the method is design invariant. The authors go onto show that other direct estimators do not possess this property. Moreover, they show that their method generalized difference estimators. The only method that comes close to the efficiency/simplicity of their estimator is the Likelihood based estimator but the authors note that the computational issues with this method are laborious and not nearly as straightforward as their simple residual technique. The method introduced here is simple and has clear advantages.

In the paper *Correlation Curves as Local Measures of Variance Explained by Regression,* the authors introduce the idea of a local correlation function that quantifies the strength of relationship in bivariate data (X, Y) which varies over the covariate space. In regression and multivariate analysis, statistics such as r^2 and ρ^2 measure the *global* relationship between two variables (X, Y). While this is sufficient in special cases (e.g. linear regression analysis), in most cases random variables have complex nonlinear relationships that have variable dependency structure. This paper attempts to address this situation by constructing a *correlation curve* which measures the correlation between variables in local regions of the variable space.

The authors construct their estimates by simply extending common notions found in the linear model literature. For instance, they note for the linear model $Y = \alpha + \beta X + \epsilon$, the correlation coefficient from this regression can be written as

$$\rho^2 = \frac{\text{var}(\alpha + \beta X)}{\text{var}(\alpha + \beta X + \epsilon)} = \frac{\sigma_1^2 \beta^2}{\sigma_1^2 \beta^2 + \sigma_\epsilon^2}$$

where $\text{var}(X) \equiv \sigma_1$ and $\text{var}(\epsilon) \equiv \sigma_\epsilon^2$. Note that this is just the proportion of variance explained by the regression by the total variance in the estimate. Moreover, these quantities are all defined as global statistics. In order to make this quantity locally adaptive, the authors replace the globally estimated variables - β, σ_ϵ - with local estimators that allow for different estimates of the strength of relationships at different X values (more on this later). Given these estimates, the authors write the squared correlation curve as

$$\rho^2(x) = \frac{\sigma_1^2 \beta^2(x)}{\sigma_1^2 \beta^2(x) + \sigma^2(x)}$$

In order to estimate this curve, the authors implement two different procedures: Neighborhood Estimates and Kernel Estimators. For the neighborhood approach, they let $I_k(x)$ be the indices of the X values closest to x such that there are equal amount of points on the left and right of the evaluation point. Then the empirical estimates for $\sigma^2(x)$ and $\beta(X)$ are

given by

$$\hat{\sigma}^2(x) = \frac{1}{k} \sum_{i \in I_k(x)} Y_i^2 - \left(\frac{1}{k} \sum_{i \in I_k(x)} Y_i \right)^2 \qquad \hat{\beta}(x) = \frac{\overline{Y}^+(x) - \overline{Y}^-(x)}{\overline{X}^+(x) - \overline{X}^-(x)}$$

where k is the number of neighbors, and $\overline{Y}^+$ and $\overline{X}^+$ are the sample means of X_i, Y_j such that $i, j \in I_k(x)$ and $X_i > x$ and $Y_i > y$ where (x, y) is a sample point. In a similar sense $\overline{X}^-$ and $\overline{Y}^-$ are the sample means of neighborhood points to the left of the point (x, y).

Using these estimates, the authors show several desirable properties of $\hat{\rho}(x)$. Namely, $\hat{\rho}(x)$ is consistent when $\hat{\beta}(x)$ and $\hat{\sigma}^2(x)$ are consistent, and under certain regulations regarding the existence of higher order derivatives, $\hat{\rho}(x)$ has bias of the order $(k/n)^2$. Moreover, when $k = cn^\delta$, then for $2/3 < \delta < 1$ the AMSE tends to zero and for $\delta = 6/7$ the rate of convergence is optimal on the order of $n^{-4/7}$. Due to this set of results, they even derive asymptotic confidence intervals for $\hat{\rho}(x)$ assuming $2/3 < \delta < 6/7$ and the Lindeberg condition.

A second estimation technique the authors use is kernel estimators. Let $w(t)$ be a kernel on $[-\tau, \tau]$ with $w(-\tau) = 0 = w(\tau)$ and $\int_{[-\tau, \tau]} w(t)dt = 1$. Let $x_1 < \ldots < x_n$ and $s_i = \frac{x_i + x_{i+1}}{2}$ for $i = 1, 2, \ldots, n$. They then define three kernels w_j for $j = 0, 1, 2$. From here they define the local estimators as

$$\tilde{\mu}_j(x) = -\sum_{i=1}^{n} \left[W_j \left(\frac{x - s_i}{b_{jn}} \right) - W_j \left(\frac{x - s_{i-1}}{b_{jn}} \right) \right] Y_i^j \qquad \tilde{\sigma}^2(x) = \tilde{\mu}_2(x) - (\tilde{\mu}_2(x))^2$$

$$\tilde{\beta}(x) = \frac{-1}{b_{0n}} \sum_{i=1}^{n} \left[w_0 \left(\frac{x - s_i}{b_{0n}} \right) - w_0 \left(\frac{x - s_{i=1}}{b_{0n}} \right) \right] Y_i$$

where the b_{jn} are a sequence of bandwidths and $W_j(x) = \int_{-\infty}^{x} w_j(x)dx$. The authors go on to show several desirable properties of these estimates such as weak and strong consistency of $\tilde{\rho}(x)$ for $\rho(x)$ under the proper conditions. Moreover, as with the neighborhood approaches they derive asymptotic confidence intervals for the the true correlation curve. The authors suggest that the bandwidth is chosen with respect to reference kernel. They do note however,

that unless the reference kernel coincides with the true model, this method will not converge to the optimal bandwidth.

Turning to applications of the local correlation estimate, we consider the paper *Empirical Evidence on Spatial Contagion Between Financial Markets*. In this paper, the authors analyze the dependency structure between financial markets in good and bad financial times. In terms of the the model $Y = m(X) + \sigma(X)\epsilon$ they say there is a contagion from market X to market Y if $\rho(x_L) > \rho(x_M)$ for x_L a low quantile of the X distribution and x_M the median of the X distribution. In words, we say a market X is contagion from market Y if X depends on Y more when X does poorly. This immediately leads to the hypothesis test $H_0 : \rho(x_L) \leq \rho(x_M)$ and $H_A : \rho(x_L) > \rho(x_M)$. After analyzing the Q-Q and P-P plots of $\hat{\rho}(x_M)$ and $\hat{\rho}(x_L)$, the authors suggest that the normal test statistic

$$Z = \frac{\hat{\rho}(x_L) - \hat{\rho}(x_M)}{\sqrt{\hat{\sigma}^2_{\hat{\rho}(x_L)} + \hat{\sigma}^2_{\hat{\rho}(x_M)}}}$$

is sufficient to test this hypothesis. They first use this statistic on US mature equity markets against other international markets. After accounting for serial dependencies and other temporal correlative structure in market variables (via GARCH modeling), the authors found that about half of the markets considered were contagion from the US market while the US market was not in contagion to any other market.

The authors also apply this method to the *flight to quality* phenomena, where investors will seek lower risk investments after a dip in market performance. In our context, the bond market is the safe investment while the equity market is the "risky" market. The authors find evidence for this phenomena as the equity market is in contagion to the bond market. Through this analysis, the authors note that while the overall correlation coefficient between the bond and equity market was $\hat{\rho} = .09$, in certain regions of the distribution $\hat{\rho}(x_M) = .43$. Therefore, by modeling the local correlative structure of the two markets, they were able to see the true dependences between the two markets.