Title
Modeling and Estimating Unpredictability with Applications in Political Economy

Author
Yang, Feng

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Modeling and Estimating Unpredictability with Applications in Political Economy

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Science in Statistics

by

Feng Yang

2019

ABSTRACT OF THE THESIS

Modeling and Estimating Unpredictability
with Applications in Political Economy

by

Feng Yang
Master of Science in Statistics
University of California, Los Angeles, 2019
Professor Chad J. Hazlett, Chair

Social science theories often make predictions not only about the mean but also about the variance of the outcome of interest. For instance, comparative political scientists argue that democracies and non-democracies have, on average, the same rate of economic growth, but the former usually has less variance in the rate than the latter. Thus, both the mean and variance of economic performance can be functions of regime types. I review four important methods to model and estimate the error variance or its function: the naive two-stage estimation, variance function regression, joint maximum-likelihood estimation, and quantile regression (plus smoothing) estimation. Using simulated data, I compare the performance of these models when the sample size is small or large, when the variance function misspecification is mild or severe, and when the mean function is misspecified. I then apply these methods in two original studies: 1) How does county leaders' in-office time affect economic volatility in Chinese counties? 2) How does the pre-WTO economic unpredictability in Chinese municipalities affect foreign direct investment (FDI) inflows after China obtained its WTO membership in 2001?

The thesis of Feng Yang is approved.

Jeffrey B. Lewis

Erin K. Hartman

Chad J. Hazlett, Committee Chair

University of California, Los Angeles

2019

To Ada Xiaoting Li,

who enriches my life in many ways.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I owe many thanks to my thesis advisor, Chad Hazlett, for his tireless guidance and mentoring. I would also like to thank Erin Hartman and Jeff Lewis for their very constructive comments and suggestions. Without their passionate participation and input, this thesis could not have been successfully completed.

Special thanks also go to Ker-Chau Li, Jingyi Jessica Li, and Robert E. Weiss, whose enlightening teaching inspired me to begin this project.

VITA

2006-2010 B.M. (Financial Management) and LL.B. (Political Science), Nankai University, China.

2010-2012 M.Phil. (Social Science), The Hong Kong University of Science and Technology (HKUST), Hong Kong.

2013- Ph.D. candidate in Political Science, University of California, Los Angeles (UCLA),USA.

PUBLICATIONS

Strategic Promotion, Reputation, and Responsiveness in Bureaucratic Hierarchies (with Xinyu Fan), *Journal of Theoretical Politics*, 2019. dio:10.1177/0951629819850638

Organizational Structure, Policy Learning, and Economic Performance: Evidence from the Chinese Commune (with Joshua Eisenman), *Socius*, 2018, 4(1): 1-18.

Overseas Students, Returnees and the Diffusion of International Norms into Post-Mao China (with David Zweig), *International Studies Review*, 2014, 16(2), 252-263.

Authoritarian Orientations and Political Trust in East Asian Societies (with Deyong Ma), *East Asia: An International Quarterly*, 2014, 31(4), 323-41.

CHAPTER 1

Introduction

Assume a vector of random variable y_i is independently drawn from normal distributions $N(\mu_i, \sigma_i^2)$, $i \in \{1, 2, \cdots n\}$. Many applications of regression analysis are interested in modeling μ as a function of observed covariates, such that $\mu_i = X_i\beta$ for each observation i, and usually treat the error variance σ_i^2 – a measure for conditional unpredictability – as constant. A violation of the constant error variance will introduce heteroscedasticity and requires ex-post adjustment of the standard error estimates of $\hat{\beta}'s$, and possibly the use of weighted least squares (WLS) for more efficient estimation.

However, many social science theories predict that not only the mean, but also the variance, is a function of observed covariates. For instance, Western and Bloome (2009) find that incarceration reduces one's average earnings but increases the variance of earnings because the job tenure usually becomes shorter after incarceration. As another example, Sah (1991) hypothesizes that economic performance are on average the same in both democratic and non-democratic countries, but the performance is more volatile in the latter due to the centralization of political power to a small number of *fallible* elites. Quinn and Woolley (2001) make the same prediction based on the observation that democratic voters are risk averse and they vote out leaders who deliver volatile growth. Hence, both mean (μ_i) and error variance (σ_i^2) of economic performance are a function of regime type of county i: $\mu_i = f(\text{Dem}_i)$, and $\sigma_i^2 = g(\text{Dem}_i)$, where `Dem` is continuous measure of democracy level or a binary indicator of democracy and $f(\cdot)$ and $g(\cdot)$ are mean and variance functions, respectively. Here, the changing error variance can be crucial to support or to falsify theories. Thus, social scientists are often interested in modeling the heterogeneity of variance and want to do more than treating it as a constant nuisance parameter.

Besides, unpredictability can have significant effects on human behavior. For instance, one may argue that all else equal, a country's foreign direct investment (FDI) inflow is negatively affected by its macroeconomic unpredictability (e.g., Aizenman 2003). Although the hypothesis is straightforward, how to empirically assess it remains a challenge in particular because of the difficulty of measuring unpredictability. One simple measure may be the standard deviation or variance of the country's recent growth. However, this measure uses unconditional variance rather than conditional variance (i.e., the error variance), thus does not accurately capture the concept of unpredictability. For instance, a large standard deviation of economic growth in a country can be due to changes in *observable* factors that predict the mean economic growth, such as pro-business institutional reforms, rather than a larger unpredictability. [1]

How can one model and estimate the error variance? The present study compares four approaches. First, the naive two-step approach runs a linear regression and obtain the residuals at the first stage, and regresses the squared residuals–an asymptotically unbiased estimator for error variance–on observable covariates at the second stage. Second, the variance function regression adopts a iterative process and uses the predicted error variance at the second stage to construct weights and re-run the first-stage regression. The process is iterated until convergence (Western and Bloome 2009). Third, the joint maximum likelihood approach specifies the contribution of both mean and variance functions to the total likelihood and search for parameters in the mean and variance functions that maximize the total likelihood. Notably, all the three approaches require modeling variance functions, such as $\log(\sigma_i^2) = g(z'\lambda)$. The advantage is two-fold. On the one hand, if the variance functional form is explicit, it is straightforward to construct confidence intervals for parameters in the variance function, which can be used to test hypotheses. On the other hand, the variance function allows borrowing information from units with similar covariates to predict the error

[1] In other words, one can decompose the total variance of y into two parts: $\text{Var}(y) = \text{Var}(X'\beta) + \text{Var}(\epsilon)$. Thus, a large variance of y can be due to a large variance of X rather than a large error variance $\text{Var}(\epsilon)$.

variance, which will augment the error variance prediction. Thus, even though one unit just has one observation, researchers can still make a reasonable estimate of the error variance when many covariates are available.

A fourth alternative approach uses the quantile regressions to predict individual error variance without requiring a parametric modeling of the variance function. Rather than estimating the mean as in usual regressions, quantile regression can estimate the outcome variable at different quantiles, such as $\tau \in \{0.05, 0.25, 0.75, 0.95\}$. Then, using the estimated quantiles, one can fit a known distribution, such as a skewed-t distribution, for each observation through minimizing the sum of squared difference between the estimated and theoretical quantiles. This approach has been recently used to measure economic vulnerability (Adrian, Boyarchenko, and Giannone 2019). As quantile regressions plus smoothing do not require identifying predictors for the variance function, it ignores the contribution of covariates in predicting the error variance. For this reason, it is immune from misspecifications of the variance function. However, notably, it does borrow information from other observation to form the quantile estimates, which indirectly augment the error variance prediction.

The present study compares the performance of the four approaches using simulated data and applies them to answer two real-world questions. First, how local economic volatility/unpredictability is affected when Chinese county leaders miss their major promotion opportunities? Second, can the variation of foreign direct investment (FDI) inflows into Chinese municipalities after 2001 be traced to the economic volatility/unpredictability before China obtained its membership in the World Trade Organization (WTO)? These two questions have not been rigorously answered before, which may be due to a lack of proper tools to model and estimate economic volatility. The two applications demonstrate how the modeled variance function can be used for testing social science theories and how the predicted error variance can be used for scientific explorations, respectively.

The rest of the paper is organized in the following way. In chapter 2, I briefly review previous works on modeling and estimating error variance. In the third chapter, I provide details about the estimation in the four discussed approaches. In chapter 4, I compare their performance using simulated data. Chapters 5 and 6 discuss two empirical applications. The last chapter concludes.

CHAPTER 2

Review of Related Works

2.1 Heteroskedaticity

The presence of heteroscedasticity–the variance of residuals varying with explanatory variables–is commonly viewed as a nuisanced to be remedied in regression analysis. Along with the growing popularity of regression analysis in testing political science theories, various methods have been proposed to mitigate the problem. These methods usually share a common purpose: to adjust standard errors of the regression coefficient estimates so that researchers can judge whether the coefficients are statistically significant or not (see King and Roberts (2015) for a recent review and critique). Though useful, unfortunately, the ex-post adjustment of standard errors does not allow testing theoretical predictions regarding the error variance.

2.2 Conditional Heteroskedasticity

In economics and finance, there are many studies on Autoregressive Conditional Heteroskedasticity (ARCH). Engle's (1982) pioneering study models the conditional variance of the error term at time t (i.e., σ_t^2) as a linear function of past error term variance, namely the error term follows the ARCH process. For instance, $\sigma_t^2 = \omega + \lambda_1\epsilon_{t-1}^2 + \lambda_2\epsilon_{t-2}^2$, where ϵ_{t-1} and ϵ_{t-2} are error variance in the recent past. Bollerslev (1986) generalized the ACH model (i.e., GARCH) by allowing the error term variance as a linear function of both unconditional and conditional variances of the error term in the recent past. Hence, using the example above, $\sigma_t^2 = \omega + \lambda_1\epsilon_{t-1}^2 + \lambda_2\epsilon_{t-2}^2 + \gamma_1\sigma_{t-1}^2 + \gamma_2\sigma_{t-2}^2$. Both models can be estimated using Maximum Likelihood. Since then, existing studies have advanced the original ARCH model by relaxing various assumptions so that the conditional density (of the outcome variable) does not need to be normal and that the variance function does not need to be linear or parametric. (See Bollerslev, Chou, and Kroner (1992) for a review.)

In general, these models aim to quantify how past information of one variable, such as inflation rate of a country, affects the current conditional variance of the same variable with time-series data. Thus, an advantage is that ARCH models does not require "specifying the causes of the changing variance" (Engle 1982, p. 988). However, due to the same reason, these models usually do not permit testing hypotheses regarding the causes of the changing variance, which is the interest of the present paper. Additionally, in cases where only cross-section data is available or the time dimension of the time-series data is short, this class of models may not be applicable.

2.3 Measuring Volatility

Besides the ARCH models, economists have introduced alternative methods to estimate volatility of economic outcomes. For instance, Blanchard and Simon (2001) focus on the unconditional variance of economic output and measure economic output volatility in the US with the standard deviation of quarterly GDP growth. Given that the mean and variance of economic output could co-evolve, others pay more attention to variance of economic outcomes conditional on its mean. For example, using cross-national panel data, Chandra and Rudra (2015) propose a growth adjusted volatility measure by first estimating the growth-instability frontier and then subtract the feasible minimum instability from the actual standard deviation of economic growth over an one-decade period. Unsurprisingly, measuring volatility with standard deviations is data demanding. For instance, the standard deviations are computed using quarterly data from a rolling sample of 20 quarters in Blanchard and Simon (2001).

Similar to other cross-national studies (e.g., Acemoglu et al. 2003; Quinn and Woolley 2001; Yang 2008), Chandra and Rudra (2015) compute standard deviation of growth for countries using annual data over some years.

More recently, researchers use a Bayesian approach and quantile regressions to measure economic volatility with the estimated error variance. For instance, Nakamura and coauthors (2017) use a Bayesian approach to estimate the evolution of the worldwide and country-specific stochastic volatility of consumption. Alternatively, Adrian, Boyarchenko, and Giannone (2019) use quantile regressions to predict GDP growth at a few different quantiles and fit the conditional quantile estimates using skewed-t distributions. The estimated parameters of the skewed-t distributions allow construction of a measure of error variance. I will provide more details about the quantile regression approach in the next section.

2.4 Variance Function Regression

There are also many sociological studies that aim to model and predict error variance using covariates. A notable example is Western and Bloome (2009) where the authors use the variance function regression to study inequality, which is measured with the variance of log earnings. Both the mean and variance are modeled as functions of some observable covariates. As will be detailed in the next section, they propose a iterative process to estimate the mean and variance functions until convergence. The result from the last iteration will be used to construct the confidence intervals for parameters in both mean and variance regressions.

This approach has been frequently used in the recent studies. For instance, Zhou (2014) applies it to decompose the overall inequality in China to different levels of observed covariates and the residual part. Relatedly, Zhou (2019) models the variance of individuals' latent political preference as a function of time in the US. As the preference variance across individuals can be conceptualized as mass polarization, the author further examines the temporal changes of mass polarization in the US.

CHAPTER 3

Modeling and Estimating Unpredictability

In this section, I will introduce more details about estimation used in the aforementioned approaches. Assume that the data generating process takes the following form in Eq. 3.1. i indicates observations. x and z are predictors for the mean and variance, which can be the same or different.

In the conventional notation (e.g., King (1998)),

$$\begin{aligned} \mathrm{Y}_i &\sim N(\mu_i, \sigma_i^2) \\ \mu_i &= x_i'\beta \\ log(\sigma_i^2) &= z_i'\lambda \end{aligned} \tag{3.1}$$

3.1 Naive Two-Step Regression (Naive two-step)

1. Estimate β with a linear regression of y on x. Obtain and save the residuals, $\hat{e}_i = y_i - \hat{\beta}x_i$, where $\hat{\beta}$ is the least squares estimate.

2. Estimate λ with a gamma regression of the squared residuals, $\hat{e}_i^2$ on z_i, using a log link function, which gives $\hat{\lambda}$ and $\hat{\sigma}_i^2 = \exp(z_i'\hat{\lambda})$.

3.2 Variance Function Regression (VFR)

Western and Bloome (2009) propose iteratively estimating the two steps in the Naive approach using maximum likelihood. Whenever the sample is small, they suggest using restricted maximum likelihood estimation (RMLE). Following Western and Bloome (2009), I propose the following four-step estimation.

1. Estimate β with a linear regression of y on x. Obtain and save the residuals, $\hat{e}_i = y_i - \hat{\beta}x_i$, where $\hat{\beta}$ is the least squares estimate.

2. Estimate λ with a gamma regression of the squared residuals, $\hat{e}_i^2$ i , on z_i, using a log link function. Obtain the fitted values $\hat{\sigma}_i^2 = \exp(z_i'\hat{\lambda})$.

3. Fit a weighted linear regression of y on x , with weights, $1/\hat{\sigma}_i^2$. Update the residuals $\hat{e}_i$, and evaluate the log-likelihood.

4. Iterate steps 2 and 3 to convergence, updating $\hat{\beta}$ and $\hat{e}_i$ from the weighted linear regression, and $\hat{\lambda}$ and $\hat{\sigma_i^2}$ from the gamma regression. The convergence is achieved if the improvement of log-likelihood from last-round estimation is less than the tolerance parameter, 10^{-6}.

A advantage of a gamma regression for the variance function is that the interpretation is straightforward, thus suitable for inference purposes. However, the variance function does not necessarily need to be estimated with generalized linear models or other parametric models. Instead, one can estimate the variance function with machine learning methods, espeicially when the major purpose is predicting the variance rather than making inferences about the data generating process of the variance. For instance, under the feasible generalized least squares' (FGLS) framework, Miller and Startz (2018) use support vector regression (SVR) to predict the error variance.

3.3 Joint Maximum-Likelihood Estimation (Joint ML)

As discussed in Aitkin (1987), each observation i's contribution to the total log-likelihood in Eq. 3.1 is:

$$
\begin{aligned}
L(\beta, \lambda, y_i) &= -\frac{1}{2}[\log(\sigma_i^2) + \frac{(y_i - \hat{y}_i)}{\sigma_i^2}] \\
&= -\frac{1}{2}[z_i'\lambda + (y_i - x_i'\beta)\exp(-z_i'\lambda)]
\end{aligned}
\tag{3.2}
$$

Thus, one can simultaneously search for a pair of $\hat{\beta}$ and $\hat{\lambda}$ that maximize the sum of the log-likelihood.[1]

3.4 Quantile Regression plus Smoothing (QR+SM)

Following Adrian, Boyarchenko, and Giannone (2019), I conduct a quantile regression and estimate $\hat{y}$ at different quantiles and then smooth the estimates using a well-known distribution. It is worth noting that they used a skewed-t distribution to smooth the quantile estimates, which is more flexible than normal distributions. But the computational cost is also large. Here, for simplicity I fit the normal distribution to the predicted y at two quantiles, $\tau \in \{0.05, 0.095\}$. Only two quantiles are needed because a normal distribution can be characterized by its two parameters, μ and σ. This approach involves three steps:

1. Run quantile regressions at the 5 and 95 percent quantiles, $\tau \in \{0.05, 0.095\}$.

2. Obtain predicted quantile of y conditional on x from that regressions:

$$\widehat{Q_{y|x}}(\tau|x) = x'\hat{\beta}_\tau$$

   Now each observation will have two predicted quantile values of y.

[1] For instance, this can be done using the `optim` function in `R`.

3. Choose a pair of μ_i and σ_i (for each observation) to match the 5 and 95 percent quantiles:

$$\{\hat{\mu}_i, \hat{\sigma}_i\} = \arg\min_{\mu_i,\sigma_i} \sum_{\tau} \left(\widehat{Q_{y|x}}\left(\tau|x\right) - F^{-1}\left(\tau; \mu_i, \sigma_i\right)\right)$$

, where $F^{-1}(\cdot)$ is the inverse cumulative distribution function (i.e. quantile function) of normal distributions. This gives $\hat{\mu}_i$ and $\hat{\sigma}_i$.

CHAPTER 4

A Simulation Example

In this section, I show some simulation results. More specifically, I demonstrate how changing sample size ($n \in \{100, 250, 500\}$) and different levels of misspecification of model and variance functions affect the relative performance of the methods mentioned above. I start with assuming that researchers correctly know the specifications for both the mean and variance functions. Next, I consider a more realistic case in which researchers do not have a good measure of the true causes of the variance, but only a noisy proxy. Finally, I also consider how the misspecified mean function affects the estimation.

$$\begin{aligned} \mu_i &= 1 - 2x_i + 10x_i^2 \\ \log(\sigma_i^2) &= 3z_i \\ \mathrm{y}_i &\sim N(\mu_i, \sigma_i^2) \end{aligned} \tag{4.1}$$

The true generating process of the data is shown in Equation 4.1: x and x^2 are two predictors for the mean regression and z is a predictor for the variance function. $\beta_0 = 1$, $\beta_1 = -2$, $\beta_2 = 10$, $\lambda_0 = 0$, $\lambda_1 = 3$. $x \sim U(-0.5, 0.5)$, and is fixed across the $N = 500$ simulations. For simplicity, $z = x$.

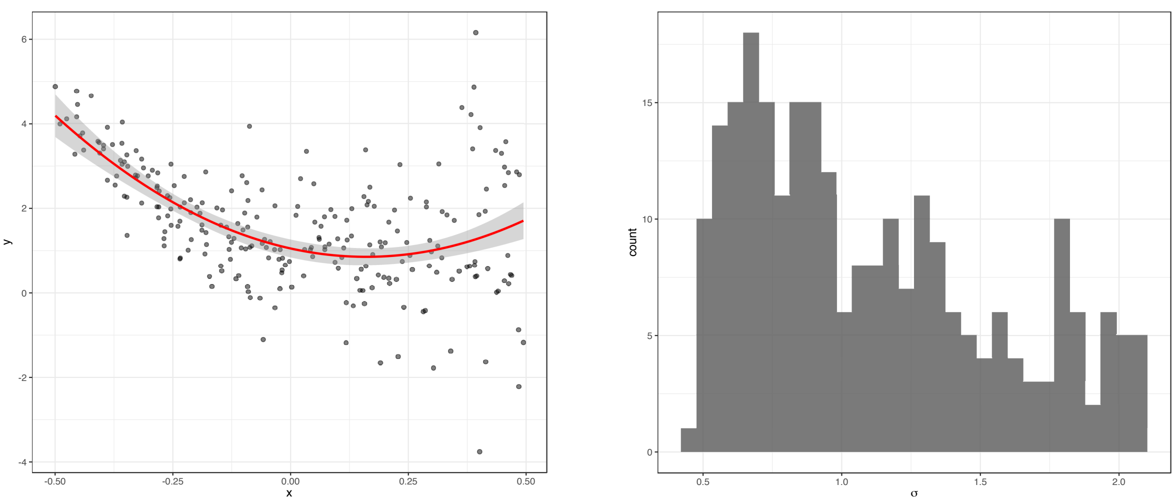
Figure 4.1 panel (a) shows the plot between x and y when $n = 250$. It is apparent that the error variance becomes larger when x increases. (Again, this is because I set $z = x$ for simplicity.) Panel (b) shows the distribution of the true σ.

To compare the performance of the four methods, I focus on three quantities of interest.

First, the bias of parameter estimates, defined as $E\big(\hat{\theta}-\theta\big)=\frac{\sum_{s=1}^{N}\hat{\theta}^{s}-\theta}{N}$. $\hat{\theta}^{s}$ is the estimated $\hat{\theta}$ from a simulation s out of $N=500$ simulations. θ can be $\beta's$ from the mean regression or $\lambda's$ from the variance regression. Second, the precision of parameter estimates, which is the sampling variance of the estimates in the 500 simulations. And finally, I compare the accuracy of the individual error variance estimates σ_i^2, which has two indicators. One is the expected value of mean squared difference between $\hat{\sigma}$ and σ: $\mathrm{E}\big(\frac{\sum_i^n(\sigma_i-\hat{\sigma_i})^2}{n}\big)$ where n is the sample size. The other is the expected value of the correlation between $\hat{\sigma}$ and σ: $\mathrm{E}\big(Cor(\hat{\sigma},\sigma)\big)$. While the former resembles the bias of $\beta's$ and $\lambda's$, the latter reveals the risk of using the predicted error variance/standard deviation as an independent variable for further scientific inquiries. For the latter, a larger expected correlation indicates better performance.

Figure 4.1: One Realization of Simulated Data ($n=250$)

(a) Plot x and y (b) Distribution of σ



4.1 Correctly Specified Mean and Variance Models

Table 4.1 compares the estimation bias and accuracy of $\hat{\sigma}$ when both the mean and variance models are correctly specified. Across different sample sizes, the bias of the parameter estimates for the mean regression remains low, because heteroskedasticity is well know not to bias the coefficient estimation in the mean regression. Because the variance function is correctly specified, the bias of $\hat{\lambda}_0$ and $\hat{\lambda}_1$ is small across different sample sizes. However, Joint ML and VFR do have smaller bias for $\hat{\lambda}_1$ than the Naive two-step.

To compare the accuracy of $\hat{\sigma}$, besides simulation results from the four methods, I also include the results from which the absolute value of residuals from the first-stage regression is used as the estimate for individual error standard deviation (see the row `Residuals` ($\hat{\sigma} = \mid \hat{e} \mid$)). Quite strikingly, the accuracy of $\hat{\sigma}$ predicted by the four methods is much higher than the naive absolute value of residuals when the sample size is not very large ($n \leqslant 500$). Also, while the accuracy of $\hat{\sigma}$ increases along with an increasing sample size in the four discussed methods, it does not get much better if one uses the naive regression residuals as the estimate for σ. This is likely because the Naive two-step, the Joint ML and the VFR all directly borrow information about the error variance from similar observations to augment the prediction of $\hat{\sigma}$ while QR+SM borrows information to form quantile estimates, which also contributes to the prediction augmentation. Also because QR+SM does not directly borrow information about error variance from other units, as shown in last two columns of Table 4.1, it performs relatively worse than the other three methods in accurately predicting σ *when researchers know the mean and variance function specifications and the data is available.*

Table 4.1: Simulation Results from Correctly Specified Mean and Variance Models

		Bias					$\hat{\sigma}$ accuracy	
Sample (n)	Method	$E(\hat{\beta}_0 - \beta_0)$	$E(\hat{\beta}_1 - \beta_1)$	$E(\hat{\beta}_2 - \beta_2)$	$E(\hat{\lambda}_0 - \lambda_0)$	$E(\hat{\lambda}_1 - \lambda_1)$	$E\left(\frac{\sum_i^n (\sigma_i - \hat{\sigma}_i)^2}{n}\right)$	$E(Cor(\hat{\sigma}, \sigma))$
	Naive two-step	-0.003	-0.039	-0.030	-0.045	-0.071	0.018	0.999
	Joint ML	0.003	-0.047	-0.121	-0.052	-0.003	0.018	0.999
n=100	VFR	0.003	-0.047	-0.121	-0.052	-0.003	0.018	0.999
	QR + SM						0.036	0.973
	Residuals ($\hat{\sigma} = \mid \hat{e} \mid$)						0.557	0.448
	Naive two-step	-0.002	-0.008	0.015	-0.015	-0.021	0.006	1.000
	Joint ML	-0.001	-0.009	-0.006	-0.019	0.011	0.006	1.000
n=250	VFR	-0.001	-0.009	-0.006	-0.019	0.011	0.006	1.000
	QR + SM						0.014	0.992
	Residuals ($\hat{\sigma} = \mid \hat{e} \mid$)						0.569	0.443
	Naive two-step	0	0.004	0.037	-0.005	-0.010	0.003	1.000
	Joint ML	0	0.004	0.038	-0.006	0.003	0.003	1.000
n=500	VFR	0	0.004	0.038	-0.006	0.003	0.003	1.000
	QR + SM						0.006	0.996
	Residuals ($\hat{\sigma} = \mid \hat{e} \mid$)						0.565	0.458

Table 4.2 shows the sampling variance of parameter estimates when both mean and variance models are correctly specified. A smaller value indicates more precise estimation. In general, all parameter estimates become more precise when sample size becomes larger. Also, for each given sample size, the Joint ML and VFR have more precise mean regression parameter estimates than the Naive two-step approach does. All approaches have similarly precise estimates for the variance regression, which again is due to the correct specification of the variance regression.

Combining findings of bias and precision, it seems that joint ML and VFR have similar performance, both of which are better than that of the Naive two-step. Thus, either joint ML or VFR can be used to test hypotheses about $\beta's$ and $\lambda's$. Because QR+SM does not model the variance function, it does not allow directly making inferences about the variance function. Instead, it provides $\hat{\sigma}$ for each observation, thus estimating $\lambda's$ will require an additional modeling using $\hat{\sigma}$ as the outcome variable and z as the independent variable.

Table 4.2: Sampling Variance of Parameter Estimates from Correct Mean and Variance Models

		Sampling Variance of Point Estimates				
Sample	Methods	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\lambda}_0$	$\hat{\lambda}_1$
n=100	Naive two-step	0.026	0.211	3.258	0.022	0.302
	Joint ML	0.020	0.205	2.190	0.021	0.313
	VFR + SM	0.020	0.205	2.190	0.021	0.313
n=250	Naive two-step	0.010	0.086	1.538	0.008	0.110
	Joint ML	0.008	0.083	0.864	0.008	0.109
	VFR + SM	0.008	0.083	0.864	0.008	0.109
n=500	Naive two-step	0.005	0.042	0.572	0.004	0.049
	Joint ML	0.004	0.040	0.357	0.004	0.049
	VFR + SM	0.004	0.040	0.357	0.004	0.049

4.2 Correct Mean Model plus Incorrect Variance Model

Recall that in the true data generating process, $\log(\sigma_i^2) = 3z_i$ and for simplicity, $x_i = z_i$. Above has assumed that researchers know specification of the variance function and have available data for z. Next, I explore a more realistic case where data of the *true* variance function predictor (z) is unavailable. Instead, only a proxy variable $\tilde{z}$ is available. I let the correlation between z and $\tilde{z}$ – namely, $\rho_{z,\tilde{z}}$ – vary and compare the relative performance of the four methods. Specifically, $\rho_{z,\tilde{z}} \in \{0.01, 0.5, 0.75, 1\}$.[1] The case $\rho_{z,\tilde{z}} = 1$ corresponds to the case where $z = \tilde{z}$ that has been discussed in the previous section. To focus on how changing $\rho_{z,\tilde{z}}$ affects the estimation, I fix the sample size at $n = 500$.

Table 4.3 summarizes the findings. As expected, incorrect specification of the variance function does not substantively affect the bias of the mean regression parameter estimates, which remains low. However, along with the declining $\rho_{z,\tilde{z}}$, the accuracy of $\hat{\sigma}$ declines as well. Interestingly, when $\rho_{z,\tilde{z}} < 0.75$, QR+SM outperforms the other three in predicting σ. The finding suggests that QR+SM can be a more reliable strategy to predict σ when researchers are uncertain about the variance function specification or the data is unavailable.

4.3 Incorrect Mean Model plus Correct Variance Model

Recall that μ is a linear function of both x and x^2 in the true data generating process. In this section and the next, I consider two cases in which the mean model is misspecified and only x is included as a predictor for μ.

This section discusses the case in which the mean model is incorrectly specified while the variance model is correct. Table 4.4 summarizes the results, which are in sharp contrast with those reported in Table 4.1 in which the mean model is correctly specified. First, because

[1] I generate a random variable $\eta \sim N(0, 0.5)$. $\tilde{z} = z$, $\tilde{z} = 0.5(\eta + 2z)$, $\tilde{z} = 0.5(\eta + z)$, and $\tilde{z} = \eta$ correspond to $\rho_{z,\tilde{z}} = 1$, $\rho_{z,\tilde{z}} = 0.75$, $\rho_{z,\tilde{z}} = 0.5$, and $\rho_{z,\tilde{z}} = 0.01$, respectively.

Table 4.3: Simulation Results from Correct Mean Model plus Incorrect Variance Model

		Bias			$\hat{\sigma}$ accuracy	
$Cor(z, \tilde{z})$	Method	$\text{E}\big(\hat{\beta}_0 - \beta_0\big)$	$\text{E}\big(\hat{\beta}_1 - \beta_1\big)$	$\text{E}\big(\hat{\beta}_2 - \beta_2\big)$	$\text{E}\big(\frac{\sum_i^n(\sigma_i - \hat{\sigma}_i)^2}{n}\big)$	$\text{E}\big(Cor(\hat{\sigma}, \sigma)\big)$
$\rho = 1.00$	Naive two-step	0	0.004	0.037	0.003	1.000
	Joint ML	0	0.004	0.038	0.003	1.000
	VFR	0	0.004	0.038	0.003	1.000
	QR + SM				0.006	0.996
$\rho = 0.75$	Naive two-step	0	0.003	0.034	0.101	0.748
	Joint ML	-0.001	0.005	0.050	0.101	0.748
	VFR	-0.001	0.005	0.050	0.101	0.748
	QR+ SM				0.006	0.996
$\rho = 0.5$	Naive two-step	0	0.003	0.034	0.171	0.500
	Joint ML	-0.001	0.005	0.046	0.171	0.500
	VFR	-0.001	0.005	0.046	0.171	0.500
	QR + SM				0.006	0.996
$\rho = 0.01$	Naive two-step	0	0.003	0.034	0.228	0.002
	Joint ML	0	0.004	0.034	0.229	0.002
	VFR	0	0.004	0.034	0.229	0.002
	QR+ SM				0.006	0.996

the mean function is misspecified in Table 4.4, the bias of $\hat{\beta}_0$ and $\hat{\beta}_1$ becomes much larger than those reported in Table 4.1. Second, the mean function misspecification also biases the estimates of λ_0 and λ_1 while such bias does not necessarily become smaller when the sample size increases. For instance, when $n = 500$, the bias of $\hat{\lambda}_1$ estimated using the VFR approach is 0.003 in Table 4.1 but becomes 0.085 in Table 4.4. Thirdly, the incorrect mean model also negatively affects the accuracy of $\hat{\sigma}$. If one predict σ using QR+SM, the "bias" of $\hat{\sigma}$ (i.e., $\text{E}\big(\frac{\sum_i^n(\sigma_i - \hat{\sigma}_i)^2}{n}\big)$) is 0.006 in Table 4.1 but becomes 0.078 in Table 4.4, which is more than ten times larger than the former. However, quite interestingly, despite of the increased "bias" of $\hat{\sigma}$, the correlation between $\hat{\sigma}$ and σ remains high when the mean function is incorrect. Thus, although $\hat{\sigma}$ does not equal the true σ when the mean function is misspecified, the strong correlation between the former and the latter may still make $\hat{\sigma}$ a meaningful measure if one wants to uses it as a variable for scientific investigations.

Table 4.4: Simulation Results from Incorrect Mean Model plus Correct Variance Model

		Bias				$\hat{\sigma}$ accuracy	
Sample (n)	Method	$\mathrm{E}\big(\hat{\beta}_0 - \beta_0\big)$	$\mathrm{E}\big(\hat{\beta}_1 - \beta_1\big)$	$\mathrm{E}\big(\hat{\lambda}_0 - \lambda_0\big)$	$\mathrm{E}\big(\hat{\lambda}_1 - \lambda_1\big)$	$\mathrm{E}\big(\frac{\sum_i^n(\sigma_i - \hat{\sigma}_i)^2}{n}\big)$	$\mathrm{E}\big(Cor(\hat{\sigma}, \sigma)\big)$
	Naive two-step	0.798	0.029	0.479	-1.410	0.092	0.996
	Joint ML	0.555	-1.960	0.368	0.053	0.112	0.999
n=100	VFR	0.556	-1.960	0.368	0.053	0.112	0.999
	QR+SM					0.097	0.983
	Residuals ($\hat{\sigma} = \mid \hat{e} \mid$)					0.702	0.334
	Naive two-step	0.741	0.352	0.494	-1.424	0.083	0.996
	Joint ML	0.533	-1.723	0.378	0.031	0.082	1.000
n=250	VFR	0.533	-1.723	0.378	0.031	0.082	1.000
	QR+SM					0.078	0.983
	Residuals ($\hat{\sigma} = \mid \hat{e} \mid$)					0.718	0.331
	Naive two-step	0.811	0.185	0.525	-1.386	0.088	0.996
	Joint ML	0.556	-1.931	0.418	0.086	0.092	1.000
n=500	VFR	0.556	-1.931	0.418	0.085	0.092	1.000
	QR+SM					0.078	0.982
	Residuals ($\hat{\sigma} = \mid \hat{e} \mid$)					0.723	0.344

4.4 Incorrect Mean and Variance Models

Now, I consider the case in which both the mean and variance models are misspecified. As in previous section, the mean model is misspecified because only x (not x^2) is included in the mean model. I then let the $\rho_{z,\tilde{z}}$ vary and investigate how the changing $\rho_{z,\tilde{z}}$ affects the estimation. Again, to focus on the effect of changing $\rho_{z,\tilde{z}}$ on estimation, I fix the sample size at $n = 500$.

As summarized in 4.5, the accuracy of $\hat{\sigma}$ declines when the misspecification of variance model becomes more severe (indicated by a decreasing $\rho_{z,\tilde{z}}$). When $\rho = 0.75$, the "bias" of $\hat{\sigma}$ is almost doubled from the case in which $\rho = 1$ if one uses the naive two-step, joint ML or the VFR estimations. Meanwhile, the correlation between σ and its estimate declines to 0.75. When ρ further decreases, both the bias and correlation indicators reflect worse performance of the three approaches. In addition, Tables 4.4 and 4.5 jointly suggest that the $\hat{\sigma}$ bias indicator is sensitive to both mean and variance models misspecification while the correlations indicator is more sensitive to variance model misspecification than to mean model misspecification. Thus, the observation from Table 4.4 that the mean function misspecification does not significantly undermine the correlation between σ and $\hat{\sigma}$ is likely because of its correctly specified variance model. Like in Table 4.3, QR+SM performs relatively better than other three approaches when the variance function is misspecified.

In sum, when both the mean function and the variance function are correctly specified, the VFR and joint ML deliver unbiased estimation of parameters (i.e., λ's) in the variance function and generate accurate predictions of σ. Thus, they can be used to construct confidence intervals and test hypotheses regarding these parameters, and/or to predict σ.

However, a misspecified mean function will bias the estimation of variance function parameters. Meanwhile, misspecified variance functions yield poor predictions of σ. On the one hand, unfortunately, there are not easy remedies for these problems other than collecting more information about the mean and variance functions to propose correct model specifi-

Table 4.5: Simulation Results from Incorrectly Specified Mean and Variance Models

		Bias		$\hat{\sigma}$ accuracy	
$Cor(z, \tilde{z})$	Method	$\text{E}\big(\hat{\beta}_0 - \beta_0\big)$	$\text{E}\big(\hat{\beta}_1 - \beta_1\big)$	$\text{E}\big(\frac{\sum_i^n(\sigma_i - \hat{\sigma_i})^2}{n}\big)$	$\text{E}\big(Cor(\hat{\sigma}, \sigma)\big)$
$\rho = 1$	Naive two-step	0.811	0.185	0.088	0.996
	Joint ML	0.556	-1.931	0.092	1.000
	VFR	0.556	-1.931	0.092	1.000
	QR+SM			0.078	0.982
$\rho = 0.75$	Naive two-step	0.811	0.185	0.177	0.749
	Joint ML	0.747	-0.659	0.173	0.749
	VFR	0.747	-0.659	0.173	0.749
	Quantile Reg			0.078	0.982
$\rho = 0.5$	Naive ML	0.811	0.185	0.250	0.499
	Joint ML	0.794	-0.104	0.249	0.500
	VFR	0.794	-0.104	0.249	0.500
	Quantile Reg			0.078	0.982
$\rho = 0.01$	Naive two-step	0.811	0.185	0.310	0.000
	Joint ML	0.811	0.179	0.310	0.000
	VFR	0.811	0.179	0.310	0.000
	Quantile Reg			0.078	0.982

cations. On the other hand, in the presence of mild misspecification of the mean function, QR+SM generates reasonably accurate prediction of σ even when the variance function is misspecified as well. Thus, when inference of the variance function is a key purpose, one may use the VFR or joint ML to test hypotheses about λ's while also using QR+SM as a robustness check for model misspecification. Application I is an example of such. Additionally, when researchers cannot specify the variance function ex ante but want to predict individual estimated error variance, QR+SM can be an appropriate strategy because it does not require much knowledge about the causes of the variance function. Application II shows how the predicted $\hat{\sigma}$ can be used as an independent variable.

CHAPTER 5

Application I: In-office Time and Economic Unpredictability

5.1 Theory

In China, economic decentralization has greatly empowered county leaders in local economy management. Thus, their competence and effort largely determine local economic growth. In addition, to motivate local bureaucrats to promote economic growth, the Chinese ruling party selects those county leaders whose jurisdiction experiences a satisfactory growth rate ($\hat{g}$) to higher-ranking political offices. In response, local bureaucrats spend efforts to meet the target $\hat{g}$ set by the ruling party. If a local bureaucrat is talented, s/he does not need to spend too much effort to meet the target; however, a incompetent local bureaucrat may need to spend much more efforts to meet the same target.

Because effort is costly, it is natural to expect that the effort spent by local bureaucrats depends on their expected promotion chance. When future promotion is possible, a bureaucrat spends certain efforts to compensate for his/her (in)competence so that local economic growth will meet the target. Thus, local economic growth all converges to the set target. However, when future promotion chance is low, bureaucrats will not spend any costly efforts and local economic growth will mainly depend on their competence, which can vary widely. Thus, the economic unpredictability – the variance of economic growth – in the latter group is larger than that in the former. (To more clearly show the mechanism, I include a simple formal model in the Appendix.)

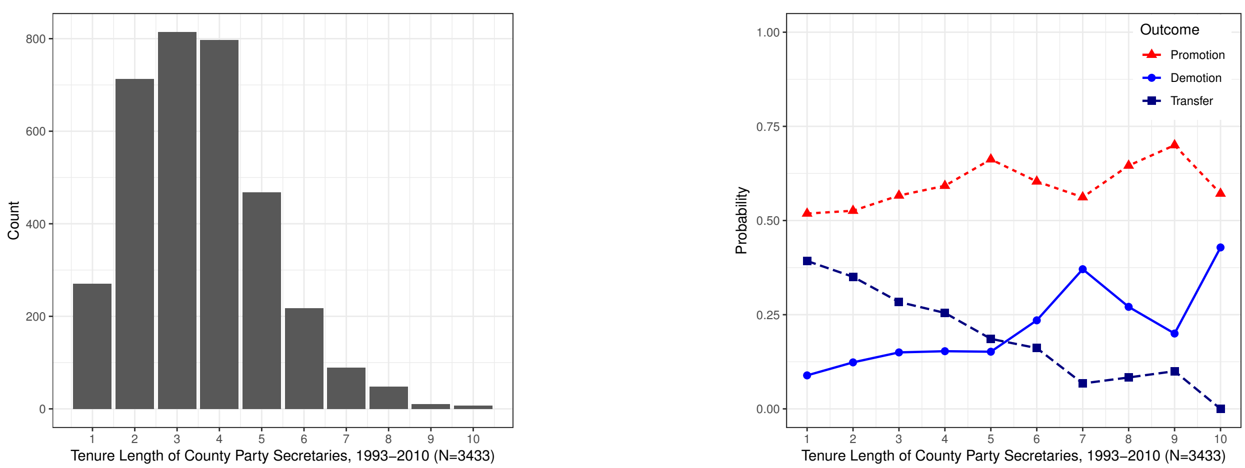
Empirical evidence suggests that Chinese county leaders' in-office time is meaningful indicator for their future promotion chance (e.g., Guo 2009). After collecting information of 3433 Chinese county party secretaries between 1993 and 2010, I show that the turnover rate of county party secretaries is quite high: many of them face turnovers within the first four years in office (panel *a*). In addition, and more importantly, their promotion chance increases with their in-office time within the first 5 years, but declines afterward (panel *b*). Thus, those who have already stayed at their position for more than five years ($t_i > 5$) should expect a low promotion probability in the future. More importantly, the low expected promotion chance may translate into larger economic unpredictability, as stated in H1.

Hypothesis 1 (H1): The unpredictability of economic growth is larger among officials who have stayed at their positions for more than five years than among those having spent less time.

Figure 5.1: Tenure Length and Turnover Outcomes of Chinese County Officials

(a) Distribution of Tenure Length

(b) Turnover Outcomes



Note: Calculations based on 3433 Chinese county party sectaries between 1993 and 2010, for whom the final turnover outcome data is available.

5.2 Data

To test H1, I collect annual economic data for 1249 Chinese counties between 2002 and 2005. After further excluding observations with missing values, 4660 county-year observations are left.

The dependent variable is annual per-capita GDP growth (i.e., $\frac{GDP_{i,t}-GDP_{i,t-1}}{GDP_{i,t-1}}$). i indicates counties while t denotes years. In the variance regression, I include the categorical measure for the in-office time of county party secretaries: `Middle` if they have stayed for three to five years, and `Long` if they have stayed for more than five years. The omitted category `Short` (1-2 yrs) is the reference. According to H1, I expect λ_2 to be positive in specification 5.1. Similarly, I also explore using polynomials of the continuous measure of in-office years.

In the mean regression, I include a set of variables (X), such as a categorical measure for the in-office time of county party secretaries, local population size, per-capita GDP in the previous year, fiscal revenue in the previous year, province dummies, and year dummies.

$$\begin{aligned} \text{growth}_{it} &\sim N(\mu_{it}, \sigma^2_{it}) \\ \mu_{it} &= X'_{it}\beta \\ \log(\sigma^2_{it}) &= \lambda_0 + \text{Middle}_{i,t}\lambda_1 + \text{Long}_{i,t}\lambda_2 \end{aligned} \tag{5.1}$$

5.3 Results

Table 5.1 shows the results estimated using VFR or QR+SM. In models (1) and (3), the signs of `Long` is positive, which supports H1. In models (2) and (4), the polynomials of the continuous measure of in-office years yields similar results. Model (4) use second-order polynomial because the third-order polynomial is not statistically significant.

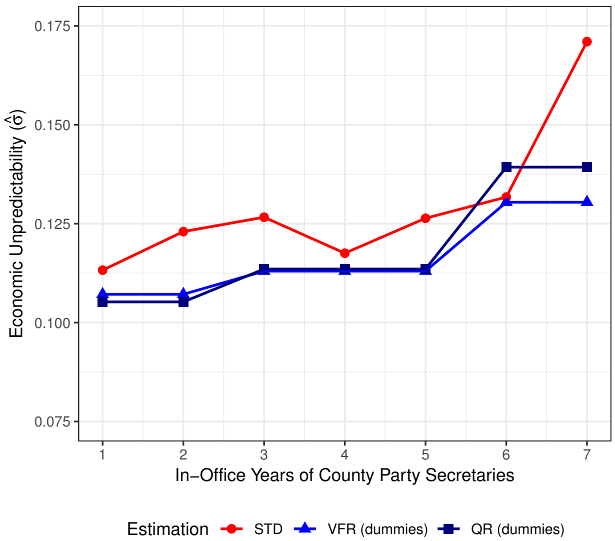
More straightforwardly, Figure 5.2 plots the predicted $\hat{\sigma}$ against in-office years. The $\hat{\sigma}$ is estimated when in-office time enters the variance function as three categories (i.e., `Short`, `Middle`, or `Long`) in panel (a) of Figure 5.2. In panel (b), $\hat{\sigma}$ is fitted when in-office time enters the variance function as polynomials in models (2) or (4). As a comparison, in each plot, I also included the standard deviation (`STD`) of economic growth for groups of county party secretaries with different in-office time. Again, the patterns are consistent with H1. Also, the estimated $\hat{\sigma}$ is usually smaller than the `STD` measure because the latter measures unconditional variance rather than conditional variance, namely error variance.

Table 5.1: Local Officials' In-Office Time and Economic Unpredictability in Chinese Counties

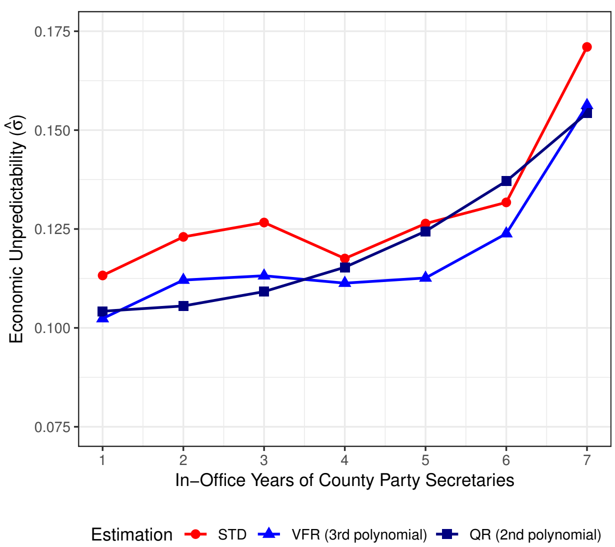
	VFR		QR+SM	
	(1)	(2)	(3)	(4)
In-office Years: (Ref.: 1-2 yr.)				
Middle (3-5 yr.)	0.107		0.152^{***}	
	(0.072)		(0.025)	
Long (> 5 yr.)	0.393^{*}		0.561^{***}	
	(0.213)		(0.073)	
In-office Years (Cont.)		0.626^{**}		-0.038
		(0.294)		(0.034)
In-office Years^2		-0.191^{**}		0.021^{***}
		(0.096)		(0.005)
In-office Years^3		0.018^{**}		
		(0.009)		
Constant	-4.467^{***}	-5.012^{***}	-4.503^{***}	-4.506^{***}
	(0.049)	(0.256)	(0.017)	(0.087)
Observations	4,660	4,660	4,660	4,660
Log Likelihood	18,491.750	18,496.590	16,356.900	16,382.970
Akaike Inf. Crit.	$-36,977.510$	$-36,985.190$	$-32,707.800$	$-32,759.930$
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure 5.2: In-Office Years and Predicted Economic Unpredictability

(a) Unpredictability Fitted with Categorical In-Office Time



(b) Unpredictability Fitted with Polynomial In-Office Time



Note: The two graphs show the predicted error standard deviation ($\hat{\sigma}$ when a county party secretary's in-office time varies. Panel (a) shows the prediction from models (1) and (3) in Table 5.1, where in-office time enters the variance function regression as a categorical measure (i.e., `Short`, `Middle`, and `Long`). Panel (b) shows the estimates from models (2) and (4), where in-office time is treated as a continuous variable and its polynomials are included in regressions. In both graphs, `STD`, the standard deviation of economic growth in groups of county party secretaries with different in-office years, is included for comparison.

CHAPTER 6

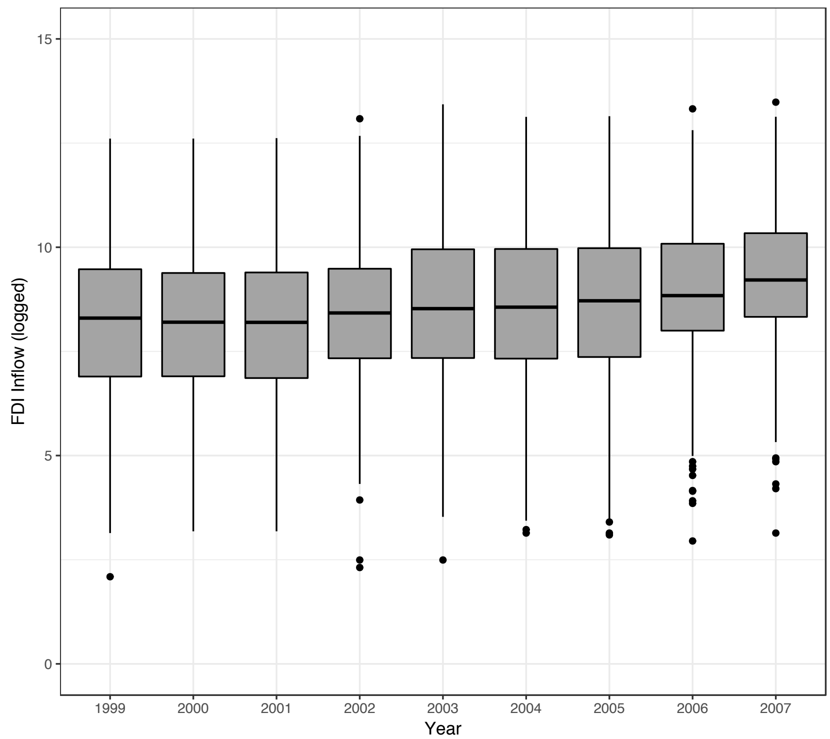
Application II: Economic Uncertainties and FDI Inflows in Chinese Municipalities

6.1 Theory

Theoretical and empirical studies have shown that political and macroeconomic uncertainties affect the risk premium of investment projects and thus the decision of multinationals to locate their investment. Hence, these uncertainties significantly reduce FDI inflows, especially in developing countries (Aizenman 2003; **?**; Busse and Hefeker 2007). To some extent, the political and economic uncertainties can have even larger effects than surely expected "bad" events. This is because the unknown signs of future changes undermine the ability of powerful multinationals to make strategic arrangements to maximize their return from investment.

China became a member of the World Trade Organization (WTO) on December 11, 2001. Afterward, thanks to the membership, China has witnessed a growth of FDI inflows (Figure 6.1). Due to economic decentralization in China, municipalities – administrative units higher than county but lower than province – compete against each other in attracting FDI. The often used tactics include providing cheap land, giving tax breaks, and reducing governmental regulations. Those providing better resources to multinationals usually attract more FDI inflows. However, the multinationals could also consider economic uncertainties in the municipalities and locate their investment to regions with similarly cheap land and similarly few governmental regulations but *less* economic uncertainties. More specifically, all else equal, the post-WTO FDI inflows can be smaller in municipalities where the economic uncertainty was large before China joined the WTO.

Figure 6.1: FDI Inflows into Chinese Municipalities



Hypothesis 2 (H2): The unpredictability of pre-WTO economic growth reduces FDI inflows into Chinese municipalities after China joined the WTO.

6.2 Data

To test H2, I collect municipal annual economic data between 1999 and 2007. Due to missing values, each year has around 260 (out of in total 330) municipalities in the sample. The first three years (1999-2001) are used to construct a measure for economic unpredictability. Then, the municipality-level unpredictability measure will be used as an independent variable to investigate how it affects FDI inflows between 2002 and 2007, right before the global economic crisis.

More specifically, the unpredictability measure is constructed using the QR+SM approach. Thus, I first run quantile regressions ($\tau \in \{0.05, 0.95\}$). The dependent variable is economic growth of municipalities in 2001 while the independent variables include FDI inflows, fiscal expenditure, population size, the number of college graduates (a proxy for skilled workers), government size and province dummies, all of which are lagged by one year. I also include per-capita GDP growth in 1999 and the growth in 2000 at the right side to capture dependence of economic growth. The quantile regression allows me to obtain quantile estimates of 2001 economic growth for each municipality. Then, I fitted a normal distribution for each municipality by minimizing the theoretical quantiles and estimated quantiles. The fitted normal distribution gives $\hat{\sigma}$– a measure for pre-WTO economic uncertainties. Then, I run another set of regressions to estimate how pre-WTO economic uncertainties affect FDI inflows between 2002 and 2007.

For comparison, I also compute an alternative measure for economic unpredictability: the standard deviation (`STD`) of economic growth in each municipality between 1999 and 2001. As will be demonstrated in the following, the `STD` measure yields findings that are entirely opposite to the theoretical prediction.

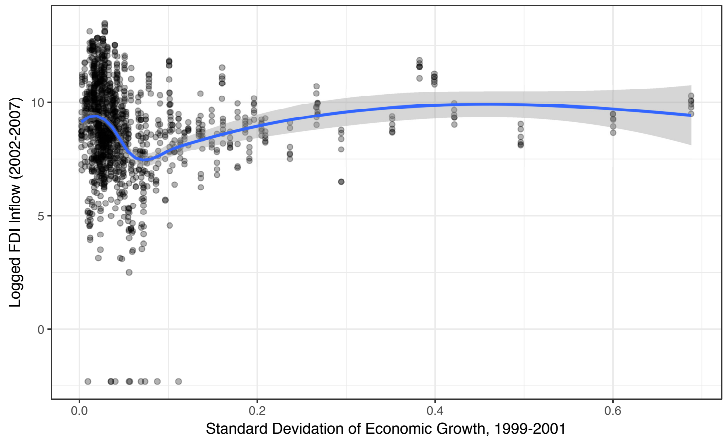
6.3 Results

Before proceeding to more rigorous regression analysis, Figure 6.2 panel (a) shows that the `STD` measure increases rather than decreases FDI inflows, which is opposite to the theoretical prediction. By contrast, in panel (b), the estimated $\hat{\sigma}$ shows the pattern predicted by the theory: higher pre-WTO uncertainties are associated with lower post-WTO FDI inflows.

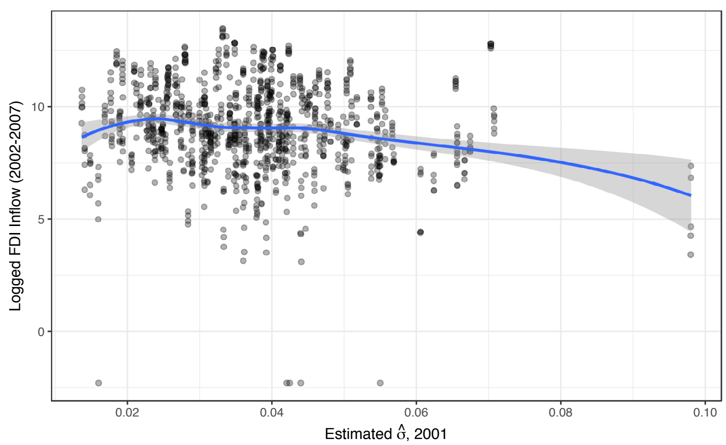
Table 6.1 summarizes the regression estimates. In models (1), $\hat{\sigma}$ has a statistically significant and negative sign, which supports H2. In models (2), I explore whether the negative effect of pre-WTO unpredictability declines as time passes by. This conjecture is reasonable because uncertainties in earlier years may play a less important role in multinational's decision. The positive interaction effect provides supportive evidence. These results are robust to the

Figure 6.2: Economic Unpredictability and FDI Inflows into China

(a) Economic Unpredictability Measured with STD of Growth



(b) Economic Unpredictability Measured with Estimated $\hat{\sigma}$



Note: `STD` is the standard deviation of municipal annual per-capita GDP growth between 1999 and 2001. $\hat{\sigma}$ is predicted error standard deviation for municipal per-capita GDP growth in 2001, which is estimated using the QR+SM approach discussed in the main text.

inclusion of province dummies in model (3) and (4).

In models (5) to (8), I replicate the analysis but use `STD` as an alternative measure for pre-WTO economic unpredictability. The coefficient of `STD` has the "wrong" sign while the interaction is not statistically significant. One possible explanation is that `STD` measures the *unconditional* variance of economic growth, thus an increase may reflect observable improvement of factor endowment rather than increasing unpredictability. Also, `STD` have more extreme values than the $\hat{\sigma}$ measure does (Figure 6.2) because QR+SM indirectly borrows information from others for prediction.

Table 6.1: Estimated Effect of Economic Unpredictability on FDI Inflows in Chinese Municipalities, 2002-2007

	Dependent variable: FDI Inflows (logged)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\hat{\sigma}$	-27.803^{***}	-36.609^{***}	-16.502^{***}	-23.176^{***}				
	(4.158)	(6.125)	(4.020)	(5.648)				
$\hat{\sigma} \times$ (yr. $-$ 2002)		3.298^{*}		2.441^{*}				
		(1.686)		(1.452)				
STD					1.916^{***}	1.639^{**}	1.101^{***}	1.021^{*}
					(0.402)	(0.704)	(0.379)	(0.602)
STD $\times$ (yr. $-$ 2002)						0.109		0.032
						(0.228)		(0.186)
yr. - 2002	-0.081^{***}	-0.204^{***}	-0.068^{***}	-0.160^{***}	-0.092^{***}	-0.099^{***}	-0.072^{***}	-0.074^{***}
	(0.025)	(0.068)	(0.023)	(0.060)	(0.026)	(0.029)	(0.023)	(0.026)
GDP per capita (lagged)	2.555^{***}	2.563^{***}	1.727^{***}	1.739^{***}	2.271^{***}	2.272^{***}	1.435^{***}	1.436^{***}
	(0.118)	(0.117)	(0.130)	(0.130)	(0.117)	(0.117)	(0.121)	(0.122)
Fiscal exp. per capita (lagged)	-0.507^{***}	-0.501^{***}	-0.159	-0.149	-0.397^{***}	-0.398^{***}	-0.023	-0.024
	(0.147)	(0.146)	(0.161)	(0.161)	(0.149)	(0.149)	(0.155)	(0.155)
Population (lagged)	1.325^{***}	1.334^{***}	1.332^{***}	1.343^{***}	1.096^{***}	1.096^{***}	1.130^{***}	1.130^{***}
	(0.080)	(0.080)	(0.083)	(0.083)	(0.071)	(0.071)	(0.067)	(0.067)
Skilled workers (lagged)	-0.070	-0.082	0.101^{*}	0.088	0.197^{***}	0.196^{***}	0.287^{***}	0.287^{***}
	(0.058)	(0.058)	(0.058)	(0.059)	(0.043)	(0.043)	(0.040)	(0.041)
Gov't size (lagged)	-0.413^{**}	-0.406^{**}	0.071	0.069	-0.877^{***}	-0.875^{***}	-0.197	-0.196
	(0.170)	(0.170)	(0.198)	(0.198)	(0.158)	(0.158)	(0.178)	(0.179)
Province dummies			X	X			X	X
Observations	1,104	1,104	1,104	1,104	1,271	1,271	1,271	1,271
R^2	0.629	0.630	0.733	0.734	0.566	0.567	0.717	0.717
Adjusted R^2	0.627	0.628	0.726	0.726	0.564	0.564	0.710	0.710
Note:								*p<0.1; **p<0.05; ***p<0.01

CHAPTER 7

Conclusion

The present study reviews four approaches – the Naive two-step, the Joint ML, the VFR, and the QR+SM – to model and estimate error variance, which can be used for inference or prediction purposes. I compare their performance using simulated data. When both the mean and variance functions are correctly specified, all approaches perform well and are better at predicting individual error variance than using the squared regression residuals. Among the four discussed approaches, quantile regressions plus smoothing (QR+SM) are less efficient than the other three approaches when the variance function can be correctly specified and estimated. The comparison result is reversed among the four when the variance function cannot be correctly estimated using available data. Thus, the variance function regression (VFR) and the joint maximum likelihood (Joint ML) seem to be proper tools to use if the key task is to make inference about the variance function. However, due to possible misspecifications of both mean and variance functions, the QR+SM approach may be used as a robustness check. When the main task is to predict the error variance and the knowledge about causes of changing variance is limited, QR+SM can outperform others.

I also use the VFR and the QR+SM to answer two research questions, which perhaps due to a lack of proper tools, have not been rigorously answered before. Application I shows that Chinese county economic growth becomes more unpredictable after county leaders miss their primary time window for promotion, namely the first five years of tenure. Application II suggests that economic uncertainties in the recent past significantly reduce FDI inflows into Chinese municipalities, but the adverse effects gradually fades away as time passes by.

Due to space limits, I have not discussed another important tool: the Bayesian approach. In fact, previous studies have shown that the VFR can achieve approximately the same performance of Bayesian modeling if the variance function is correctly specified (Western and Bloome 2009). For future studies, one may compare the Bayesian results and those from the discussed approaches when variance function is misspecified. Relatedly, when inference of the variance function is not a priority, one could use machine learning tools to enhance the conditional variance prediction. Finally, because estimating variance usually requires more data than estimating the mean does, all the variance functions discussed in this paper have fewer parameters than the mean functions do. One may explore cases in which the mean and variance have the same functional form. An extra advantage of such exercise is that the variance function may reveal non-linearities that are missed in the mean function.

CHAPTER 8

Appendix: A Simple Model for Application I

In China, economic decentralization has greatly empowered county leaders in local economy management. Thus, their competence (c) and effort (e) largely determine local economic growth. Each county leader's competence is a life-long feature and is thus fixed. However, as effort is costly, county leaders' effort may vary and can be sensitive to their future promotion probability. Let's assume each official's competence is drawn from $c_i \sim N(\bar{c}, \sigma_c^2)$ where $\sigma_c^2 > 0$; the effort that each chooses falls into the interval $e_i \in [\underline{e}, \bar{e}]$.

Additionally, for new officials ($t_i \leqslant 5$), if the local economic growth is equal to or lager than the threshold set by the central government ($\hat{g}$), they will more likely be promoted. Thus, again for simplicity, I assume the promotion probability takes the following form: $P = \mathbf{1}\{c_i + e_i > \hat{g}\} \cdot \mathbf{1}\{t_i \leqslant 5\}$. Thus,

$$P = \begin{cases} 0, \text{ if } t_i > 5 \text{ or } c_i + e_i < \hat{g} \\ 1, \text{ if } t_i \leqslant 5 \text{ and } c_i + e_i \geqslant \hat{g} \end{cases} \tag{8.1}$$

Because any effort is costly, officials with $t_i > 5$ will spend zero effort because any extra effort does not improve their promotion probability. Accordingly, the local economic growth of regions governed by these officials will be $c_i + 0 = c_i$. For officials with $t_i \leqslant 5$, they will spend effort to compensate for their competence so that $c_i + e_i = \hat{g}$.[1] Thus, the local economy will be $\hat{g}$.

[1]Because effort is costly, no one wants to spend more effort than needed. Thus, they will not so much effort e_i such that $c_i + e_i > \hat{g}$.

It is straightforward that the relationship between (conditional) means of economic growth of both groups of officials will depend on the relationship between $E(c_i) = \bar{c}$ and $E(\hat{g}) = \hat{g}$, which lacks a clear empirical prediction because of the unknown $\hat{g}$. However, it is more certain that the (conditional) variance of economic growth is larger in the former group (i.e. $t_i > 5$) than in the latter group (i.e. $t_i \leqslant 5$) because $Var(c_i) = \sigma_c^2 > Var(\hat{g}) = 0$. Thus, a theory that local officials make their governance effort based on their expected promotion chance is not falsifiable by looking at the mean economic growth in the two groups of officials but can be tested by comparing the variance in the two groups (H1).

Bibliography

Acemoglu, Daron, Simon Johnson, James Robinson, and Yunyong Thaicharoen. 2003. "Institutional causes, macroeconomic symptoms: volatility, crises and growth." *Journal of monetary economics* 50 (1): 49–123.

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone. 2019. "Vulnerable growth." *American Economic Review* 109 (4): 1263–89.

Aitkin, Murray. 1987. "Modelling variance heterogeneity in normal regression using GLIM." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 36 (3): 332–339.

Aizenman, Joshua. 2003. "Volatility, employment and the patterns of FDI in emerging markets." *Journal of Development Economics* 72 (2): 585–601.

Blanchard, Olivier, and John Simon. 2001. "The long and large decline in US output volatility." *Brookings papers on economic activity* 2001 (1): 135–174.

Bollerslev, Tim. 1986. "Generalized autoregressive conditional heteroskedasticity." *Journal of econometrics* 31 (3): 307–327.

Bollerslev, Tim, Ray Y Chou, and Kenneth F Kroner. 1992. "ARCH modeling in finance: A review of the theory and empirical evidence." *Journal of econometrics* 52 (1-2): 5–59.

Busse, Matthias, and Carsten Hefeker. 2007. "Political risk, institutions and foreign direct investment." *European journal of political economy* 23 (2): 397–415.

Chandra, Siddharth, and Nita Rudra. 2015. "Reassessing the links between regime type and economic performance: Why some authoritarian regimes show stable growth and others do not." *British Journal of Political Science* 45 (2): 253–285.

Engle, Robert F. 1982. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation." *Econometrica: Journal of the Econometric Society*: 987–1007.

Fails, Matthew. 2014. "Leader Turnover, Volatility, and Political Risk." *Politics & Policy* 42 (3): 369-399.

Giuliano, Paola, Prachi Mishra, and Antonio Spilimbergo. 2013. "Democracy and reforms: evidence from a new dataset." *American Economic Journal: Macroeconomics* 5 (4): 179-204.

Guo, Gang. 2009. "China's Local Political Budget Cycles." *American Journal of Political Science* 53 (3): 621-632.

King, Gary. 1998. *Unifying political methodology: The likelihood theory of statistical inference*. University of Michigan Press.

King, Gary, and Margaret E Roberts. 2015. "How robust standard errors expose methodological problems they do not fix, and what to do about it." *Political Analysis* 23 (2): 159–179.

Miller, Steve, and Richard Startz. 2018. "Feasible generalized least squares using machine learning.". Available at SSRN: `https://ssrn.com/abstract=2966194` or `http://dx.doi.org/10.2139/ssrn.2966194`.

Nakamura, Emi, Dmitriy Sergeyev, and Jon Steinsson. 2017. "Growth-rate and uncertainty shocks in consumption: Cross-country evidence." *American Economic Journal: Macroeconomics* 9 (1): 1–39.

Powell, Jonathan, and Clayton Thyne. 2011. "Global instances of coups from 1950 to 2010: A new dataset." *Journal of Peace Research* 48 (2): 249-259.

Quinn, Dennis P, and John T Woolley. 2001. "Democracy and national economic performance: the preference for stability." *American journal of political science*: 634–657.

Rodrik, Dani. 1991. "Policy uncertainty and private investment in developing countries." *Journal of Development Economics* 36 (2): 229-242.

Sah, Raaj Kumar. 1991. "Fallibility in human organizations and political systems." *Journal of Economic Perspectives* 5 (2): 67–88.

Western, Bruce, and Deirdre Bloome. 2009. "Variance function regressions for studying inequality." *Sociological Methodology* 39 (1): 293–326.

Yang, Benhua. 2008. "Does democracy lower growth volatility? A dynamic panel analysis." *Journal of Macroeconomics* 30 (1): 562–574.

Zhou, Xiang. 2014. "Increasing returns to education, changing labor force structure, and the rise of earnings inequality in urban China, 1996–2010." *Social Forces* 93 (2): 429–455.

Zhou, Xiang. 2019. "Hierarchical Item Response Models for Analyzing Public Opinion." *Political Analysis*: 1–22.