ESSAYS ON NONPARAMETRIC ECONOMETRICS WITH APPLICATIONS TO
CONSUMER AND FINANCIAL ECONOMICS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the
Graduate School of The Ohio State University

By

Yi Zheng, M.A.

The Ohio State University

2008

Dissertation Committee:

Professor Abdoul Sam, Advisor

Professor Matthew Roberts

Professor Stanley Thompson

Approved by

Advisor

Graduate Program in Agricultural,
Environmental and Development
Economics
ABSTRACT

This dissertation is composed of three chapters centering on nonparametric econometrics with applications to consumer demand system analysis, value-at-risk analysis of commodity future prices, and credit risk analysis of home mortgage portfolios.

The first chapter, based on my joint research with Abdoul Sam considers a semiparametric estimation model for a censored consumer demand system with micro data. A common attribute of disaggregated household data is the censoring of commodities. Maximum likelihood and existing two-step estimators of censored demand systems yield biased and inconsistent estimates when the assumed joint distribution of the disturbances is incorrect. This essay proposes a semiparametric estimator that retains the computational advantage of the two-step methods while circumventing their potential distributional misspecification. The key difference between the proposed estimator and existing two-step counterparts is that the parameters of the binary censoring equations are estimated using a distribution-free single-index model. We implement the proposed estimator using household-level data obtained from the Hainan province in China. Horowitz and Härdle (1994)'s specification test lends support to our approach.

The second chapter is an empirical application of a nonparametric estimator of Value-at-Risk on the cattle feeding margin. Value-at-Risk, known as VaR is a common measure of downside market risk associated with an asset or a portfolio of assets. It has been used as a standard tool of predicting potential portfolio losses for twenty years in the financial industry. Recently VaR has gained popularity in agricultural economics literature since the market price risks associated with agricultural commodities are under
evaluation. As initial empirical findings suggest that the performance of any VaR estimation technique is sensitive to the types of data set (portfolio composition) used in developing and evaluating the estimates, agricultural data provides a unique laboratory to further explore VaR and its estimation approaches. This essay as a first attempt applies a distribution-free nonparametric kernel estimator of VaR in an agricultural context, the cattle feeding margin using futures data. The empirical results suggest that the nonparametric VaR estimates enjoy a significant efficiency gain without losing much accuracy compared to the parametric estimates.

The third chapter measures credit risks associated with residential mortgage loans. Credit risk is the primary source of risk for real estate lenders. Recent advancements in the measurement and management of credit risk give lenders with sophisticated internal risk models a significant comparative advantage over other lenders in terms of capital optimization and risk controlling. This manuscript helps understand the determinants of credit risk and acquire perspectives on how it is distributed in the current or future loan portfolios. This essay contributes to the existing volume of literature as it incorporates the nonparametric estimation technique into default risk analysis. The CreditRisk⁺ model is modified and estimated using the consumer side of information. The model identifies the factors determining household default risks and generates a full loan loss distribution at the portfolio level using consumer finance survey data. In the end, portfolio management strategies are discussed.
Dedicated to my family
ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor, Professor Abdoul Sam. The knowledge I learned in his class and the ideas inspired during his insightful talks made me interested in applied econometrics. Professor Sam provides me constant guidance, invaluable advice and warm encouragement through out three years working on my research at The Ohio State University. It is his intellectual support that makes this thesis possible. He also showed valuable patience and prudence in correcting the mistakes in my dissertation. His admirable personality and professional ability are of great benefit to my study and my life in the future.

I also want to thank Professor Matthew Roberts and Professor Stan Thompson. This dissertation benefits greatly from their careful and close review. I appreciate their valuable suggestion and comments on this dissertation and also on conducting decent research.

My special thanks are dedicated to my dear parents and sister. Their trustful and unlimited love supported me in every stage of my life. Without them, none of my achievements would have been possible.

Last but not least, I would like to thank Jing Han, my very best friend, who lends me courage and endless help in all aspects of my life. I truly believe without him my life won’t be so delightful and meaningful as today.
VITA

March 8, 1980 .......................... Born in Inner Mongolia, China

2003 .................................... B.A. in Economics,
    University of Colorado-Denver

2004 .................................... M.A. in Economics,
    The Ohio State University

2008 .................................... M.A.S. in Statistics,
    The Ohio State University

FIELDS OF STUDY

Major Field: Agricultural, Environmental and Development Economics
# TABLE OF CONTENTS

Abstract ........................................................................................................................ ii
Dedication ......................................................................................................................... iv
Acknowledgments ............................................................................................................. v
Vita ...................................................................................................................................... vi
List of Tables ................................................................................................................... ix
List of Figures .................................................................................................................. x

Chapters:

1. Semiparametric Estimation of Consumer Demand Systems with Micro Data 1
   1.1 Introduction: Literature Review ................................................................. 1
   1.2 A Semiparametric Estimation of Censored Demand System ................. 4
   1.3 Empirical Application ................................................................................. 7
       1.3.1 Estimation of First Stage Equations .................................................. 9
       1.3.2 Demand Elasticities .......................................................................... 16
   1.4 Conclusion ....................................................................................................... 23
   1.5 Appendix: Horowitz and Hardle Test ......................................................... 23
   1.6 References ....................................................................................................... 26

2. Nonparametric Inference of Value at Risk for Commodity Future Prices .......... 30
   2.1 Introduction ..................................................................................................... 30
   2.2 Value at Risk .................................................................................................. 32
       2.2.1 Definition ............................................................................................ 32
       2.2.2 Methods of Estimation ...................................................................... 33
   2.3 Nonparametric Estimation of VaR ................................................................. 36
       2.3.1 Theoretical Construct of the Estimator ................................................ 36
       2.3.2 Statistical Properties .......................................................................... 37
       2.3.3 Implementation .................................................................................... 36
   2.4 Application to the Cattle Feeding Margin ..................................................... 40
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1: Variable Definitions and Sample Statistics</td>
<td>8</td>
</tr>
<tr>
<td>Table 1.2: Estimates for the Sample Selection Model</td>
<td>10</td>
</tr>
<tr>
<td>Table 1.3: Horowitz and Hardle Test Results</td>
<td>11</td>
</tr>
<tr>
<td>Table 1.4: AIDS Parameter Estimates</td>
<td>18</td>
</tr>
<tr>
<td>Table 1.5: Semiparametric Elasticity Estimates</td>
<td>19</td>
</tr>
<tr>
<td>Table 1.6: Parametric Elasticity Estimates</td>
<td>22</td>
</tr>
<tr>
<td>Table 2.1: Parametric Estimates of the GARCH (1,1) Model</td>
<td>46</td>
</tr>
<tr>
<td>Table 2.2: Daily VaRs for the Four Returns Series and for Different Confidence Levels (90%, 95%, 99%)</td>
<td>50</td>
</tr>
<tr>
<td>Table 3.1: Variable Definitions and Sample Statistics</td>
<td>64</td>
</tr>
<tr>
<td>Table 3.2: Parameter Estimates of Default Rate Models</td>
<td>70</td>
</tr>
<tr>
<td>Table 3.3: Loss Given Default Ratings and Loss Rates (percent)</td>
<td>75</td>
</tr>
<tr>
<td>Table 3.4: Loan Loss Distribution Summary</td>
<td>76</td>
</tr>
<tr>
<td>Table 3.5: Economic Capital at Various Confidence Levels</td>
<td>77</td>
</tr>
<tr>
<td>Table 3.6: Stress Testing at the 99.97\textsuperscript{th} Percentile</td>
<td>78</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>Figure 1.1 Estimates of the Cumulative Distribution Function for Beef</td>
<td>14</td>
</tr>
<tr>
<td>Figure 1.2 Estimates of the Probability Density Function for Beef</td>
<td>14</td>
</tr>
<tr>
<td>Figure 1.3 Estimates of the Cumulative Distribution Function for Fish</td>
<td>15</td>
</tr>
<tr>
<td>Figure 1.4 Estimates of the Probability Density Function for Fish</td>
<td>15</td>
</tr>
<tr>
<td>Figure 2.1 Graphical Demonstration of Value at Risk</td>
<td>33</td>
</tr>
<tr>
<td>Figure 2.2.a CBOT May Corn Daily Returns and Its Autocorrelation Function</td>
<td>44</td>
</tr>
<tr>
<td>Figure 2.2.b CME May Feeder Cattle Daily Returns and Its Autocorrelation Function</td>
<td>44</td>
</tr>
<tr>
<td>Figure 2.2.c CME October Live Cattle Daily Returns and Its Autocorrelation Function</td>
<td>44</td>
</tr>
<tr>
<td>Figure 2.2.d The Cattle-feeding Margin Return Series and Its Autocorrelation Function</td>
<td>45</td>
</tr>
<tr>
<td>Figure 2.3.a Kernel Density Estimates for Corn Returns</td>
<td>47</td>
</tr>
<tr>
<td>Figure 2.3.b Kernel Density Estimates for Feeder Cattle Returns</td>
<td>47</td>
</tr>
<tr>
<td>Figure 2.3.c Kernel Density Estimates for Live Cattle Returns</td>
<td>48</td>
</tr>
<tr>
<td>Figure 2.3.d Kernel Density Estimates for Margins</td>
<td>48</td>
</tr>
<tr>
<td>Figure 3.1 Loan Loss Characteristics</td>
<td>58</td>
</tr>
<tr>
<td>Figure 3.2 The CreditRisk+ Model Structure</td>
<td>61</td>
</tr>
</tbody>
</table>
Figure 3.3 Estimates of the Cumulative Distribution Function for Default Rate ........................................... 71

Figure 3.4 Estimates of the Probability Density Function for Default Rate ........ 71
CHAPTER I

SEMIPARAMETRIC ESTIMATION OF CONSUMER DEMAND SYSTEMS
WITH MICRO DATA

1.1 Introduction

A common attribute of disaggregated household data is the censoring of commodities stemming from corner optima for demand functions or the infrequency of purchase. Ignoring the censoring leads to biased and inconsistent parameter estimates of the demand system, and throwing away limit observations is inefficient and could also lead to inconsistent results if the reduced observations do not constitute a random sample (Lee and Pitt, 1987). These concerns, along with the increased reliance on micro data to estimate consumer demand equations have stimulated a growing body of theoretical and empirical research on the econometric treatment of censored demand equations. Wales and Woodland (1983), Lee and Pitt (1986), and Chiang and Lee (1992) among others, have proposed maximum likelihood models which account for the presence of zero consumption in a system of demand equations. While the maximum likelihood approach has the advantage of being consistent with the theory of consumer choice, consistency of parameter estimates is contingent on the correctness of the assumed joint distribution function. Furthermore, the practical potential of the maximum likelihood approach is hampered by its computational cost when censoring occurs for
several commodities, as it requires the evaluation of multidimensional integrals.\footnote{We note however, that recent developments in the simulation of multivariate probability density functions (Hajivassiliou (1993), (Kotz et al. (2000)) have renewed interest in the maximum likelihood approach; see Yen and al (2003), Yen (2005), and Dong et al. (2004) for applications.}

Shonkwiler and Yen (1999)(henceforth SY) proposed an estimator in the same realm as Heckman’s two-step sample selection approach (1979). While inefficient, the SY estimator represents a computationally expeditious alternative to maximum likelihood estimators. The SY estimator relies on a stochastic binary censoring rule to explore the determinants of commodity purchase by a household during the survey period. In the first step, probit regressions are used to recover estimates of the parameters of the binary censoring equations. In the second step, each censored demand equation is augmented with an inverse-Mills ratio—constructed with the probit estimates—and then weighted by the normal cumulative distribution function to account for both the censoring and the omitted variable bias. The two-step model is easy to implement and thus has gained significant attention in applied work. However, there are two sources of inconsistency of the SY estimator that merit attention. First, the SY estimator is derived under the assumption that the error processes in the binary censoring and the demand equations for each commodity follow a bivariate normal distribution to recover consistent estimates of the demand system. As such the estimator is prone to distributional misspecification. Specifically, when the underlying distribution between the error processes is joint normal then the SY estimator yields unbiased and consistent estimates. On the other hand, if the wrong joint distribution is assumed then the parameter estimates are biased and inconsistent (Schafgans(2004), Martins (2001)). Second, the SY two-step model is derived with
the assumption that the disturbance terms in the censoring equations are homoscedastic, which is an untenable assumption when dealing with cross-sectional micro data. Consequently, the SY estimator as well as the maximum likelihood estimators may yield erroneous estimates with potentially significant economic implications.

Drawing from theoretical advances in the nonparametric econometrics literature, this paper proposes a two-step semiparametric approach for the estimation of a censored demand system that is similar in spirit to the SY estimator but is exempt from distributional misspecification and accommodates a certain form of heteroscedasticity. The key difference between our proposed and the SY estimator lies in the estimation of the first step parameters. Our approach uses Klein and Spady’s semiparametric single-index model (1993) instead of a probit. The advantage of the Klein and Spady estimator is that it generates consistent and efficient estimates without relying on distributional assumptions and accommodates heteroscedasticity of a certain form in the error term. The Klein and Spady estimator (hereafter KS) has been applied in several empirical studies (e.g., Newey, Powell, and Walker (1990), Newey (1991), Martins (2001), Ker and Tolga (2007)).

The paper is organized as follows. In section 2 we construct our proposed semiparametric model for censored demand system estimation. Section 3 implements the proposed estimator using a cross-sectional data set of 1,237 households from the Hainan province in China. Concluding remarks are presented in Section 4.
1.2 A Semiparametric Estimator of Censored Demand Systems

Let $Y_{ij}$ and $d_{ij}$ be two response variables, $X_{ij}$ and $W_{ij}$ two vectors of regressors such that

\begin{align*}
Y_{ij} &= d_{ij} \ast (g(X_{ij}, \beta_i) + \epsilon_{ij}) \\
d_{ij} &= I(W_{ij}^t \gamma_i + v_{ij} > 0), \quad \text{for } i = 1, 2, ..., n; \quad j = 1, 2, ..., J.
\end{align*}

$I(\omega)$ denotes an indicator function of event $\omega$, $\beta_i$ and $\gamma_i$ are the model parameters, and $\epsilon_{ij}$ and $v_{ij}$ are zero-mean and finite variance error processes. The subscript $i$ indexes commodities and the subscript $j$ indexes observations. We concern ourselves with the estimation of the conditional mean:

\begin{equation}
E(Y_{ij}|X_{ij}, W_{ij}) = E(Y_{ij}|X_{ij}, W_{ij}; d_{ij} = 1) \cdot P(\text{prob}(d_{ij} = 1)
\end{equation}

Let $F_i(W_{ij}^t \gamma_i)$ denote the unknown cumulative distribution function of the error term $v_{ij}$. Then, given our conceptual framework above, the system of regression equations of interest is:

\begin{align*}
Y_{ij} &= (g(X_{ij}, \beta_i) + \lambda_i(W_{ij}^t \gamma_i)) \cdot F_i(W_{ij}^t \gamma_i) + \eta_{ij}, \quad i = 1, 2, ..., n; \quad j = 1, 2, ..., J \quad (1)
\end{align*}

where $\lambda_i(W_{ij}^t \gamma_i) = E(\epsilon_{ij}|v_{ij} > -W_{ij}^t \gamma_i)$, $\eta_{ij} = Y_{ij} - E(Y_{ij}|X_{ij}, W_{ij})$. The functional form of $\lambda_i(\cdot)$ is mathematically tractable. If for example we assume that the joint distribution of the error terms $\epsilon_{ij}, v_{ij}$ is the bivariate normal with covariance parameter $\theta_i$ (as in SY), then the corresponding censored demand system is

\begin{align*}
Y_{ij} &= \left( g(X_{ij}, \beta_i) + \theta_i \frac{\phi(W_{ij}^t \gamma_i)}{\Phi(W_{ij}^t \gamma_i)} \right) \Phi(W_{ij}^t \gamma_i) + \eta_{ij}, \quad i = 1, 2, ..., n; \quad j = 1, 2, ..., J \quad (2)
\end{align*}

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote respectively the standard normal cumulative distribution and probability density functions. SY propose that the $\beta_i$’s in (3) be estimated in
two steps. First, estimate $\gamma_i$ by Probit to obtain $\tilde{\gamma}_i$; then estimate the $J$ equations jointly as a system of seemingly unrelated regressions (SUR) after substituting $\tilde{\gamma}_i$ for $\gamma_i$.

There are two sources of misspecification in the SY estimator that we focus on in this paper. First, consistency of the SY estimator hinges in part on the correctness of the assumed distribution of the disturbances (Schafgans (2004), Martins (2001)). Second, the SY estimator is inconsistent if the error term in the censoring equations, $v_{ij}$, is heteroscedastic (Poirier and Ruud (1983), Donald (1995)). However the homoscedasticity assumption is likely to be violated when disaggregated household data is used to estimate the demand equations. Thus, the SY estimator, while computationally convenient has the potential to generate erroneous elasticity estimates.

Our aim in this paper is to propose an estimator which retains the computational advantage of the two-step estimator but circumvents the two aforementioned sources of inconsistency. To achieve this aim, we estimate the parameters of the first step using the KS semiparametric single-index model. The KS estimator does not make any assumption about the distribution of the error term in the binary selection equation, instead it estimates the distribution function nonparametrically using the Kernel method. However, it assumes a linear index function to circumvent the curse of dimensionality which plagues nonparametric techniques. Briefly, the KS estimator of $\gamma_i$ is obtained by maximizing the quasi-likelihood function

$$\ell(\gamma_i) = \sum_{j=1}^{J} \left( d_{ij} \log(\tilde{F}_i(W_{ij}' \gamma_i)) + (1 - d_{ij}) \log(1 - \tilde{F}_i(W_{ij}' \gamma_i)) \right)$$

with respect to $\gamma_i$, where

$$\tilde{F}_i(v_{is}) = \frac{\sum_{l=1}^{J} d_{il} K_h(v_{il} - v_{is})}{\sum_{l=1}^{J} K_h(v_{il} - v_{is})}, \quad v_{is} = W_{is}' \gamma_i, \quad K_h(u) = \frac{1}{h} K(u/h)$$
and $h$ is a non-stochastic smoothing parameter satisfying the condition $J^{-1/16} < h < J^{-1/8}$. As for the probit, a location-scale normalization is needed to ensure identification of the parameter vector. For the probit, the location-scale normalization requires setting the first and second moments of the error term to zero and 1 respectively. For the KS estimator, the location-scale normalization is imposed by constraining the intercept to zero and one of the coefficients on continuous regressors to a constant. KS show that the resulting estimator, $\hat{\gamma}_i$, is both consistent and achieves the semiparametric efficient bound under certain regularity conditions. Additionally and unlike the probit, the KS estimator is consistent even if the errors are heteroscedastic provided that the heteroscedasticity depends on the matrix $W_i$ of regressors only via the index or is of a known general form. This feature of the KS estimator is particularly attractive in empirical settings such as the estimation of censored demand systems where cross-sectional household data is used.

Our two-step approach to estimate the demand system (2) proceeds as follows. First, we obtain the estimates of $\gamma_i$ and the link function $F_i(v_{ij})$ using the KS method for each censored equation. Second, $\hat{\gamma}_i$ and $\hat{F}_i(\hat{v}_{ij})$ are substituted for $\gamma_i$ and $F_i(v_{ij})$ respectively in (2). Following Newey (1991) and Martins (2001), $\lambda_i(\hat{v}_i)$ is approximated with a series based on orthogonal polynomials of the first-stage index, i.e. $\lambda_i(\hat{v}_i) \simeq \sum_{l=1}^{L} \alpha_l \hat{v}_i^{l-1}$ where the $\alpha_l$’s are unknown coefficients to be estimated. Hence, the second step consists of estimating the following system of nonlinear equations:

$$Y_i = \left( g(X_i, \beta_i) + \sum_{l=1}^{L} \alpha_l \hat{v}_i^{l-1} \right) \hat{F}_i(W_i^{t} \hat{\gamma}_i) + \xi_i, \quad i = 1, 2, \ldots, n.$$  

The system of equations (4) can be consistently estimated as a system of unrelated regressions. However, Murphy and Topel (1985) show that the resulting standard
errors are incorrect since they do not account for the additional variability introduced by the two-step nature of the estimation process. The inconsistency of the SUR standard errors can be overcome by either adjusting the variance as in Murphy and Topel or bootstrapping the parameter estimates (Green et al. (1987)). Since the variance adjustment procedure is tedious to implement for a system of equations, we chose to bootstrap our estimates in the ensuing empirical application.

1.3 Empirical Application

In this section we apply our proposed econometric model using a household survey carried out by the Chinese National Bureau of Statistics in urban areas of the Hainan Province in 2003. The survey contains information on quantities and prices of various food products purchased by a sample of 1,237 urban households along with socioeconomic characteristics of the households. As in Yen & Lin (2006), we limit our empirical study to the demand for meat products: beef, pork, fish, and poultry. During the survey, pork and poultry are consumed by nearly all (over 99%) households, fish is consumed by 93.5% of households, and only 50.8% of households consumed beef. From the reported expenditure and quantity of each meat product consumed, price was derived as the unit value. In addition to income and prices of the four meats, we also include in our model three socioeconomic variables which are the number of wage earners in a household (NOWE), educational level of the head of household (EDUC), and the size of the household (HSIZE). Definitions of variables and sample descriptive statistics are presented in Table 1.1. The summary statistics show that on average fish is the least expensive product, which makes sense since Hainan is a costal province, beef is the most expensive, while pork is the most consumed.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantities (Kg. per person per annum)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beef (Consuming households: 50.8% of sample)</td>
<td>2.64</td>
<td>5.60</td>
</tr>
<tr>
<td>Pork</td>
<td>42.90</td>
<td>27.32</td>
</tr>
<tr>
<td>Fish (Consuming households: 93.5% of sample)</td>
<td>11.76</td>
<td>13.33</td>
</tr>
<tr>
<td>Poultry</td>
<td>18.34</td>
<td>17.52</td>
</tr>
<tr>
<td>Expenditures (Yuan per person per annum)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beef (Consuming households: 50.8% of sample)</td>
<td>36.66</td>
<td>77.46</td>
</tr>
<tr>
<td>Pork</td>
<td>461.82</td>
<td>290.69</td>
</tr>
<tr>
<td>Fish (Consuming households: 93.5% of sample)</td>
<td>83.29</td>
<td>97.97</td>
</tr>
<tr>
<td>Poultry</td>
<td>208.67</td>
<td>187.75</td>
</tr>
<tr>
<td>Prices (Yuan/Kg.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beef</td>
<td>14.38</td>
<td>2.404</td>
</tr>
<tr>
<td>Pork</td>
<td>10.83</td>
<td>1.12</td>
</tr>
<tr>
<td>Fish</td>
<td>7.32</td>
<td>1.96</td>
</tr>
<tr>
<td>Poultry</td>
<td>12.16</td>
<td>3.75</td>
</tr>
<tr>
<td>NOWE (number of wage earners)</td>
<td>1.48</td>
<td>0.89</td>
</tr>
<tr>
<td>HSIZE (size of household)</td>
<td>3.05</td>
<td>0.87</td>
</tr>
<tr>
<td>EDUC (educational level of household head)</td>
<td>5.34</td>
<td>1.63</td>
</tr>
</tbody>
</table>


Table 1.1: Variable Definitions and Sample Statistics (Sample Size: 1,237)
1.3.1 Estimation of First Stage Equations

We estimate two binary regression equations, one for beef, and the other for fish by Probit and KS estimators respectively. The dependent variables in these equations are dichotomous variables that take the value 1 if the household makes a purchase and zero otherwise. The design variables are the natural logarithms of all meat prices, the natural logarithm of total household income, and the three socioeconomic variables discussed above. For the KS estimator, the maximization of the quasi-maximum likelihood function is undertaken using the probit estimates as the starting values and with the constraint that bandwidth \( h \in (0.3, 0.41) \) as required. Per the location restriction, the KS estimator does not have an intercept. To satisfy the scale restriction and facilitate comparison of the probit and KS estimates, we set the coefficient on EDUC to be equal its probit estimate (as in Ker and Tolga (2007)). The results of the estimations are displayed in Table 1.2 below.
<table>
<thead>
<tr>
<th>Variables</th>
<th>Probit</th>
<th></th>
<th>Klein-Spady</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beef</td>
<td>Fish</td>
<td>Beef</td>
<td>Fish</td>
</tr>
<tr>
<td>Intercept</td>
<td>-3.740*** 1.053</td>
<td>-0.562 1.739</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Income</td>
<td>0.600*** 0.065</td>
<td>-735*** 0.096</td>
<td>0.163*** 0.015</td>
<td>0.505*** 0.056</td>
</tr>
<tr>
<td>Beef price</td>
<td>-0.836*** 0.200</td>
<td>-768** 0.404</td>
<td>-0.073*** 0.009</td>
<td>-365*** 0.112</td>
</tr>
<tr>
<td>Pork price</td>
<td>0.623* 0.391</td>
<td>-287 0.598</td>
<td>0.037* 0.023</td>
<td>-0.92 0.109</td>
</tr>
<tr>
<td>Fish price</td>
<td>0.675*** 0.175</td>
<td>-379* 0.285</td>
<td>0.113*** 0.011</td>
<td>-0.103 0.110</td>
</tr>
<tr>
<td>Poultry price</td>
<td>-0.261** 0.134</td>
<td>-111 0.199</td>
<td>-0.019*** 0.007</td>
<td>0.041 0.043</td>
</tr>
<tr>
<td>NOWE</td>
<td>-0.025 0.050</td>
<td>0.001 0.083</td>
<td>-0.020*** 0.004</td>
<td>0.059** 0.028</td>
</tr>
<tr>
<td>HSIZE</td>
<td>0.000 0.049</td>
<td>0.090 0.086</td>
<td>-0.036*** 0.004</td>
<td>0.003 0.021</td>
</tr>
<tr>
<td>EDUC</td>
<td>-0.011 0.024</td>
<td>0.073** 0.038</td>
<td>-0.011 n/a</td>
<td>0.073 n/a</td>
</tr>
<tr>
<td>Log(L)</td>
<td>-840.56</td>
<td>-871.20</td>
<td>-849.25</td>
<td>-837.94</td>
</tr>
</tbody>
</table>

Note: Triple(***), double(**), and single(*) asterisks indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 1.2: Estimates for the Sample Selection Model
Table 1.3: Horowitz and Härdle Test Results

First, we note that the coefficient on the log of income is positive and statistically significant at the 1% level in both equations regardless of the estimation method. We also note that for both probit and KS results, prices of all four meats are statistically significant determinants of household’s decision to purchase beef with the expected negative sign on the coefficient of beef price. The probit estimates of the three non-price variables are not significant at any conventional level except for the EDUC variable, which is significant at the 5% level in the fish equation. The positive coefficient on EDUC in the fish equation suggests that the more educated heads of households are more likely to purchase fish perhaps because they are more aware of the nutritional benefits of a diversified diet that includes fish. The KS estimates show that bigger households and those with more wage earners are less likely to consume beef. The fish equation has fewer significant coefficients. Both the probit and KS estimates of the fish equation show that the decision to purchase fish is negatively
related the price of fish, however the coefficient is not statistically significant in the KS results. Additionally, the KS results indicate that the higher the number of wages earners in a household, the more likely that household is to consume fish. Finally, we note that the KS estimates have generally the same sign as the probit estimates but are much smaller and more efficient.

We test whether the assumption of normal errors (probit model) is consistent with our data, using a specification test proposed by Horowitz and Härdle (1994). The test is based on the distance between the probit link, $\Phi(.)$ and the nonparametric estimator of link function, $\hat{F}(.)$. Under the null hypothesis that the link function is specified correctly as a standard normal cumulative distribution function, the test statistic is

$$HH = \sqrt{h} \sum_{j=1}^{J} u(W_{ij}' \gamma_i) \left( d_{ij} - \Phi(W_{ij}' \gamma_i) \right) \left( \hat{F}_{i}(W_{ij}' \gamma_i) - \Phi(W_{ij}' \gamma_i) \right)$$

where $\gamma_i$ is the probit estimator of $\gamma_i$, and $u(W_{ij}' \gamma_i)$ is a function that downweights extreme index values. The test is asymptotically normally distributed with mean zero and variance $\sigma_H^2$.\(^2\)

We implemented the test using five values of the smoothing parameter $h = .2, .3, .4, .5,$ and $h = .6$. The results of the specification test are displayed in Table 1.3. To investigate the effect of the weighting function $u(.)$ on the test results, we

\(^2\)The variance of the $HH$ test statistic $\sigma_H^2 = \frac{2C_k}{N} = \sum_{j=1}^{J} \left( \frac{u(W_{ij}' \gamma_i) \hat{F}_{i}(W_{ij}' \gamma_i)[1-\hat{F}_{i}(W_{ij}' \gamma_i)]^2}{P_{h}(W_{ij}' \gamma_i)} \right)$ and $C_k = \int_{-\infty}^{\infty} K(x)^2dx = \frac{1}{2\pi \sigma^2}$ if the standard normal is used as the Kernel function. $\hat{F}_{h}(.)$ is the nonparametric (Nadaraya-Watson) estimator of the probability density function. In practice, since the the nonparametric estimator is biased, $\hat{F}(.)$ is replaced with a linear combination of two nonparametric estimators of the link function estimated with different bandwidths $h$ and $s$ respectively, $\hat{F}(.) = \frac{\hat{F}_{h}(.) - (h/s)^2 \hat{F}_{s}(.)}{1 - (h/s)^2}$, $h = cn^{-1/5}$ and $s = cn^{-\zeta/5}$, $c > 0, 0 < \zeta < 1$ (Bierens, 1987).
consider two cases. First \( u(.) \) is assumed to be 1 for all values within the 10\( ^{th} \) and 90\( ^{th} \) percentile of the fitted (probit) index and 0 otherwise. Second, \( u(.) \) is assumed to equal 1 for all values within the 5\( ^{th} \) and 95\( ^{th} \) percentile of the fitted index and zero otherwise. For the beef equation, the test results show that for higher values of \( h \), the null hypothesis of normal errors is rejected at the 10\% level or better. This is consistent with the findings in Martins (2001) and corroborates Proenca’s (1993) simulation study that shows using higher values of the smoothing parameter may increase the power of the test. For the fish equation, the test statistic clearly rejects of the null of normality for all values of \( h \) with one exception \( (h = .3 \) and for the second weighting function). Additionally, visual inspection of the plots of the estimated cumulative distribution and probability density functions (see figures 1.1 and 1.2 below) show significant differences between the probit and nonparametric estimates. For example, the nonparametric estimate of the probability density function for beef is bimodal. We conclude that the normality assumption of the probit model is inappropriate for our data, a conclusion that lends support to our distribution-free approach.
Figure 1.1a. Estimates of the Cumulative Distribution Function for Beef

Figure 1.1b. Estimates of the Probability Density Function for Beef
Figure 1.2a. Estimates of the Cumulative Distribution Function for Fish

Figure 1.2b. Estimates of the Probability Density Function for Fish
1.3.2 Demand Elasticities

In the second stage, the Almost Ideal Demand System (AIDS) (Deaton and Muellbauer, 1980) was chosen to model our partially censored demand system for meats. Let \( w_i \) represent the expenditure share of commodity \( i \), \( x \) the household’s expenditure on meats, and \( p_i \) the price of \( i^{th} \) commodity, then the functional form of the AIDS is:

\[
\begin{align*}
    w_i &= \alpha_i + \beta_i \left( \log \frac{x}{P} \right) + \sum_{k=1}^{n} \gamma_{ik} \log p_k, \quad \text{for } i = 1, 2, \ldots, n; \quad k = 1, 2, \ldots, J \\
    \log P &= \alpha_0 + \sum_{i=1}^{n} \alpha_i \log p_i + 0.5 \sum_{l=1}^{L} \sum_{k=1}^{n} \gamma_{lk} \log p_l \log p_k
\end{align*}
\]

with the theoretical restrictions:\(^3\)

\[
\sum_{i=1}^{n} \alpha_i = 1, \quad \sum_{i=1}^{n} \beta_i = 0, \quad \sum_{i=1}^{n} \gamma_{ik} = 0 \quad \text{(adding-up)}
\]

\[
\sum_{k=1}^{n} \gamma_{ik} = 0 \quad \text{(homogeneity)}
\]

\[
\gamma_{ik} = \gamma_{ki} \quad \text{(symmetry)}
\]

Therefore our proposed methodology consists of estimating the following partially censored demand system (see equation 4):

\[
w_i = \begin{cases} 
    \alpha_i + \beta_i \left( \log \frac{x}{P} \right) + \sum_{k=1}^{n} \gamma_{ik} \log p_k, & \text{if } i = \text{pork, poultry} \\
    \left( \alpha_i + \beta_i \left( \log \frac{x}{P} \right) + \sum_{k=1}^{n} \gamma_{ik} \log p_k + \sum_{l=1}^{L} \alpha_l (W_{i'}^{'J_i})^{l-1} \right) \widehat{F_i}(W_{i'}^{'J_i}), & \text{if } i = \text{beef, fish}
\end{cases}
\]

The number of terms in the approximation of \( \lambda(.) \), \( L = 3 \) is found by minimizing the mean squared error for each of the censored equations.

\(^3\)It is well-known that the adding-up constraint cannot be imposed in our censored demand model with parametric restrictions. Following Pudney (1989) and Yen, Lin, and Smallwood (2003) we impose adding-up of the observed shares by treating one of commodities, poultry, simply as a residual commodity. The predicted shares for poultry are obtained via the following identity:

\[
w_{\text{poultry}} = 1 - w_{\text{pork}} - w_{\text{beef}} - w_{\text{fish}}.
\]
We estimate (10) as a system of seemingly unrelated regression (SUR) using a canned procedure in the SAS statistical software. One econometric issue with the proposed estimator is the difficulty of identifying the intercept of the second-stage regression (the $\alpha_i$'s) separately from the intercept in the series approximation of the selectivity variable $\lambda_i(W_i'\hat{\gamma}_i)$. Andrews and Schafgans (1998) have proposed an approach to estimate the second-stage intercept consistently. However, since for our purposes the AIDS intercepts ({$\alpha_i$}’s) are not needed separately to compute elasticities, we did not attempt to recover their estimates. We therefore present the parameter estimates of the $\beta$’s and $\gamma$’s of the AIDS model in Table 4 along with their bootstrapped standard errors.\footnote{Because of the two-step estimation procedure in our case, it is well known that the standard errors of parameter estimates need to be adjusted to account for the added variability due to the first step estimation (Murphy and Topel, 1985). We circumvent this issue by bootstrapping our sample. Specifically, we obtained 250 bootstrapped samples from our data; performed our two-step estimation for each sample; obtained the parameter estimates of the AIDS for each bootstrapped sample and constructed standard errors from the empirical distribution of the bootstrapped estimates.}
<table>
<thead>
<tr>
<th></th>
<th>Beef</th>
<th>Pork</th>
<th>Fish</th>
<th>Poultry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beef</td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.017)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pork</td>
<td>0.030**</td>
<td>-0.032</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td></td>
<td>(0.018)</td>
<td>(0.034)</td>
<td></td>
</tr>
<tr>
<td>Fish</td>
<td>0.005</td>
<td>-0.002</td>
<td>0.037***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.013)</td>
<td>(0.015)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>Poultry</td>
<td>-0.039**</td>
<td>0.007</td>
<td>-0.040***</td>
<td>0.072***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.015)</td>
<td>(0.021)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Beta</td>
<td>0.004</td>
<td>-0.017*</td>
<td>0.012**</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.009)</td>
<td>(0.010)</td>
<td>(0.006)</td>
</tr>
</tbody>
</table>

Bootstrapping standard errors are in parenthesis. Triple(***), double(**), and single(*) asterisks indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 1.4: AIDS Parameter Estimates
<table>
<thead>
<tr>
<th>Product</th>
<th>Price of</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beef</td>
<td>Pork</td>
</tr>
<tr>
<td><strong>Uncompensated elasticities</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beef</td>
<td>-0.878***</td>
<td>0.308*</td>
</tr>
<tr>
<td></td>
<td>(0.164)</td>
<td>(0.209)</td>
</tr>
<tr>
<td>Pork</td>
<td>0.048**</td>
<td>-1.031***</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.061)</td>
</tr>
<tr>
<td>Fish</td>
<td>0.060</td>
<td>-0.050</td>
</tr>
<tr>
<td></td>
<td>(0.093)</td>
<td>(0.157)</td>
</tr>
<tr>
<td>Poultry</td>
<td>-0.174***</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.079)</td>
</tr>
<tr>
<td><strong>Compensated elasticities</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beef</td>
<td>-0.833***</td>
<td>0.912***</td>
</tr>
<tr>
<td></td>
<td>(0.162)</td>
<td>(0.194)</td>
</tr>
<tr>
<td>Pork</td>
<td>0.091***</td>
<td>-0.456***</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.060)</td>
</tr>
<tr>
<td>Fish</td>
<td>0.107</td>
<td>0.584***</td>
</tr>
<tr>
<td></td>
<td>(0.092)</td>
<td>(0.151)</td>
</tr>
<tr>
<td>Poultry</td>
<td>-0.128***</td>
<td>0.604***</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.073)</td>
</tr>
</tbody>
</table>

Table 1.5: Semiparametric Elasticity Estimates
SUR yields consistent parameter estimates which we then use to compute the income and Marshallian elasticities. For the censored meats, beef and fish, the elasticities are calculated by differentiating the unconditional mean of the expenditure share, i.e. the second step equation. Table 1.5 displays the elasticities estimated with our semiparametric two-step method. Following Green et al. (1987), bootstrapping standard errors are presented. All uncompensated own-price elasticities are negative and significant at the 1% level. Beef and pork have slightly above unity own-price elasticities, and fish and poultry have below unity own-price elasticities. The uncompensated cross-price elasticities between beef and pork and pork and poultry are positive implying gross substitutability between beef and pork and pork and poultry. On the other hand, the uncompensated cross-price elasticities between beef and fish, beef and poultry, and pork and fish are all negative suggesting gross complementarity among these meats. All expenditure elasticities are positive and significant at the 1% level. Furthermore, all four expenditure elasticities are close to unity indicating that the four meats considered here are normal goods. Similar to the uncompensated cross-price effects, the compensated elasticities indicate net substitution between beef and pork and pork and poultry, and also net complementarity between beef and poultry. Pork and fish have a negative uncompensated cross-price effect but positive compensated cross-price effect. All compensated own-price elasticities are negative and significant at the 1% level, and also smaller in absolute values than their uncompensated counterparts due to the positive expenditure elasticities.

Parametric elasticity estimates are presented in Table 1.6 for comparison purpose. Similar to semiparametric estimates, the total expenditure elasticities for all four meat are positive and significant at the 1% level. The expenditure elasticities for pork, fish,
and poultry are very close to the unit value, which indicates that they are normal goods. However, the parametric expenditure elasticity estimate for beef is a little more below the unit value, which demonstrates the different implication about beef consumption by the semiparametric estimate (1.019) and the parametric estimate (0.681). For the uncompensated price elasticities, most of the parametric estimates are very similar to their semiparametric counterparts except for the uncompensated cross-price elasticity between fish and beef. The semiparametric elasticity estimate of fish with respect to the price of beef is positive and insignificant but the parametric estimate is on the other hand negative and significant at the 1% level. In a another word, the semiparametric estimate suggest that beef is a substitutive meat for fish but the parametric estimate indicates that beef is a complementary food for fish. For the compensated price elasticities between the four meat, the semiparametric and the parametric estimates lead to almost the same results.
<table>
<thead>
<tr>
<th>Product</th>
<th>Beef</th>
<th>Pork</th>
<th>Fish</th>
<th>Poultry</th>
<th>Expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Uncompensated elasticities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beef</td>
<td>-0.729***</td>
<td>0.463**</td>
<td>-0.058</td>
<td>-0.370***</td>
<td>0.681***</td>
</tr>
<tr>
<td></td>
<td>(0.098)</td>
<td>(0.212)</td>
<td>(0.165)</td>
<td>(0.131)</td>
<td>(0.095)</td>
</tr>
<tr>
<td>Pork</td>
<td>0.041</td>
<td>-1.017***</td>
<td>0.019</td>
<td>-0.013</td>
<td>0.970***</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.060)</td>
<td>(0.024)</td>
<td>(0.034)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Fish</td>
<td>-0.391***</td>
<td>-0.087</td>
<td>-0.623***</td>
<td>-0.331***</td>
<td>0.955***</td>
</tr>
<tr>
<td></td>
<td>(0.130)</td>
<td>(0.132)</td>
<td>(0.096)</td>
<td>(0.073)</td>
<td>(0.051)</td>
</tr>
<tr>
<td>Poultry</td>
<td>-0.154***</td>
<td>-0.080</td>
<td>-0.110***</td>
<td>-0.711***</td>
<td>1.056***</td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.076)</td>
<td>(0.035)</td>
<td>(0.068)</td>
<td>(0.048)</td>
</tr>
<tr>
<td><strong>Compensated elasticities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beef</td>
<td>-0.699***</td>
<td>0.867***</td>
<td>0.010</td>
<td>-0.192</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.097)</td>
<td>(0.208)</td>
<td>(0.164)</td>
<td>(0.131)</td>
<td></td>
</tr>
<tr>
<td>Pork</td>
<td>0.084***</td>
<td>-0.441***</td>
<td>0.117***</td>
<td>0.241***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.060)</td>
<td>(0.024)</td>
<td>(0.034)</td>
<td></td>
</tr>
<tr>
<td>Fish</td>
<td>-0.349**</td>
<td>0.480***</td>
<td>-0.526***</td>
<td>-0.082</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.131)</td>
<td>(0.128)</td>
<td>(0.096)</td>
<td>(0.074)</td>
<td></td>
</tr>
<tr>
<td>Poultry</td>
<td>-0.108**</td>
<td>0.546***</td>
<td>-0.004</td>
<td>-0.435***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.077)</td>
<td>(0.034)</td>
<td>(0.069)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.6: Parametric Elasticity Estimates
1.4 Conclusion

The increased reliance on household data to estimate demand systems has generated a body of research on the econometric treatment of zero consumption. For a large system with several censored equations the SY two-step estimator is a computationally expeditious alternative to the maximum likelihood approach as it avoids the evaluation of multidimensional integrals. However, the SY estimator generates inconsistent estimates if the wrong joint distribution is assumed or if the errors in the censored equations are heteroscedastic. This paper contributes to the empirical literature of consumer demand by proposing a semiparametric estimator that is free of distributional assumptions and accommodates heteroscedasticity of a certain form in the censored equations.

Our proposed semiparametric estimator is implemented using a household survey data set of meat consumption in the Chinese province of Hainan conducted in 2003. The Horowitz and Härdle (1994) test rejects the null of normal errors for each of the two censored equations (beef and pork) therefore lending support to our semiparametric approach.
Appendix: Horowitz and Härdle (1994)

Horowitz and Härdle (1994) proposed a method for testing the adequacy of a probit (parametric) model of the mean of a random variable $Y$ conditional on a vector of explanatory variables $X$ against a semiparametric alternative that can be used for binary response models. Specifically the procedure is for testing the specification of the model:

$$E(Y|X = x) = f(x, \beta)$$

in another word the functional form of $f(x, \beta)$. $Y$ is the binary response variable taking the value zero or one, $X$ is the a vector of explanatory variables and $\beta$ is a finite vector of parameters. If we assume that the general function $f(x, \beta)$ can be written as a function, $F$, of $v(x, \beta)$, where $F$ and $v$ are known real functions, then

$$E(Y|X = x) = F[v(x, \beta)]$$

where $F$ is the link function. In a binary choice framework, the regression function $F[v(x, \beta)]$ can be written as a single index function as $E(Y|X = x) = P(Y = 1|X = x) \times 1 = \Phi(x'\beta)$. When the model is a single index, $v(x, \beta) = x'\beta$ and the function $F[v(x, \beta)] = F(x'\beta)$, the authors suggest testing the specification of the single-index model according to the hypothesis:

$$H_0 : E(Z|X'\beta = v) = F(v)$$

$$H_1 : E(Z|X'\beta = v) = H(v) \text{ where } H \text{ is an unknown function}$$

When the link function $F$ is a probit one, under the null and some regularity
conditions the test statistic $T$ has the following property

$$T = \frac{T_n}{\sigma_T} = \frac{\sqrt{n} \sum_{i=1}^{n} u(x_i'\hat{\beta})\{Y_i - \Phi(x_i'\hat{\beta})\}\{\hat{F}_i(x_i'\hat{\beta}) - \Phi(x_i'\hat{\beta})\}}{\sqrt{2C_k n \sum_{i=1}^{n} \{u(x_i'\hat{\beta})\hat{F}_i(x_i'\hat{\beta})[1-F_i(x_i'\hat{\beta})]\}^2 / P_h(x_i'\hat{\beta})}} \sim N(0, 1)$$

where $\hat{F}(.)$ is the nonparametric CDF estimator defined by Bierens (1987) as a linear combination of two regression kernels with different bandwidths, say, $h$ and $s$ as following:

$$\hat{F}(.) = \frac{\hat{F}_h(.) - (h/s)^2\hat{F}_s(.)}{1 - (h/s)^2}, \quad h = cn^{-1/5} \text{ and } s = cn^{-\zeta/5}, c > 0, 0 < \zeta < 1$$

$\hat{P}(.)$ is the nonparametric estimator of the probability density function. $\hat{\beta}$ is a vector of probit estimates of parameters. $C_k$ is defined as

$$C_k = \int_{-\infty}^{\infty} K(x)^2 dx = \int_{-\infty}^{\infty} \phi(x)^2 dx = \int_{-\infty}^{\infty} \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \right)^2 dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-x^2} dx = \frac{1}{2\sqrt{\pi}}$$

if the standard normal is chosen as the Kernel function.
Bibliography


CHAPTER II

NONPARAMETRIC INFERENCE OF VALUE-AT-RISK FOR COMMODITY FUTURE PRICES

2.1 Introduction

Value-at-Risk (VaR) determines the maximum loss of the portfolio value that will occur under “normal” market movement over a certain period of time for a particular level of confidence. Losses greater than VaR are suffered only under “unusual” market conditions with a specified small probability. Because of its simplistic construction, VaR has become one of most popular standard quantitative risk reporting methods. It has been primarily designed for the needs of financial institutions especially for measuring risks associated with highly market sensitive assets such as derivatives (Linsmeier and Pearson, 1997). Recently VaR has gained attention in the agricultural economics literature as the agricultural sector faces the market price risks of agricultural commodities. Value-at-Risk was initially introduced for agricultural applications by Boehlje and Lins (1998) and Gloy and Baker (2001). Manfredo and Leuthold (2001) and Odening and Hinrichs (2003) use the VaR approach to investigate the market risks in US cattle feeding and German hog production respectively.

Since being introduced in financial risk management, VaR has been estimated by either parametric or full valuation procedures. Each of these two categories of estimation have been criticized as having serious limitations. Parametric methods are constrained by fundamental assumptions about the underlying return distribu-
tion while full valuation procedures are criticized for not being able to capture the time-varying volatility which is often found in return time series. Initial empirical findings (Mahoney 1996, Hendricks 1996, Jackson, Maude, and Perraudin 1997) suggest that the performance of any VaR estimation technique is sensitive to the types of data set (portfolio composition) used in developing the estimates besides the other two parameters associated with any VaR: the confidence level and the time horizon. Manfredo and Leuthold (2001) propose a variety of conventional parametric models to estimate the VaR in the context of the cattle feeding margin\(^5\), while Odening and Hinrichs (2003) use extreme value theory (also parametric) to estimate VaR in hog production. This paper adopts a recently developed nonparametric kernel estimator of VaR and applies it to US cattle feeding risk management. This essay distinguishes itself from Manfredo and Leuthold (2001)’s work by implementing a nonparametric VaR estimator in the cattle feeding margin instead of a parametric one using futures data\(^6\). This research is important and unique since it provides insights into the performance of nonparametric econometrics for the evaluation of risk measures in many other agricultural contexts besides the cattle feeding margin.

### 2.2 Value-at-Risk

#### 2.2.1 Definition

Briefly stated, Value-at-Risk is a single, summary, statistical measure of possible portfolio losses under “normal” market movement. There are two analogous definitions

\(^5\)The parametric methods Manfredo and Leuthold applied in their study include: 1) a long-run historical average; 2) a 150-week historical moving average; 3) a GARCH (1,1) \(^t\); 4) implied volatilities from options on futures contracts.

\(^6\)Cash prices of all commodities are used in Manfredo and Leuthold’s work.
for VaR. One is defined on the distribution of portfolio values and the other is defined on the distribution of portfolio returns. Jorion (1996, 1997) first defines the VaR on the portfolio value distribution. Specifically, the end of period portfolio value is $W = W_0(1 + R)$ where $W_0$ is initial portfolio value, and $R$ is the portfolio return. Hence the critical end of period portfolio value, denoted as $W^*$ is named VaR where $W^* = W_0(1 + R^*)$ when the worst possible portfolio return $R^*$ occurs with a small probability $p$, which is associated with a predetermined confidence level $1 - p$ (e.g. 95%). Given the predetermined confidence level $1 - p$, these returns should not be encountered more than $p$ percent of the time. Therefore, for a general distribution of portfolio values, Jorion (1996, 1997) defines VaR as:

$$p = \int_{-\infty}^{W^*} g(W) dW = G(W^*)$$

where $g(.)$ and $G(.)$ are probability density and cumulative distribution function of portfolio values respectively.

VaR can also be defined on the distribution of return series. Let $\{Y_t\}_{t=1}^n$ be the market value of an asset over $n$ periods of time, and let $R_t = \log(Y_t/Y_{t-1})$ be the log returns. Suppose $\{R_t\}_{t=1}^n$ is a strictly stationary dependent (or independent) process with a cumulative distribution function $F$. Then given a positive number $p$, which is close to zero, the $1 - p$ level VaR, denoted by $v_p$ is defined as

$$v_p = \inf \{v : F(v) \geq p\}$$

where $v$ is an auxiliary variable.

Figure 2.1 illustrates the concept graphically in the context of two types of definitions. It is simply a way to describe the magnitude of the likely losses on the portfolio. For example, a VaR estimate of twenty thousand dollars at the 95% level
of confidence indicates that portfolio losses should not exceed twenty thousand dollars more than 5% of the time over the given holding period. Therefore the calculation of VaR consists of finding the \( p \)-quantile of the distribution of portfolio dollar values or portfolio returns.

\[ \text{VaR} \]

\[
f(W \text{ or } V)\]

\[
\text{Prob} = p
\]

The PDF for
Def.1: portfolio value \( W \)
Def.2: profit and loss \( V \)

Figure 2.1: Graphical demonstration of Value-at-Risk

### 2.2.2 Methods of Estimation

VaR is generally estimated using two types of methods, i.e. parametric and full-valuation procedures. Parametric procedures determine estimates of volatility under the assumption of normality while full-valuation procedures attempt to model the entire empirical return distribution. Parametric procedures are often referred as variance-covariance methods. Full-valuation procedures usually include historical simulation and Monte Carlo simulation. Each type of estimation procedure shows specific advantages and disadvantages. Detailed discussions about these standard procedures can be found in Jorion (1996, 1997), Dowd (1998), Manfredo and Leuthold.
(1999), and Linsmeier and Pearson (1996, 1997). This paper provides a brief overview and summary of different conventional VaR estimation procedures and further proposes a nonparametric VaR estimator for agricultural return series.

Parametric (also known as variance-covariance or delta-normal) procedures assume the general distribution in equation 2 is a normal distribution, then standard mathematical properties of the normal distribution are used to determine the loss that will be equaled or exceeded $p$ percent of the time, i.e. the Value-at-Risk. For instance, a standard property of the normal distribution is that outcomes less than or equal to 1.645 standard deviation below the mean occur only 5 percent of the time. That is, a 95% VaR is equal to 1.645 times the standard deviation of portfolio returns. In a general formula,

$$ VaR = c \sigma $$

where $c$ denotes the $p$-quantile of the standard normal distribution, and $\sigma$ is the standard deviation of portfolio returns$^7$. Given a predetermined level of confidence, it is apparent that the essential element of parametric VaR estimator (equation 3) is the estimate of portfolio standard deviation $\sigma$, also referred to as portfolio volatility. A main advantage of the parametric approach is its ability to incorporate time-varying volatility often found in financial return series. However the normality assumption about the return series used to construct volatility and correlation estimates is frequently criticized by practitioners. In fact, there is strong evidence that financial as well as agricultural price return distributions are fat tailed (Yang and Brorsen, 1992). Since VaR attempts to explain information in the lower tail of a probability

$^7$For greater details about the derivation of the formulae, see Jorion (1996, 1997), and Manfredo and Leuthold (1998, 2001).
distribution, estimates of VaR can be distorted in the presence of leptokurtosis. To circumvent this problem, many practitioners suggest the use of alternative cumulative distributions such as the Student’s t distribution or Pareto distribution. Nevertheless, a predetermined distribution is a crucial assumption in parametric models and also a plausible cause for model misspecification.

Full-valuation procedures that have been suggested for developing VaR estimates include historical simulation (Mahoney, Butler and Schachter, 1992) and Monte Carlo simulation (Jorion 1997, Linsmeier and Pearson 1996, 1997). Both approaches attempt to model the entire return distribution instead of providing a point estimate of volatility. Historical simulation is the most simplistic of the full-valuation procedures. The basic idea of historical simulation is to expose the portfolio positions to past observations of the risky positions over a given historical period. In other words, the distribution of profits and losses is constructed by taking the current portfolio and subjecting it to the actual changes in the market factors experienced in the past, say $N$ periods. After $N$ hypothetical portfolio values are obtained, they are ranked from the smallest to the largest. The portfolio value that represents the designated risk tolerance level becomes the VaR estimate. This approach has been praised for its few assumption about the underlying distribution of portfolio values. However this procedure does implicitly assume a constant distribution of the market factors and ignore time variation of the variance of the distribution (Jorion 1997). Consequently, VaR estimates based on historical simulation are very sensitive to changes in the data sample (Odening and Hinrichs, 2003).

Monte Carlo simulation shares the same spirit as historical simulation. The procedure consists of generating pseudorandom values of the risky market factors of the
portfolio based on a predetermined data generating process, obtaining the portfolio values with the generated market factors and then ordering the portfolio values from smallest to largest. Same as historical simulation, the VaR estimate is the \( p \)-quantile of the ordered series. Monte Carlo simulation is claimed to be the most flexible VaR estimation technique (Jorion 1997, Linsmeier and Pearson 1996). However, because of the design of this method, it is very prone to specification error especially with complex portfolios. A complex portfolio may contain a large number of correlated risk factors, which makes the choice of data generating process a considerably difficult task.

Considering the limitations of the traditional VaR estimation methods stated above, this manuscript adopted a nonparametric VaR estimator. The nonparametric kernel estimator of VaR considers the entire return distribution thus should belong to the full-valuation procedure category. Greater details about the nonparametric estimation of VaR are presented in the following section.

2.3 Nonparametric Estimator of VaR

2.3.1 Theoretical Construct of the Estimator

The nonparametric kernel estimator of VaR was inspired by Dowd (2001)’s sample quantile estimator commonly used in statistics. The idea of the sample quantile estimator is to replace the theoretical cumulative distribution function \( F \) in equation (2) with the empirical one \( F_n \), where \( F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x) \). As the VaR is estimated in the tail region of the distribution where the amount of data information is usually thin, Gourieroux, Laurent, and Scaillet (2000) utilized kernel smoothing on
the empirical distribution function $F_n$:

$$
\hat{F}_{n,h}(x) = \frac{1}{n} \sum_{i=1}^{n} G\left(\frac{x - X_i}{h}\right)
$$

where $G$ is the smoother to replace the indicator function $I$ in the formulation of $F_n$ and $h$ is a smoothing bandwidth that controls the amount of smoothing in the estimation of $F$. Therefore, a kernel estimator of VaR, denoted by $\hat{\nu}_p$, satisfies

$$
\frac{1}{n} \sum_{i=1}^{n} G\left(\frac{\hat{\nu}_p - X_i}{h}\right) = p
$$

This kernel VaR estimator is first introduced by Gourieroux, Laurent, and Scaillet (2000) in the context of VaR estimation. Chen and Tang (2005) develop the statistical properties of the nonparametric estimator and apply it to financial returns, which exhibit temporal dependence.

### 2.3.2 Statistical Properties

Next, we give a brief overview of the statistical properties of the nonparametric kernel VaR estimator, $\hat{\nu}_p$. Certain conditions have to be satisfied in studying the properties of the kernel VaR estimator.

1. The series $\{X_i\}_{i=1}^{n}$ has to be continuously stationary and $\alpha$-mixing, and has density function $f$ and cumulative distribution function $F$.

2. The density function $f$ has continuous second derivative in a neighborhood and $f(\nu_p) > 0$.

---

8 Specifically, the daily log-return series of the Nasdaq index and Microsoft from January 1, 1999 to December 31, 2002, which consist of four years of data (n=1000).

9 For the log return series $\{X_t\}_{t=1}^{n}$ let $F_k^t$ be the $\sigma$-algebra of events generated by $\{X_t, k \leq t \leq l\}$ for $l \geq k$. The $\alpha$-mixing coefficient (Rosenblatt, 1956) is defined as $\alpha(k) = \sup_{A \in F_k^t, B \in F_{t+k}^t} |P(AB) - P(A)P(B)|$. The series is said to be $\alpha$-mixing if $\lim_{k \to \infty} \alpha(k) = 0$. A series is geometric $\alpha$-mixing if $\alpha(k) \leq c \rho^k$ for $k > 1$, some constant $c > 0$ and $\rho \in (0, 1)$. 

---
3. \( G \) is a univariate cumulative distribution function and corresponds to a probability density function \( K \), which satisfies the following moment conditions:

\[
\int_{-\infty}^{\infty} uK(u)du = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} u^2K(u)du = \sigma_K^2
\]

4. The smoothing parameter \( h \) satisfies \( nh^{3-\beta} \to \infty \) as \( h \to 0 \) for any \( \beta > 0 \), and \( nh^4\log^2(n) \to 0 \) as \( n \to \infty \).

Assumption 1 indicates the types of dependent series which can be used for the kernel VaR estimation, whereas assumption 2 contains standard conditions for quantile estimation. Assumption 3 and 4 are commonly imposed conditions in kernel smoothing. Under these conditions, as \( n \to \infty \), the bias and variance of the kernel estimator are respectively (Chen and Tang, 2005)

\[
E(\hat{v}_p) = v_p - \frac{1}{2}h^2\sigma_K^2f'(v_p)f^{-1}(v_p) + o(h^2)
\]

\[
\text{Var}(\hat{v}_p) = \frac{1}{n}f^{-2}(v_p)\sigma_h^2(p; n) - \frac{2}{n}hf^{-1}(v_p)b_K + o\left(\frac{h}{n}\right)
\]

where \( b_K = \int uK(u)G(u)du, \sigma_h^2(p; n) = \{p(1-p) + 2\sum_{k=1}^{n-1}(1 - \frac{k}{n})\gamma_h(k)\}, \) and \( \gamma_h(k) = \text{cov}\left\{G(\frac{v_p-X_1}{h}), G(\frac{v_p-X_{k+1}}{h})\right\} \).

Though the kernel VaR estimator is biased which is true for every nonparametric estimator, there is strong evidence of convergence of the kernel estimator \( \hat{v}_p \) to \( v_p \) at a rate given above. The biggest gain from using the kernel VaR estimator is the reduction in variability of VaR estimates. It has been shown that both the kernel and the sample quantile estimators share the same leading asymptotic variance term\(^{10}\) (Chen and Tang, 2005). However, the kernel estimator reduces the variance in the second order of \( h/n \) as \( b_h > 0 \). This second-order reduction is still significant

\(^{10}\)The sample VaR estimator \( \hat{v}_p^s \) has the following properties: \( E(\hat{v}_p^s) = v_p + O(n^{-3/4}) \), \( \text{Var}(\hat{v}_p^s) = \)

38
considering that the data are thin in the tail. It also clearly demonstrates that the variance of the kernel estimator depends on the dependence (covariance) structure of the return series, and the variability increases when \( p \) gets smaller as \( f(v_p) \) tend to become thinner. Chen and Tang further establish the asymptotic normality of \( \hat{\nu}_p \) such that

\[
\sqrt{n}(\hat{\nu}_p - \nu_p) \xrightarrow{d} N(0, \sigma^2(p) f^{-2}(v_p))
\]

where \( \sigma^2(p) = \lim_{n \to \infty} \sigma_h^2(p; n) \), which can be used to construct asymptotic confidence intervals for \( \nu_p \) as well as to carry out tests on hypotheses regarding \( \nu_p \).

### 2.3.3 Implementation

In practice, this nonparametric VaR is estimated from \( n \) observations by replacing the unknown distribution of the portfolio return, \( G(.) \) by a smoothed approximation as in Gourieroux, Laurent, and Scaillet (2000). For this purpose, we introduce a Gaussian kernel and define the estimated VaR, denoted by \( \hat{\nu}_p \), as

\[
\frac{1}{n} \sum_{i=1}^{n} \Phi\left( \frac{\hat{\nu}_p - X_i}{h} \right) = p
\]

where \( \Phi \) is the c.d.f. of the standard normal distribution and \( h \) is the selected bandwidth. Equation (4) is solved numerically by a Gauss-Newton algorithm. Let \( \hat{\nu}_p^{(m)} \) denote the approximation at the \( m^{th} \) step of the algorithm, the updating is then given

\[
n^{-1} f^{-2}(v_p) \sigma^2(p; n) \{1 + o(1)\} \quad \text{where} \quad \sigma^2(p; n) = \{p(1-p) + 2 \sum_{k=1}^{n-1} (1 - \frac{k}{n}) \gamma(k)\} \quad \text{and} \quad \gamma(k) = \text{cov}\{I(X_1 < v_p), I(X_{k+1} < v_p)\} \quad \text{for positive integers} \quad k.
\]

Under the conditions of Assumptions 1-4, \( |\sigma_h^2(p; n) - \sigma^2(p; n)| = o(h) \), which indicates that \( \sigma_h^2(p; n) \) differs from the unsmoothed \( \sigma^2(p; n) \) by an amount of \( o(h) \).
by the recursive formula:

$$\hat{v}_p^{(m+1)} = \hat{v}_p^{(m)} + \frac{1}{n} \sum_{i=1}^{n} \phi\left(\frac{\hat{v}_p^{(m)} - X_i}{h}\right) - p$$

where $\phi$ is the p.d.f. of the standard normal distribution.

The starting value for the algorithm can be set equal to the VaR obtained under a Gaussian assumption or the historical VaR (empirical quantile), and the algorithm stops when $\hat{v}_p^{(m)}$ converges to a steady state, which means that the increment from $\hat{v}_p^{(m)}$ to $\hat{v}_p^{(m+1)}$ reaches a very small value, say $10^{-5}$. Other choices than the Gaussian kernel may also be made without affecting the procedure substantially. The Gaussian kernel has the advantage of being easy to integrate and differentiate from an analytical point of view, and to implement from a computational point of view.

### 2.4 Application to the Cattle Feeding Margin

#### 2.4.1 Model and Data

The cattle feeding margin is described as in Manfredo and Leuthold (2001) incorporating fixed feeding technology. It is assumed that feeding each head of feeder cattle weighing 650 pounds to 1100 pounds requires 45 bushels of corn. Thus the cattle feeding margin is defined as

$$\text{margin}($/head) = \text{live cattle price} \times 11 - \text{feeder cattle price} \times 6.5 - \text{corn price} \times 45^{11}$$

Usually cattle are continuously marketed and placed on feed (Davies and Widawsky, 1995). It is reasonable to assume cattle feeding is a continuous process with decision makers examining the volatility of feeder cattle (young), live cattle (market-ready). \[11\] Other feeding costs such as vet costs are assumed to be zero.
and corn on a regular basis in a portfolio framework. To hedge the market risk of commodities, financial instruments such as futures or options can be utilized in the cattle feeding margin analysis. As futures have different maturity months, in this essay, we suppose that it takes five months for the inputs, corn and feeder cattle to turn into the output live cattle. Therefore we will use May corn and May feeder cattle to produce October live cattle.

As seen in equation (5), price series are needed. The price series used in this paper are daily closing futures prices of corn, feeder cattle, and live cattle, which are published by the Chicago Board of Trade (CBOT) and the Chicago Mercantile Exchange (CME). The three time series span the period from November 1971 through May 2000 providing nearly 28 years of daily data (5010 observations)\textsuperscript{12}. All futures contracts with maturities of one year or less were used. The margin series is then constructed by equation (5). All four return series including the three commodity price returns as well as the margin return are defined as daily log return. In what follows, we display the VaRs for the corn prices (from the perspective of the corn producer), for the feeder cattle prices (from the perspective of the specialized feeder cattle producer), for the live cattle prices (from the perspective of the farrow-to-finish operation), and for the cattle feeding margin (from the perspective of the specialized cattle producer who buys feeder cattle and sells live cattle).

For comparative purpose, the most popular parametric procedure GARCH (1,1) model and a simple full-valuation procedure historical simulation are also developed

\textsuperscript{12}1. A few randomly missing values are deleted.

2. The future prices of May corn, May feeder cattle, and October live cattle are matched by observation dates.
for the 90%, 95%, and 99% confidence levels (table 3). The GARCH (1,1) model is specified as

$$R_t = \mu_t + \sigma_t \eta_t, \quad \eta_t \sim \text{iid } N(0, 1)$$

$$\sigma^2_{t+1} = \alpha_0 + \alpha_1 R^2_t + \beta_1 \sigma^2_t$$

We apply the univariate GARCH (1,1) model to each of the four return series respectively and report the parameter estimates as well as their standard errors in table 2. Inserting the parameter estimates in table 2.1 into (6) yields daily volatility forecasts. VaR at any given period $t$ is then:

$$\text{VaR}_t = \alpha \hat{\sigma}_{t+1}$$

where $\hat{\sigma}_{t+1}$ is the volatility forecast and $\alpha$ is the scaling factor corresponding to the desired confidence level.

The historical simulation method models the entire return distribution with the VaR designated as the quantile associated with the desired level of confidence. The historical simulation procedure used in this manuscript is similar to the methods of Linsmeier and Pearson (1996). The historical simulation is implemented as follows. First, on May 12, 2000, the cattle feeding margin is calculated using the concurrent prices of corn, feeder cattle, and live cattle as in equation (5). Second, the prices of corn, feeder cattle, and live cattle observed on May 12, 2000 are exposed to their respective previous 5,000 days of actual percentage change so that three new price series

\footnote{We refrain from estimating a multivariate GARCH model for the three-component portfolio. Instead a univariate GARCH model for each of the three components and the margin is estimated. This corresponds to the nonparametric procedure that is used later as the nonparametric kernel estimator is only applicable to univariate distributions.}

are generated. We refrain from estimating a multivariate GARCH model for the three-component portfolio. Instead a univariate GARCH model for each of the three components and the margin is estimated. This corresponds to the nonparametric procedure that is used later as the nonparametric kernel estimator is only applicable to univariate distributions.
for corn, feeder cattle, and live cattle are generated. Third, the cattle feeding margin is recalculated using these new component prices (after experiencing the percentage changes), creating 5,000 new values of the cattle feeding margin. Next, each of these new values of the cattle feeding margin is subtracted from the actual feeding margin realized on May 12, 2000, yielding 5,000 differences between the cattle feeding margin on May 12, 2000 and the simulated values of the feeding margin previously generated. Finally, the sample quantile from the distribution of these differences associated with desired confidence level becomes the VaR estimate.

2.4.2 Empirical Results

In line with the discussion above, three estimation procedures of VaR are applied to the cattle feeding data. They are respectively the nonparametric kernel method, the variance-covariance method, and the historical simulation (HS). Before we carry out estimation, we plot the four daily log-return series as well as their corresponding autocorrelation functions in figure 2.2a-2.2d, which indicate some dependence in the feeder cattle and live cattle series. Thus, a GARCH (1,1) model is appropriate for estimating the series volatility forecasts so as to parametrically estimate the VaRs. Parameter estimates of the GARCH (1,1) model are displayed in table 2.1.
Figure 2.2a. CBOT May corn daily returns and its autocorrelation function

Figure 2.2b. CME May feeder cattle daily returns and its autocorrelation function

Figure 2.2c. CME October live cattle daily returns and its autocorrelation function

Figure 2.2d. The cattle-feeding margin return series and its autocorrelation function
Table 2.1 presents the parameter estimates of the GARCH (1,1) model using respectively four time series of the returns of corn, feeder cattle, live cattle, and the cattle-feeding margin. All parameter estimates are significant at the 1% level, which reinforces the previous results of time-varying volatilities of the return series.

To gain insights into the dynamics of these four return series, we plot in figure 2.3a-2.3d the kernel estimates of the return densities for the four series with the normal densities for comparative purpose. It is apparent that the kernel densities for corn, feeder cattle, and the margin substantially deviate from the normal densities, and the kernel density estimates suggest that for a specified confidence level, the nonparametric VaR estimates should be smaller than their parametric counterparts. The kernel density estimates for live cattle, on the other hand, are somehow close to its normal density estimates. In general, the observed deviation from normality

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Corn</th>
<th>Feeder Cattle</th>
<th>Live Cattle</th>
<th>Margin</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>2.35E-5***</td>
<td>3.85E-6***</td>
<td>2.16E-6***</td>
<td>2.44E-5***</td>
</tr>
<tr>
<td></td>
<td>(3.72E-7)</td>
<td>(1.35E-7)</td>
<td>(9.13E-8)</td>
<td>(3.88E-7)</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.280***</td>
<td>0.095***</td>
<td>0.255***</td>
<td>0.290***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.003)</td>
<td>(0.005)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.774***</td>
<td>0.892***</td>
<td>0.787***</td>
<td>0.769***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
</tbody>
</table>

Note: standard errors are in parenthesis. ***level of significance 99%
shown in the four density plots implies the inappropriateness of parametric models, i.e. GARCH (1,1) in this study, which make the normality assumption in the first place.
Figure 2.3a. Kernel Density Estimates for Corn Returns

Figure 2.3b. Kernel Density Estimates for Feeder Cattle Returns

Figure 2.3c. Kernel Density Estimates for Live Cattle Returns

Figure 2.3d. Kernel Density Estimates for Margins
We carried out the 90%, 95% and 99% VaR estimation using all three methods. The results are summarized in table 2.2. Take the margin return series as an example, the nonparametric 90% one-day VaR estimate is 0.007 suggesting that the maximum one day loss for our cattle feeding portfolio is 0.7% under normal market movements in the prices of corn, feeder cattle, and live cattle. The parametric estimate for the same VaR is however 0.023, which clearly indicates a higher level of risk for our portfolio. For the corn and margin returns, we see increasing VaR estimates by using nonparametric estimator, historical simulation, and parametric approach at all three levels except a minor violation at the 95% level. There is little change among the three VaR estimates at all three levels for the feeder cattle returns, and these were all around 0.01 at the 90% level, and 0.02 at the 95% and 99% level after removing the negative sign. For the live cattle returns, we observe quite large discrepancies between the nonparametric kernel VaR estimates and the other two estimates. Specifically, the kernel estimates are much higher than the other two estimates at the 90% and 95% level, but a little lower than the other two at the 99% level. We also present bootstrapping standard errors for the kernel VaR estimates and the HS estimates. The standard errors for the parametric VaR estimates are obtained by finding the standard deviation of the time-varying VaR estimates, \( \text{VaR}_t = \alpha \tilde{\sigma}_{t+1} \) where the volatility forecasts were traced out from estimating the GARCH (1,1) model. For all series, the variability of the kernel estimates is uniformly smaller than that of the parametric estimates.
<table>
<thead>
<tr>
<th>Description</th>
<th>Corn</th>
<th>Feeder Cattle</th>
<th>Live Cattle</th>
<th>Margin</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>90% 95% 99%</td>
<td>90% 95% 99%</td>
<td>90% 95% 99%</td>
<td>90% 95% 99%</td>
</tr>
<tr>
<td>Nonparametric Kernel Density Estimation</td>
<td>0.008 0.022 0.030</td>
<td>0.009 0.017 0.022</td>
<td>0.017 0.021 0.032</td>
<td>0.007 0.022 0.028</td>
</tr>
<tr>
<td>Std. Error</td>
<td>0.012 0.011 0.010</td>
<td>0.011 0.010 0.012</td>
<td>0.011 0.010 0.012</td>
<td>0.012 0.011 0.010</td>
</tr>
<tr>
<td>Historical Simulation</td>
<td>0.012 0.018 0.032</td>
<td>0.011 0.017 0.024</td>
<td>0.009 0.014 0.025</td>
<td>0.012 0.018 0.033</td>
</tr>
<tr>
<td>Std. Error</td>
<td>0.003 0.004 0.008</td>
<td>0.004 0.004 0.006</td>
<td>0.003 0.006 0.005</td>
<td>0.006 0.008 0.002</td>
</tr>
<tr>
<td>Variance-Covariance Method</td>
<td>0.022 0.029 0.041</td>
<td>0.014 0.018 0.026</td>
<td>0.012 0.015 0.021</td>
<td>0.023 0.029 0.041</td>
</tr>
<tr>
<td>Std. Error</td>
<td>0.016 0.021 0.029</td>
<td>0.007 0.009 0.013</td>
<td>0.007 0.009 0.013</td>
<td>0.016 0.021 0.030</td>
</tr>
</tbody>
</table>

Note: standard errors for nonparametric kernel estimation and historical simulation are obtained by nonparametric bootstrap.

Table 2.2: Daily VaRs for the Four Returns Series and for Different Confidence Levels (90%, 95%, 99%)
2.4.3 Summary and Conclusions

A nonparametric method proposed by Chen and Tang (2005) provides an easily implemented and few assumption demanded alternative in calculating daily VaR for the examined portfolio, the cattle feeding margin. However, other parametric estimator, i.e. GARCH (1,1) and the simple historical simulation (full-valuation procedure), also offer well-calibrated VaR measures for all three confidence levels. It is most likely because of the fact that the cattle feeding margin, as defined in this study, is a portfolio of linear instruments (future prices). Overall, it is concluded that at least for this portfolio, the kernel estimates appear to be more efficient than the parametric ones, which is a significant advantage in practice due to the fact that a small reduction in variance can translate to a large reduction of provision in the absolute dollar term.

This study is the first known attempt at empirically examining the applicability as well as performance of the nonparametric kernel estimator of VaR in the context of an agricultural enterprise portfolio, i.e. cattle feeding. The nonparametric estimator has the advantage of being free of distributional assumptions on return series, while being able to capture fat-tail and asymmetry distribution of returns automatically. In other words, the model risk of using a nonparametric approach is lower than that of a parametric approach. However some disadvantages of the nonparametric VaR estimator should be mentioned. One drawback is that the nonparametric estimator is basically designed for the analysis of univariate distributions. In the cattle feeding example, a linear relationship between margin and input & output (fixed feeding technology) is assumed. Therefore, this shortcoming of nonparametric VaR estimator is naturally overcome. Another disadvantage is the requirement of a reasonable sample
size to ensure good performance. Again in our cattle feeding example, daily futures prices are used to estimate the VaR. Agricultural futures are actively traded at the Chicago Board of Trade and the Chicago Mercantile Exchange. Abundant data are stored at these trading agencies.
Bibliography


CHAPTER III

CREDIT RISK ANALYSIS OF HOME MORTGAGE PORTFOLIOS USING CONSUMER SURVEY DATA

3.1 Introduction

During past a few years, rapid price appreciation in real estate markets created substantial equity for homeowners. Under these favorable conditions, mortgage defaults were relatively low, and the accumulation of equity cushioned lenders against large losses. However, with the recent emerging subprime mortgagors, slumping housing market and flagging home prices, it becomes clear that managing loan portfolios of all types requires more advanced analytical tools to assess risk in changing environments. With better analytical tools, financial institutions can establish appropriate reserves against credit losses, suggest pricing differentials to compensate for risk, identify market segments from which they would like to attract new business, and make strategic lending decisions.

The main improvement proposed by the New Basel Capital Accord\textsuperscript{14} is a shift from rules-based to process-oriented regulatory practice and methods (Karacadag and Taylor, 2000). This means a transfer of emphasis from the strict categories and

\textsuperscript{14}New Basel Capital Accord, also named Basel II is the second of the Basel Accords, which are recommendations on banking laws and regulations issued by the Basel Committee on Banking Supervision. Basel II was initially published in June 2004 and was constantly updated until November 2007.
formula that are employed before to a more refined and flexible approach. Banks are allowed to develop internal rating systems and use industry-sponsored models for determining the amount of regulatory capital they must hold.

This study concentrates on estimating the distribution of loan losses which are caused by borrowers’ credit risk, which is the primary source of risk for a lender since it captures the risk of loss from borrower defaults. A continuous-form representation of the loan loss distribution is illustrated in figure 3.1 (Credit Suisse Financial Products, 1997). It is characterized by a smooth distribution with a fact right tail since small losses have a lower bound of zero and may occur with higher probability, while large losses typically occur with low probabilities. The probability density function of loan losses for the whole portfolio vary among different portfolios, but they tend to be both highly skewed and leptokurtic (Ong, 1999). The expected loss is a long-run average that is accounted in loan pricing and covered by the loan loss reserve (allowance for loan losses). Unexpected loss is the maximum potential loss at a given level of confidence (usually 99% to 99.99%). The unexpected loss is not accounted in loan pricing but it requires an amount of financial resources to cover the loss. Extreme losses (larger than the value-at-risk) occur so rarely so that it is generally assumed that it is too costly to hold enough capital to fully cover them.
Considering the loan loss characteristics, credit risk models allow portfolio managers to quantify risk at both the portfolio and individual loan contribution levels. These models can be used to estimate a lender’s probability density function (shown in above graph) for credit losses, which can be utilized to obtain the expected loss (that will be used for loan pricing) and the unexpected loss (that conventionally refers to the Value-at-Risk amount of loss). The difference between these two amounts of losses is defined as the economic capital, which is an amount of financial resources serving as a cushion to the unexpected losses.

In recent years new methods and models have been developed to quantify credit risk on a portfolio basis. The four most prominent credit risk models are CreditMetrics™ (RiskMetrics Group of J.P. Morgan, released in 1997)\textsuperscript{15}, PortfolioManager™ (Moody’s

\textsuperscript{15}See J.P. Morgan (1997).
KMV Corporation, released in 1993\textsuperscript{16}, CreditRisk\textsuperscript{+} (Credit Suisse Financial Products, released in 1997)\textsuperscript{17}, and CreditPortfolioView\textsuperscript{TM} (McKinsey and Company, 1997)\textsuperscript{18}. Despite the differences in their distributional assumptions and solution techniques, recent studies conclude that these models are more or less compatible when they are parameterized consistently and the models are correctly specified (Kern and Rudolph, 2001; Gordy, 2000; Finger, 1999; Koyluoglu and Hickman, 1998). Based on consumer survey data availability and the ability to satisfy model assumptions, CreditRisk\textsuperscript{+} appears to be the most appropriate model for our home mortgage loan portfolio.

The objective of this essay is to apply the CreditRisk\textsuperscript{+} model to a residential mortgage loan portfolio so as to generate a loss distribution for the entire portfolio and further calculate the capital reserve for maintaining such a portfolio. The essay differs from the existing literature that applies the CreditRisk\textsuperscript{+} model to various loan portfolios since it incorporates the nonparametric estimation techniques to predict the borrower’s default probability, which in reality does not necessarily have a normal (or another specified) distribution. In such cases, conventional parametric estimation methods such as the Probit or the Logit model yields biased and inconsistent estimates when a wrong distribution is assumed.

The paper is organized as follows. In section 2 we give a brief overview about the CreditRisk\textsuperscript{+} model (mathematical derivation is given in the appendix at the end of the essay). Section 3 presents the data used in this study as well as the preliminary analysis of the data. In section 4 we introduce and estimate each single input parameter in the CreditRisk\textsuperscript{+} model, and also compare the parametric and

\textsuperscript{16}See Kealhofer (1998).
\textsuperscript{17}See Credit Suisse Financial Products (CSFP, 1997).
nonparametric estimation methods. Section 5 presents our model output, which is mainly a full loan loss distribution at the portfolio level. Concluding remarks are presented in section 6.

### 3.2 CreditRisk⁺ Overview

CreditRisk⁺ model\(^{19}\), introduced by Credit Suisse First Boston is based on the insurance approach that uses survival analysis to model a sudden event of borrower default. No assumptions are made about the cause of default. Credit defaults occur as a sequence of independent events in such a way that it is not possible to forecast the exact time of any one default nor the exact total number of defaults. The default event is modeled as a discrete random variable with a probability distribution, i.e. a Poisson distribution, as there is exposure to default losses from a large number of borrowers and the probability of default by any particular borrower is small. By introducing the probability generating function, it is possible to calculate the distribution of portfolio losses analytically without the need to perform Monte Carlo simulations (see CSFP, 1997 for more details). Even though the model assumes no causal link between any two defaults, background factors, such as the state of economy may cause the incidences of default to be correlated. The model considers the effect of possible background factors by using sector analysis.

The model works as follows. It generally takes a few input parameters and

\(^{19}\)CreditRisk⁺ is a trademark of Credit Suisse Financial Products, a subsidiary of Credit Suisse First Boston. CreditRisk⁺ methodology is freely released to the public. CSFP’s website contains the technical document and a spreadsheet implementation of the model, which is capable of handling up to 4,000 loans and 8 macro factors.
transforms these inputs into a full loan loss distribution at the portfolio level. The CreditRisk+ model allows explicit calculation of the loss distribution of a portfolio of credit exposures. Figure 3.2 shows a brief overview of the model structure. The model inputs are credit exposures, loss rates, default rates and their volatilities. All these inputs are at the individual level. By representing the default events as a Poisson process and utilizing the probability generating function\(^{20}\), CreditRisk+ model calculates the probability that a loss of a certain multiple of the chosen unit of exposure will occur, and builds a recurrence relationship of the probabilities that certain amounts of loss occur (see appendix for technical details). This allows a full loss distribution at the portfolio level to be generated.

<table>
<thead>
<tr>
<th>Inputs (Individual Level)</th>
<th>Output (Portfolio Level)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Default Probabilities</td>
<td>Loan Loss Distribution</td>
</tr>
<tr>
<td>✓ Default Probability Volatilities</td>
<td></td>
</tr>
<tr>
<td>✓ Credit Exposures</td>
<td>Expected Loss</td>
</tr>
<tr>
<td>✓ Loss Rates</td>
<td>99\textsuperscript{th} VaR</td>
</tr>
</tbody>
</table>

Figure 3.2: The CreditRisk+ Model Structure

\(^{20}\)The probability generating function is defined in terms of an auxiliary variable \(z\) by \(F(z) = \sum_{n=0}^{\infty} p(n \text{ defaults}) z^n\).
3.3 Data

The data used in this study is called Consumer Finance Monthly Private data. It was and has been continuously collected by Center for Human Resource Research (CHRR) of The Ohio State University. The database contains the survey data collected from February 2005 to December 2007. It provides relatively more recent household information including home mortgage, credit card history, track of bill and loan payment, measures of debt stress, household assets and liabilities and so on. The total sample size is 10,986, among which 35.8% of consumers (3,933) carry at least one residential mortgage loan during the survey period. 68.6% of households (7,533) report the market values of their residential properties based on their own knowledge. 74.7% of households (8,212) earn annual income including salary from all family members living in the residential property, investment income and all other earnings. 66.6% of households (7,317) obligate to monthly payment, consisting of payment on primary mortgage, student loan, installment loan, auto loan, and required minimum payment on all credit cards that the household owned. 23.6% of consumers (2,596) plan to pay back their mortgages earlier than maturity. 2.3% of households’ (253) files have been sent to collection. 1.2% of households (132) have been more than 60 days late on their mortgage payment. 9.5% of households (1,048) have filed for bankruptcy in last 12 months.

Our analysis focuses on concurrent home mortgage loan carriers only.21 In the full sample of home mortgage loan carriers (2,852), 10.66% of them have ever been more than 60 days late on their mortgage payment or filed for bankruptcy in last 12 months.

21One limitation of this original data is caused by missing observations. As a result, the number of observations that could be used in estimations is less than 3,933.
These households are classified as the default group. Table 3.1 displays the definitions of variables and their sample statistics (mean and standard deviation) for both default and non-default groups. The default group of households has a slightly higher average level of mortgage loan volume and outstanding balance on mortgage comparing to the non-default group, however the market values of residential properties of the defaulted households are hardly beaten by those of non-defaulted households. While the default group has slightly less monthly payment on all debt obligations, their total household income level are also lower than the non-default group on average. Not surprisingly the default group has significantly more missed payments on their debt obligations, more files sent to collection, and has higher stress level over their debt comparing to the non-default group. The default group of households also has significantly more children (under 18) in the house and less education than the non-default group. The default households are slightly older than the non-default group on average but the age difference is not quite significant. Furthermore, the defaulted households appear to dwell in states with higher unemployment rates. One interesting thing is that more households in the default group seem to plan prepayment of their mortgages, which reflects the result that defaulted households have higher debt stress levels.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Default (Std Dev)</th>
<th>Non-default (Std Dev)</th>
<th>Mean Diff.</th>
<th>Test p-value&lt;sup&gt;22&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>60+ days delinquent or filed for bankruptcy</td>
<td>10.66% (-)</td>
<td>89.34% (-)</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>volume of primary mortgage loan</td>
<td>120,443.43 (146,707.57)</td>
<td>127,142.59 (121,860.93)</td>
<td>0.097</td>
<td></td>
</tr>
<tr>
<td>outstanding balance on mortgage</td>
<td>120,438.49 (146,709.17)</td>
<td>127,009.44 (121,830.51)</td>
<td>0.099</td>
<td></td>
</tr>
<tr>
<td>required monthly payment on all debt</td>
<td>1,464.59 (1,046.22)</td>
<td>1,595.28 (1,285.25)</td>
<td>0.089</td>
<td></td>
</tr>
<tr>
<td>household total income</td>
<td>88,118.23 (146,339.21)</td>
<td>114,668.39 (168,890.84)</td>
<td>0.557</td>
<td></td>
</tr>
</tbody>
</table>

<sup>22</sup>These p-values indicate the significance of the difference of means between two groups.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Default</th>
<th>Non-default</th>
<th>Mean Diff. Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Std Dev)</td>
<td>(Std Dev)</td>
<td>p-value</td>
</tr>
<tr>
<td>market value of residential property (if sold)</td>
<td>224,899.10</td>
<td>328,912.63</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(323,600.28)</td>
<td>(497,065.82)</td>
<td></td>
</tr>
<tr>
<td>number of missed payment on all debt</td>
<td>0.62</td>
<td>0.25</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(1.80)</td>
<td>(0.94)</td>
<td></td>
</tr>
<tr>
<td>plan to prepay the mortgage</td>
<td>0.61</td>
<td>0.57</td>
<td>0.596</td>
</tr>
<tr>
<td>1=Yes</td>
<td>(0.49)</td>
<td>(0.50)</td>
<td></td>
</tr>
<tr>
<td>0=No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>files sent to collection</td>
<td>0.23</td>
<td>0.09</td>
<td>0.004</td>
</tr>
<tr>
<td>1=Yes</td>
<td>(0.42)</td>
<td>(0.29)</td>
<td></td>
</tr>
<tr>
<td>0=No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stress over debt (level: 1-5)</td>
<td>2.91</td>
<td>2.37</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(1.23)</td>
<td>(1.14)</td>
<td></td>
</tr>
<tr>
<td>Variable</td>
<td>Default</td>
<td>Non-default</td>
<td>Mean Diff.</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>----------</td>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td></td>
<td>(Std Dev)</td>
<td>(Std Dev)</td>
<td>(Std Dev)</td>
</tr>
<tr>
<td>Number of children (&lt;18) in the family</td>
<td>1.11</td>
<td>0.92</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>(1.30)</td>
<td>(1.22)</td>
<td></td>
</tr>
<tr>
<td>Age of household head</td>
<td>48.45</td>
<td>48.02</td>
<td>0.304</td>
</tr>
<tr>
<td></td>
<td>(10.64)</td>
<td>(12.62)</td>
<td></td>
</tr>
<tr>
<td>Gender of household head</td>
<td>1.54</td>
<td>1.53</td>
<td>0.985</td>
</tr>
<tr>
<td>1=male</td>
<td>(0.5)</td>
<td>(0.5)</td>
<td></td>
</tr>
<tr>
<td>0=female</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education of household head (level: 0-20)</td>
<td>14.22</td>
<td>15.46</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(2.53)</td>
<td>(2.60)</td>
<td></td>
</tr>
<tr>
<td>Local unemployment rate</td>
<td>4.75</td>
<td>4.71</td>
<td>0.841</td>
</tr>
<tr>
<td></td>
<td>(0.89)</td>
<td>(0.92)</td>
<td></td>
</tr>
</tbody>
</table>


Table 3.1: Variable Definitions and Sample Statistics (Sample Size: 2,852)
3.4 Model Parameterization

Each of the four input parameters in the CreditRisk\(^+\) model are explored in details in this section. The definitions, estimation methods, and assumptions about these parameters are presented in order. Consumer Finance Monthly Private data for 2005 through 2007 is used for deriving the model parameters and to estimate the economic capital requirements for the current portfolio.

The Basel II definition of default indicates that a borrower is in default if she/he files for bankruptcy, foreclosure occurs, or if one of more of his/her loans or leases meet at least one of the following conditions: 1) become nonaccrual, 2) are delinquent 90 days or more, 3) have a charge-off, or 4) become subject to distressed restructuring. This study makes a conservative judgement in terms of default based on data availability\(^23\). Specifically, a debtor is in default if she/he files for bankruptcy, or if she/he is more than 60 days late on mortgage payment.

3.4.1 Default Probabilities

Probability of default (PD) of a single obligor is his/her likelihood of default. The CreditRisk\(^+\) model requires a PD estimate for each obligor over a time horizon usually one year. In practice, the model uses a range of risk-rating grades\(^24\) to represent all debtors. Therefore, The annual default probabilities and their deviations have to

---

\(^{23}\) The Center for Human Resource Research at The Ohio State University only asks the survey question: whether the household has ever been late on the mortgage payment for more than 60 days.

\(^{24}\) Credit rating grades are usually denoted from the highest quality to the loss grade by AAA, AA, A, BBB, BB, B, CCC, CC, C, and so forth. Acceptable risk ratings are from AAA to BBB, a rating of BB is special mention, ratings of B to CC are unacceptable ratings, and C (or below) is a loss loan.
be calculated for each risk rating. Estimates of PD must represent a conservative measure of a long-run average PD. Using historical data series to calculate the actual PD, which is in this case the ratio of the defaulted loans over a given year to the total number of loans at the beginning of the year may be difficult, since annual frequency of observations does not allow for long time series. What’s more, there may not be any defaults among high-quality obligors over a year, and the fact that no defaults have been observed does not necessarily imply a zero probability of default. Consequently, regression analysis has been adopted in the literature to estimate a PD for each rating grade, i.e. \( \ln(PD) = a + b \times \text{Risk Rating} \) (Zech and Pederson, 2006).

Since this paper uses empirical observations on household debts and assets from the Consumer Finance survey data instead of the bank’s loan data, no internal risk ratings are assigned to the borrowers. However, the cross-sectional data allows us to investigate the mortgagor’s default rate as a function of several financial variables associated with the mortgagor. Specifically, the method used in this study focuses on a relationship between default and the outcomes of financial choices that consumers make. Emerging as primary explanatory variables are: the homeowners’ equity position (usually measured by the loan-to-value ratio, or LTV), the homeowners’ capacity to repay the mortgage loan (measured by back-end monthly payment-to-monthly income ratio\(^{25}\)), and the homeowners’ credit reputation, i.e. credit scores. It has been documented in numerous studies (Mints 2006, Katchova and Barry 2005, Smith 1994).

\(^{25}\)Taking into account the fact that a huge number of households have other debt obligations besides just the mortgage loan, a "back-end" ratio refers to the ratio between all monthly debt (such as mortgage loan, car payments, student loans, etc.) and monthly household income. A "front-end" ratio by convention evaluates the borrower’s monthly housing payment as a share of total household monthly income.
that a strong correlation exists between the probability of the mortgagor’s default and these three factors. For instance, an analysis of loans purchased by Freddie Mac\textsuperscript{26} between 1985 and 1989 demonstrated that borrowers of loans with LTV of 95-99% were 5 times more likely to default than the borrowers having loans with LTVs below 80%, and that borrowers with back-end ratio greater than 36% were twice as likely to enter foreclosure as those with ratios below 30% (Mints, 2006). As no credit scores are recorded in our survey data, I would use another credit indicator to represent the borrower’s credit reputation: the number of payments missed on all debt positions including credit cards, primary mortgage, home equity, student loan, installment loans, and auto loans. Lawrence, Smith, and Rhoades (1992) studied for manufactured homes, found that borrower delinquency patterns are important indicators of the likelihood of default.

Three binary choice models are applied to the historical consumer survey data to estimate the default probability of each mortgagor in our portfolio. The parameter estimates associated with the three explanatory variables as well as their standard deviations are presented in table 3.2. The probit and logit estimates of the LTV ratio and the number of missed payments are all significant at the 1% level, while the monthly payment-to-income ratio seems not to be quite significant as suggested by the probit and the logit model. Similarly, the nonparametric estimates of the LTV ratio and the number of missed payments are both significant at the 1% level.\textsuperscript{27}

\textsuperscript{26}Freddie Mac is a Government Sponsored Enterprise (GSE) charged with providing a secondary market for home mortgage loans.

\textsuperscript{27}The last parameter, which is associated with the number of missed payment variable cannot be obtained by estimating the Klein and Spady’s nonparametric model as one parameter has to be fixed to a certain value in Klein and Spady’s model.
As expected these three independent variables are all positively associated with the probability of default, suggesting that higher levels of LTV ratio, payment-to-income ratio, and more numbers of missed payments imply higher default probabilities. The PD for each mortgagor is obtained as the fitted value of the binary regression.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Probit</th>
<th>Logit</th>
<th>Klein-Spaydy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>S.E.</td>
<td>Coeff.</td>
</tr>
<tr>
<td>Intercept</td>
<td>-1.856***</td>
<td>0.087</td>
<td>-3.332***</td>
</tr>
<tr>
<td>Loan-to-value</td>
<td>0.971***</td>
<td>0.131</td>
<td>1.913***</td>
</tr>
<tr>
<td>No.of missed payment</td>
<td>0.092***</td>
<td>0.024</td>
<td>0.166***</td>
</tr>
<tr>
<td>Payment-to-income</td>
<td>0.255</td>
<td>0.175</td>
<td>0.457</td>
</tr>
<tr>
<td>Log(L)</td>
<td>-925.082</td>
<td>-924.847</td>
<td>-920.334</td>
</tr>
</tbody>
</table>

Triple(***), double(**), and single(*) asterisks indicate significance at the 1%, 5% and 10% levels, respectively.

Table 3.2: Parameter Estimates of Default Rate Models

To gain more insight into the dynamics of the default rate distribution, we plot in figure 3.3 and 3.4 the estimated cumulative distribution function and the probability density function for the default rate by applying three different econometric models. Clearly the probit and logit estimates of the CDF and PDF of default rates are quite similar except that the logit density estimates have a fatter tail. However, there is significant difference between nonparametric estimates of the distribution functions and the parametric ones. The nonparametric PDF has a bimodal shape and a very much fat left tail. We conclude that the normality assumption of the probit model as well as the logistical assumption of the logit model are not appropriate for our data, a conclusion that lends support to our distribution-free approach.
Figure 3.3. Estimates of the Cumulative Distribution Function for Default Rates

Figure 3.4. Estimates of the Probability Density Function for Default Rates
3.4.2 Default Rate Volatilities

Published default statistics (Carty & Lieberman, Moody’s Investors Service Global Credit Research, 1996) demonstrate that one-year default rates show significant variation year on year. During periods of economic recession, the number of defaults can be many times the level observed at other times. Therefore, default rate volatilities are usually incorporated in credit risk modeling, reflecting the high fluctuations observed during economic cycles. The Internal Ratings Based (IRB) approach in the Basel II Capital Accord requires a minimum of five years of historical observations to calculate PDs and their variance. Due to insufficient data availability, this manuscript has to assume fixed default rate for each obligor at this moment. However, the consumer finance survey is still going on. If more data become available, the default rate volatility can be obtained as the variance of a set of annual PDs.

3.4.3 Credit Exposure

Credit exposure refers to a total value that a bank is exposed to at the time of default, namely exposure at default (EAD). EAD can simply be the unpaid amount of loan at the time of default. Conventionally, the expected loss that will arise at default is often measured over one year, EAD is accordingly estimated as the outstanding balance of a mortgage loan at the beginning of a specified year. While quantifying losses may be simple, in some situations it may be quite difficult and require the analysis of multiple variables. For example, if Bank A issues a mortgage of $400,000 to household X and X defaults on the note, Bank A’s loss isn’t necessarily $400,000. This is because Bank A may resell the house after the foreclosure with X. When all relevant variables are factored in, Bank A may have lost substantially less than the
original $400,000 loan. Therefore, another concept in credit risk modeling has to be introduced, the loss given default.

### 3.4.4 Loss Given Default

Loss given default or LGD is a common parameter in risk models and it is also an important attribute of any exposure on bank’s client. LGD is defined as a fraction of exposure at default (EAD) that will not be recovered following default. Thus, LGD can be replaced by \( (1 - \text{recovery rate}) \). LGD is facility-specific as such losses are generally understood to be influenced by key transaction characteristics such as the presence of collateral and the degree of subordination. Theoretically, LGD is calculated in different ways, but the most popular is “Gross” LGD, where total losses are divided by EAD. Another method is to divide losses by the unsecured portion of an exposure (where security, usually collateral covers a portion of EAD). This is known as “Blanco” LGD. Gross LGD is most popular among academics because collateral values often are unknown or irrelevant in bond market data, which is the only accessible data by researchers. Blanco LGD on the other hand is mostly adopted by industry practitioners (banks) as banks often have many secured facilities, i.e. mortgage, home equity loans.

Under Basel II to determine the required capital for a bank or lending institution, the institution has to use an estimate of the LGD for each exposure. There are two approaches for deriving this estimate: a foundation approach and an advanced approach. Under the foundation approach, Basel II Supervisor prescribes fixed LGD ratios for certain classes of exposures based on the presence of collateral and the degree of subordination. Under the advanced approach, namely IRB approach the
bank itself determines the appropriate LGD to be applied to each exposure, on the basis of abundant historical data and robust estimation methods. Using internal LGD estimates for capital purposes allows banks to differentiate LGD values according to a wider set of transaction characteristics as well as borrower characteristics, usually resulting in substantial reduction in capital reserve. However this internal estimation and analysis must be validated both by the bank itself and by supervisors and would also be expected to represent a conservative view of long-run averages.

Due to the limitation of consumer survey data, it is unrealistic to estimate LGD grounded in historical experience and empirical evidence for this study. To circumvent the problem of insufficient historical data yet consist with the New Basel Capital Accord (Basel, 2001), a conservative pre-defined LGD measure under the foundation approach is used in this study same as in Zech and Pederson (2004). Specifically, by classifying four LGD grades (see table 3.3), each loan in the current portfolio is assigned a LGD rating according to its collateral-to-loan value. LGD rating 1 is assigned to loans with collateral-to-loan values over 150%. Loans with collateral-to-loan values between 100% and 150% are assigned LGD rating 2. An LGD rating of 3 is assigned to loans with collateral-to-loan values between 50% and 100% and a LGD rating of 4 is assigned to loans with collateral-to-loan values below 50%.
<table>
<thead>
<tr>
<th>LGD Rating</th>
<th>Loss Given Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.00</td>
</tr>
<tr>
<td>2</td>
<td>20.00</td>
</tr>
<tr>
<td>3</td>
<td>50.00</td>
</tr>
<tr>
<td>4</td>
<td>75.00</td>
</tr>
</tbody>
</table>

Table 3.3: Loss Given Default Ratings and Loss Rates (percent)

3.5 Model Output

The main output of the credit risk model is the full loan loss distribution. Table 3.4 provides summary statistics for the analyzed portfolio and a summary of the resulting loan loss distribution. For comparative purpose, column 2, 3, and 4 presents the characteristics of the portfolio loan loss distribution by using probit, logit, and KS probability of default estimates respectively. Total volume is the sum of individual exposures, which are household’s outstanding balance on home mortgage loans. Maximum loss exposure is the sum of exposures multiplied by LGD rates. This value measures the total losses if all mortgagors in the portfolio default. The distribution mean is the expected loss on this portfolio and tail percentiles show the value-at-risk, the entire required risk funds to cover both the expected and unexpected losses.
Summary Data

<table>
<thead>
<tr>
<th></th>
<th>Probit PD Estimates</th>
<th>Logit PD Estimates</th>
<th>KS PD Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of exposures</td>
<td>2,852</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of nondefaulted exposures</td>
<td>2,548</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total volume</td>
<td>$445,455,345</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum loss exposure</td>
<td>$33,129,790</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Loan Loss Distribution Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Probit PD Estimates</th>
<th>Logit PD Estimates</th>
<th>KS PD Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>$5,108,600</td>
<td>$5,164,100</td>
<td>$4,471,200</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>$501,550</td>
<td>$505,420</td>
<td>$463,140</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.17</td>
<td>0.17</td>
<td>0.19</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>3.04(^{28})</td>
<td>3.04</td>
<td>3.05</td>
</tr>
<tr>
<td>95th percentile</td>
<td>$5,959,000</td>
<td>$6,011,000</td>
<td>$5,258,000</td>
</tr>
<tr>
<td>99th percentile</td>
<td>$6,340,000</td>
<td>$6,395,000</td>
<td>$5,614,000</td>
</tr>
<tr>
<td>99.97th percentile</td>
<td>$6,989,000</td>
<td>$7,049,000</td>
<td>$6,219,000</td>
</tr>
</tbody>
</table>

Table 3.4: Loan Loss Distribution Summary

3.6 Capital Adequacy

As defined above, the expected loss represents the required allowance for loan loss. The difference between the value-at-risk at the selected percentile, i.e. 99.97% and the mean is the required amount of economic capital, which serves as a financial cushion for unexpected credit losses on a portfolio. Table 3.5 displays the economic capital

\(^{28}\text{Normal}=3\)
requirements at various confidence levels. The allowance and VARs are generated by the CreditRisk\textsuperscript{+} model using KS estimates of default rates. Typical confidence levels range from 99.00\% to 99.99\%. The choice of confidence levels depends on the lending institution’s level of risk aversion. In commercial banking, with sufficient knowledge about loan characteristics and borrower characteristics, the choice of the confidence level depends on the target debt rating of the financial institution. For example, a 99.90\% capital level corresponds to a single-A debt rating. The Basel II Capital Accord uses the 99.50\textsuperscript{th} percentile in deriving the economic capital that targets a triple-B rating, while many commercial banks adopt 99.97\textsuperscript{th} percentile (the equivalent of a double-A rating), which is also used as a primary confidence level in this study.

<table>
<thead>
<tr>
<th>Loss Percentile</th>
<th>Allowance</th>
<th>Value-at-Risk</th>
<th>Economic Capital</th>
</tr>
</thead>
<tbody>
<tr>
<td>90.00</td>
<td>4,471,200</td>
<td>5,075,000</td>
<td>603,800</td>
</tr>
<tr>
<td>95.00</td>
<td>4,471,200</td>
<td>5,258,000</td>
<td>786,800</td>
</tr>
<tr>
<td>99.00</td>
<td>4,471,200</td>
<td>5,614,000</td>
<td>1,142,800</td>
</tr>
<tr>
<td>99.50</td>
<td>4,471,200</td>
<td>5,747,000</td>
<td>1,275,800</td>
</tr>
<tr>
<td>99.90</td>
<td>4,471,200</td>
<td>6,028,000</td>
<td>1,556,800</td>
</tr>
<tr>
<td><strong>99.97</strong></td>
<td><strong>4,471,200</strong></td>
<td><strong>6,219,000</strong></td>
<td><strong>1,747,800</strong></td>
</tr>
<tr>
<td>99.99</td>
<td>4,471,200</td>
<td>6,383,000</td>
<td>1,911,800</td>
</tr>
</tbody>
</table>

Table 3.5: Economic Capital at Various Confidence Levels

Table 3.5 (column 1) shows the expected credit losses for our mortgage portfolio, which are the same for various selected loss percentiles. Value-at-risk (column 2) is the required total risk funds to cover losses at a given loss percentile. Economic
capital is equal to the value-at-risk less the loan loss allowance\textsuperscript{29}. As indicated in

table 3.5 (column 3) the amount of economic capital nearly doubles as the confidence
level increases from 99.00\% to 99.99\%, which implies the choice of confidence level is
an important parameter.

### 3.7 Sensitivity Analysis

Stress testing is widely used as a supplement for value-at-risk models. It is generally a
way to measure and monitor the consequences of extreme movements in parameters,
which can be historical or hypothetical. Stress testing scenarios are required by the
Basel II Capital Accord. Table 3.6 shows model results under various economic sce-
narios. All of the scenarios are analyzed under the assumption of a 99.97th confidence
level.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Allowance</th>
<th>Value-at-Risk</th>
<th>Economic Capital</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>4,471,200</td>
<td>6,219,000</td>
<td>1,747,800</td>
</tr>
<tr>
<td>Mild Recession</td>
<td>7,744,800</td>
<td>10,590,000</td>
<td>2,845,200</td>
</tr>
<tr>
<td>Moderate Recession</td>
<td>20,653,000</td>
<td>25,870,000</td>
<td>5,217,000</td>
</tr>
<tr>
<td>Severe Recession</td>
<td>30,979,000</td>
<td>37,300,000</td>
<td>6,321,000</td>
</tr>
<tr>
<td>Crisis of 1980s</td>
<td>33,144,000</td>
<td>39,540,000</td>
<td>6,396,000</td>
</tr>
</tbody>
</table>

Table 3.6: Stress Testing at the 99.97th Percentile

The “Base” scenario repeats the results described earlier. The “Mild Recession”
scenario assumes that mortgagor’s default rates do not change but all four LGD rates
increase by 50\% (LGD for rating 4 is capped at 100\%), representing the fact that
\textsuperscript{29}In some documentations, economic capital needs to cover market and operational risks in addi-
tion to credit risk. This study focuses on credit risk only as no bank-side data is available.
collateral values may decline or collateral may become less liquid during a recession. This scenario increases the economic capital requirement by around one million dollars comparing to the “Base” scenario. The “Moderate Recession” scenario assumes that all default probabilities and the LGD rates double (LGD for rating 3 and 4 is capped at 100%). This scenario increases the need for risk funds near three times over that in the “Base” scenario. The “Severe Recession” scenario assumes default probabilities triple and the LGD rates double. Under this scenario, the economic capital has to be raised more than four million dollars comparing to the “Base”. The “Crisis of 1980s” scenario assumes that default probabilities are 30% for all loans, and all LGD ratings migrate to the next lower level. Thus, all loans owning a LGD rating 1 become loans having a LGD rating 2, all LGD rating 2 loans become LGD rating 3 ones, and so on. This scenario generates similar results as the “Severe Recession”.

3.8 Portfolio Management

Credit risk models are able to quantify credit risks at the portfolio level, therefore they are powerful instruments for portfolio managers. To achieve efficient capital allocation and higher profitability risk managers need to propose strategies for constructing exposure portfolios by comparing the risks and returns associated with the portfolio segments that may be bought or sold in the secondary market. One way to reduce financial reserves is through diversification. Portfolio diversification can be measured from the loan loss distribution. If the portfolio were less diversified, the spread of the distribution curve would be wider (a long and fat tail) and a higher level of economic capital would be required. On the contrary, if the portfolio were more diversified, a lower level of economic capital would be required.
The risk of a particular exposure is determined by the size of exposure, the maturity of the exposure, the probability of default and the systematic risk\textsuperscript{30} of the obligor. For managing risks on a portfolio basis, with the aim of creating a diversified portfolio, a different measurement that incorporates all those individual factors is required. This specific measurement is called risk contribution. The risk contribution of an exposure is defined as the incremental effect on a chosen percentile level of the loss distribution, i.e. 99.97th percentile, when the exposure is removed from the existing portfolio. In another word, for a chosen percentile level the risk contribution is the marginal impact of the obligor on the amount of economic capital required to support the portfolio. By ranking obligors in decreasing order of risk contribution, the obligors that requires the most economic capital can easily be identified. The removal of the obligors with the largest risk contributions from a portfolio and focusing on the management of a small number of obligors with large risk contributions lead to a significant reduction in the economic capital.

3.9 Concluding Remarks

Managing credit risk in financial institutions requires the ability to forecast aggregate losses on existing loans and analyze the expected performance of particular segments in the existing portfolio. Various credit risk models serve these functions. A mortgage lender may use this type of model to manage and monitor portfolio risk, analyze the effects of changes in portfolio composition and diversification, evaluate risk-adjusted profitability, and perform risk-based loan pricing. This manuscript is the first attempt to incorporate a nonparametric method in estimating the probability of default, one

\textsuperscript{30}Refers to the effect of background factors, i.e. the state of the economy, countries, or industries.
of the key inputs in all credit risk models, and applies the well-known CreditRisk$^+$ model to a survey data set of home mortgage loans. The results of estimating the CreditRisk$^+$ model are presented and discussed in details. As a final note, our intention is not to replace or discharge other estimation methods for probability of default, since time has shown that they are indeed simple and useful. Instead, we hope to suggest another, admittedly more ambitious, alternative.
Appendix: Technical Documents for CreditRisk$^+$ Model (CSFP, 1997)

This appendix summarizes the technical document created by the Credit Suisse First Boston Corporation in 1997, and presents an analytical technique for generating the full distribution of losses from a portfolio of credit exposures. The technique is valid for any portfolio where the default rate for each obligor is small. Also the original technical documents for the CreditRisk$^+$ Model contain more generalized cases such as in which the default rate is stochastic. In this essay, we only consider the case in which the mean default rate for each obligor in the portfolio is fixed.

Credit defaults occur as a sequence of events in such a way that it is not possible to forecast the exact time of occurrence of any one default or the exact total number of defaults. In this section we derive the basic statistical theory of such processes in the context of credit default risk.

Consider a portfolio consisting of $N$ obligors. It is assumed that each exposure has a definite known probability of defaulting over a one-year time horizon. Thus

$$p_A = \text{Annual probability of default for obligor A} \tag{1}$$

To analyze the distribution of losses arising from the whole portfolio, we introduce the probability generating function defined in terms of an auxiliary variable $z$ by

$$F(z) = \sum_{n=0}^{\infty} p(n \text{ defaults}) z^n \tag{2}$$

An individual obligor either defaults or does not default. The probability generating function for a single obligor is easy to compute explicitly as

$$F_A(z) = 1 - p_A + p_A z = 1 + p_A (z - 1) \tag{3}$$
As a consequence of independence between default events, the probability generating function for the whole portfolio is the product of the individual probability generating functions. Therefore

\[ F(z) = \prod_A F_A(z) = \prod_A (1 + p_A(z - 1)) \quad (4) \]

It is convenient to write this in the form

\[ \log F(z) = \sum_A \log(1 + p_A(z - 1)) \quad (5) \]

Given that the probabilities of default are uniformly small, powers of those probabilities can be ignored. Thus, the logarithms can be replaced using the expression

\[ \log(1 + p_A(z - 1)) = p_A(z - 1) \quad (6) \]

and, in the limit, equation (5) becomes

\[ F(z) = e^{\sum_A p_A(z-1)} = e^{\mu(z-1)} \quad (7) \]

where we write

\[ \mu = \sum_A p_A \quad (8) \]

for the expected number of default events in one year from the whole portfolio. To identify the distribution corresponding to this probability generating function, we expand \( F(z) \) in its Taylor series:

\[ F(z) = e^{\mu(z-1)} = e^{-\mu} e^{\mu z} = \sum_{n=0}^{\infty} \frac{e^{-\mu} \mu^n}{n!} z^n \quad (9) \]

Thus if the probabilities of individual default are small, although not necessarily equal, then from equation (9) we deduce that the probability of realizing \( n \) default events in the portfolio in one year is given by

\[ \Pr(n \text{ defaults}) = \frac{e^{-\mu} \mu^n}{n!} \quad (10) \]
Under our initial assumptions, the distribution of numbers of defaults in a portfolio of exposures in one year has been obtained. However our main objective is to understand the likelihood of suffering given levels of loss from the portfolio, rather than given numbers of defaults. The first step in obtaining the distribution of losses from the portfolio in an amenable form is to group the exposures in the portfolio into bands. This has the effect of significantly reducing the amount of data that must be incorporated into the calculation.

The following table shows the notations which are used for the exposure banding described above.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obligor</td>
<td>A</td>
</tr>
<tr>
<td>Exposure</td>
<td>$L_A$</td>
</tr>
<tr>
<td>Probability of Default</td>
<td>$P_A$</td>
</tr>
<tr>
<td>Expected Loss</td>
<td>$\lambda_A$</td>
</tr>
</tbody>
</table>

In order of perform the calculations, a unit amount of exposure $L$, denominated in a base currency, is chosen. For each obligor A, define numbers $\varepsilon_A$ and $\upsilon_A$ by writing

\[
L_A = L \times \upsilon_A \\
\lambda_A = L \times \varepsilon_A
\]  

(11)

Thus, $\upsilon_A$ and $\varepsilon_A$ are the exposure and expected loss respectively, of the obligor, expressed as multiples of the unit.

The key step is to round each exposure size $\upsilon_A$ to the nearest whole number. This step replaces each exposure amount $L_A$ by the nearest integer multiple of $L$. The
portfolio can then be divided into $m$ exposure bands, indexed by $j$, where $1 \leq j \leq m$.

With respect to the exposure bands, we make the following definitions

<table>
<thead>
<tr>
<th>Reference</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common exposure in Exposure Band $j$ in units of $L$</td>
<td>$v_j$</td>
</tr>
<tr>
<td>Expected loss in Exposure Band $j$ in units of $L$</td>
<td>$\varepsilon_j$</td>
</tr>
<tr>
<td>Expected number of defaults in Exposure Band $j$</td>
<td>$\mu_j$</td>
</tr>
</tbody>
</table>

The following relations hold, expressing the expected loss in terms of the probability of default events

$$\varepsilon_j = v_j \times \mu_j$$

$$\mu_j = \frac{\varepsilon_j}{v_j} = \sum_{A: v_A = v_j} \frac{\varepsilon_A}{v_A}$$

As in equation (8), let $\mu$ stand for the total expected number of default events in the portfolio in one year. Since $\mu$ is the sum of the expected number of default events in each exposure band, we have

$$\mu = \sum_{j=1}^{m} \mu_j = \sum_{j=1}^{m} \frac{\varepsilon_j}{v_j}$$

We have analyzed the distribution of default events previously. We now proceed to derive the distribution of default losses. Let $G(z)$ be the probability generating function for losses expressed in multiples of the unit $L$ of exposure:

$$G(z) = \sum_{n=0}^{\infty} p(\text{aggregate losses}=n \times L)z^n$$

The exposures in the portfolio are assumed to be independent. Therefore, the exposure bands are independent, and the probability generating function can be written
as a product over the exposure bands

\[ G(z) = \prod_{i=1}^{m} G_i(z) \]  

(15)

However, by treating each exposure band as a portfolio and using equation (9), we obtain

\[ G_j(z) = \sum_{n=0}^{\infty} p(n \text{ defaults}) z^{nv_j} = \sum_{n=0}^{\infty} \frac{e^{-\mu_j} \mu_j^n}{n!} z^{nv_j} = e^{-\mu_j + \mu_j z^{v_j}} \]  

(16)

Therefore

\[ G(z) = \prod_{j=1}^{m} e^{-\mu_j + \mu_j z^{v_j}} = e^{- \sum_{j=1}^{m} \mu_j + \sum_{j=1}^{m} \mu_j z^{v_j}} \]  

(17)

This is the desired formula for the probability generating function for losses from the portfolio as a whole.

By defining a polynomial \( P(z) \) as follows

\[ P(z) = \frac{\sum_{j=1}^{m} \mu_j z^{v_j}}{\mu} = \frac{\sum_{j=1}^{m} \left( \frac{\xi_j}{v_j} \right) z^{v_j}}{\sum_{j=1}^{m} \left( \frac{\xi_j}{v_j} \right)} \]  

(18)

the probability generating function in equation (17) can now be expressed as

\[ G(z) = e^{\mu(P(z) - 1)} = F(P(z)) \]  

(19)

This functional form for \( G(z) \) expresses mathematically the compounding of two sources of uncertainty arising, respectively, from the Poisson randomness of the incidence of default events and the variability of exposure amounts within the portfolio.

In the following section, the actual distribution of credit losses is derived from equation (17). For an integer \( n \), let \( A_n \) be the probability of a loss of \( nL \), or \( n \) units from the portfolio. Comparing the definition in equation (14) with the Taylor series expansion for \( G(z) \), we have

\[ p(\text{loss of } nL) = \frac{1}{n!} \frac{d^n G(z)}{dz^n} \bigg|_{z=0} = A_n \]  

(20)
In our case $G(z)$ is given in closed form by equation (17). Using Leibniz’s formula we have

$$
\frac{1}{n!} \frac{d^n G(z)}{dz^n} \bigg|_{z=0} = \frac{1}{n!} \frac{d^{n-1}}{dz^{n-1}} \left\{ G(z) \frac{d}{dz} \sum_{j=1}^{m} \mu_j z^{v_j} \right\} \bigg|_{z=0} 
$$

$$
= \frac{1}{n!} \sum_{k=0}^{n-1} \left( \frac{n-1}{k} \right) \frac{d^{n-k-1}}{dz^{n-k-1}} G(z) \frac{d^{k+1}}{dz^{k+1}} \left( \sum_{j=1}^{m} \mu_j z^{v_j} \right) \bigg|_{z=0}
$$

However

$$
\frac{d^{k+1}}{dz^{k+1}} \left( \sum_{j=1}^{m} \mu_j z^{v_j} \right) \bigg|_{z=0} = \begin{cases} 
\mu_j (k + 1)! & \text{if } k = v_j - 1 \text{ for some } j \\
0 & \text{otherwise}
\end{cases} 
$$

and by definition

$$
\frac{d^{n-k-1}}{dz^{n-k-1}} G(z) \bigg|_{z=0} = (n - k - 1)! A_{n-k-1}
$$

Therefore

$$
A_n = \sum_{k=0}^{n-1} \frac{1}{n!} \binom{n-1}{k} (k+1)! (n-k-1)! \mu_j A_{n-k-1} = \sum_{j: v_j \leq n} \frac{\mu_j v_j}{n} A_{n-v_j}
$$

Using the relation $\varepsilon_j = v_j \times \mu_j$, the following recurrence relationship is obtained

$$
A_n = \sum_{j: v_j \leq n} \frac{\varepsilon_j}{n} A_{n-v_j}
$$

This recurrence relationship allows quick computation of the distribution. In order to commence the computation, we have the following formula for the first term, which expresses the probability of no loss arising from the portfolio

$$
A_0 = G(0) = F(P(0)) = e^{-\mu} = e^{-\sum_{j=1}^{m} \varepsilon_j}
$$
Bibliography


Bibliography


