Essays on Trade Agreements, Agricultural Commodity Prices and Unconditional Quantile Regression

by

Na Li

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of
Doctor of Philosophy
in
Food, Agricultural and Resource Economics

Guelph, Ontario, Canada

ABSTRACT

ESSAYS ON TRADE AGREEMENTS, AGRICULTURAL COMMODITY PRICES AND UNCONDITIONAL QUANTILE REGRESSION

Na Li
University of Guelph, 2013

Advisor:
Professor A. P. Ker

My dissertation consists of three essays in three different areas: international trade; agricultural markets; and nonparametric econometrics. The first and third essays are theoretical papers, while the second essay is empirical.

In the first essay, I developed a political economy model of trade agreements where the set of policy instruments are endogenously determined, providing a rationale for countervailing duties (CVDs). Trade-related policy intervention is assumed to be largely shaped in response to rent seeking demand as is often shown empirically. Consequently, the uncertain circumstance during the lifetime of a trade agreement involves both economic and rent seeking conditions. The latter approximates the actual trade policy decisions more closely than the externality hypothesis and thus provides scope for empirical testing.

The second essay tests whether normal mixture (NM) generalized autoregressive conditional heteroscedasticity (GARCH) models adequately capture the relevant properties of agricultural commodity prices. Volatility series were constructed for ten agricultural commodity weekly cash prices. NM-GARCH models allow for heterogeneous volatility dynamics among different market regimes. Both in-sample fit and out-of-sample forecasting tests confirm that the two-state NM-GARCH approach performs significantly better than the traditional normal GARCH model. For each commodity, it is found that an expected negative price change corresponds to a higher volatility persistence, while an expected positive price change arises in conjunction with a greater responsiveness of volatility.

In the third essay, I propose an estimator for a nonparametric additive unconditional quantile regression model. Unconditional quantile regression is able to assess the possi-

ble different impacts of covariates on different *unconditional quantiles* of a response variable. The proposed estimator does not require d-dimensional nonparametric regression and therefore has no curse of dimensionality. In addition, the estimator has an oracle property in the sense that the asymptotic distribution of each additive component is the same as the case when all other components are known. Both numerical simulations and an empirical application suggest that the new estimator performs much better than alternatives.

ACKNOWLEDGMENTS

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This dissertation consists of three manuscripts on three diversified areas: trade agreements, agricultural markets and unconditional quantile regression. The first and third essays are theoretical papers, while the second essay is empirical.

1.1   Trade Agreements: A Political Economy Perspective

Trade protection policies especially those implemented in agricultural sector stand out as a divergence of trade policy from classical trade theory which promotes free trade. Trade agreements remains uneasy to negotiate (e.g. World Trade Organization (WTO) Doha round) and not always effectively enforced. Economists have provided frameworks/models that explain the structure of optimal trade agreements under varying assumptions.

Two different avenues are prominent in the literature. The first approach treats trade agreements as *incomplete contracts* and utilizes contract theory (see Copeland (1990); Horn (2006); Horn et al. (2010)). The second approach explicitly accounts for political pressure and argues trade-related policy intervention is largely shaped in response to rent seeking demand. This approach is consistent with empirical evidence in that many free-trade-resistant industries such as the agricultural and food sector maintain significant rent

seeking activities. The milestone work of Grossman and Helpman (1994, 1995a,b) "protection for sale" (PFS) model brings rent seeking behavior into the realm of trade policies and concludes tariff rates are affected both by a *political support* motive and a *terms-of-trade* motive. The central predictions emphasize determinants of cross-sectional differences in protection. First, the relationship between trade protection and import penetration depends fundamentally on whether or not an industry is politically organized. Second, protection depends inversely on import demand elasticity. Schleich and Orden (1996) extended the original PFS model by incorporating domestic production policies and concluded that production subsidies can substitute for trade policies that would have otherwise resulted from rent seeking efforts. As a result, without trade agreements, the tariff rate represents only the *terms-of-trade* motive given the presence of production subsidies.

Both approaches contribute to our understanding of the structure of trade agreements. The incomplete-contract approach assumes that production and consumption externalities give rise to policy intervention. However, externalities are difficult to quantify and thus there is very little empirical evidence supporting the theory. In addition, an implicit assumption is that governments are benevolent and immune from political pressure. As noted by Bagwell and Staiger (1999), *it is thus important to consider further the rationale for a trade agreement, within a richer model in which governments may have political concerns*. The political economy approach lends itself to empirical verification (see for example, Goldberg and Maggi (1999); Gawande and Bandyopadhyay (2000); McCalman (2004); Eicher and Osang (2002)). However, the political economy literature does not explain why trade agreements feature discretion over a production subsidy and/or an internal consumption tax.

In the first essay, a political economy model of trade agreements which accounts for uncertainty and contracting costs is developed. Adding incomplete contract theory to the political economy framework yields new insights into the mechanisms that drive trade

agreements. First, if contracting costs are zero it is optimal to use a tariff to offset a foreign export subsidy and restore free trade, as domestic production subsidies can substitute for trade policies and still meet lobbyists' rent seeking demand. This provides rationale for countervailing duties (CVDs) in WTO. Second, if contracting costs are non-zero, an agreement should consider a trade-off between contracting costs and including more state contingent policy constraints. This essay argues that the uncertainty during the lifetime of a trade agreement comes not only from economic conditions but also from political pressures and that the latter may dictate trade policy more than the externality framework.

This study follows Grossman and Helpman (1995a) by assuming that each lobby sets contribution schedules to maximize total net payoff of its members, and the incumbent government maximizes an expected weighted sum of aggregate social welfare and total political contributions received from the lobbies. This study follows an approach similar to Battigalli and Maggi (2002) by assuming that contracting costs are increasing in the number of state variables and policies included in the agreement. Within a competitive two-country setting, the choice of contract form is characterized endogenously as in Horn et al. (2010). In addition, conditions that restrict the *trade volume* effect to be positive is identified; this is in contrast to the original Horn et al. (2010) result with a benevolent government, in which an increasing *trade volume* always has positive effect on the gains of constraining production/consumption policy. In addition, the *monopoly power* effect as defined by Horn et al. (2010) is decomposed into a *trade volume* effect and the effect of *price sensitivity of import demand*. This decomposition is simple yet provides new interpretation for empirical investigation.

1.2   Price Volatility in Agricultural Markets

One issue in the agricultural markets that has long intrigued agricultural economists is the agricultural commodity price volatility, which has been exceptionally high during the last

decade. (FAO and UNCTAD (2011)). Large and unpredictable price variations create a level of uncertainty which increases risks for producers, traders, consumers and governments. Using options and forward contracts to manage risk is more costly for producers and processors when prices are exceptionally volatile. Furthermore, large price uncertainty raises risks to investment and production decisions, particularly where the physical production cycle is long. In addition, volatile prices pose significant problems for market regulators and governments as they need greater skills to manage markets in a volatile state, this is especially the case in underdeveloped countries where households may suffer severe food scarcity and food security problems. Volatile prices also affect consumers as it becomes more difficult for them to make proper consumption decisions when the discretionary income left is uncertain.

The challenge that high commodity price volatility brings highlights the need to better understand its causes, patterns, impacts and measures available to mitigate them. Modelling commodity price volatility helps to forecast the absolute magnitude, quantiles, and in fact, the entire distribution of price changes. Such forecasts are widely used in risk management, derivative pricing and hedging, portfolio selection, among other economic activities.

A better fitting discrete time volatility model can help producers and traders manage price risk and make better production and investment decisions in a volatile market state. When budgeting for production levels, a producer of one commodity may want to know today the likelihood that the price of the product will decline in the future. To hedge a contract, an option trader will want to know the the expected volatility over the life of the contract, as well as how volatile is this option-implied volatility. To make a profit and limit their losses to a comfortable level, a futures trader may want to use a sell or buy stops based on volatility and density prediction.

Modelling commodity price volatility following price spikes is essential for market regulators and governments, particularly when prices are highly volatile. A proper volatility

model help monitor and predict the movement of commodity prices. Market regulation and other policies that aims to manage the risks and mitigate the impacts of high volatility can be developed accordingly.

While there is a large body of literature on price volatility since the introduction of generalized autoregressive conditional heteroscedasticity (GARCH) framework by Engle (1982) and Bollerslev (1986), the usual assumption in the majority of the proposed volatility models applied to commodity prices is that the error process is conditionally normally distributed. However, normal GARCH models do not adequately capture heavy tails, large kurtosis, and occurrence of extreme events, which are often key features of financial time series. As an extension, Bollerslev (1987) proposed modelling the innovations via a GARCH model with a Student's t-distribution and Fernández and Steel (1998) extended it further considering the skewed t-distribution.

A few recent studies have proposed using a mixture of two normal distributions to model volatility. Among them, Haas et al. (2004) introduced the general symmetric Normal Mixture(NM) GARCH model, and Alexander and Lazar (2006) further investigated the property of NM-GARCH(1,1) model and provided empirical evidence that the generalized two-component NM-GARCH(1,1) models perform better than both symmetric and skewed Student's t-GARCH models for modelling exchange rates. A clear advantage of the NM-GARCH model over Students' t-GARCH models is the capability to model time-varying conditional skewness and kurtosis. Another advantage of NM-GARCH model is that it accounts for multiple states, which may contribute to economic interpretation. Haas et al. (2004), for example, pointed out that an NM-GARCH model accommodates the possibility of distinct types of responses to heterogeneous market shocks. Alexander and Lazar (2009) argued that a component with relative low variance could represent a "usual" state, which generally occurs, while a component with high variance could represent a "crash" state which rarely occurs.

An important empirical regularity of equity markets is the fact that volatility increases more after price declines than after price increases by the same magnitude, such an asymmetric return-volatility relationship is documented as a financial leverage effect in early influential studies (Black, 1976; Christie, 1982; Engle and Ng, 1993; Glosten et al., 1993). Glosten et al. (1993) introduced the GJR-GARCH(1,1) model allowing unequal response weight for negative and positive shocks. In commodity markets, contrary to equity markets, an "inverse leverage effect" may exist, i.e., a rise in the price level has stronger impact on the price volatility than a drop in the price. This is understandable, as increased prices of commodities generally bring panic and give rise to higher volatility. Previous studies, such as Geman and Shih (2009) and Chang (2012) found such an effect in energy markets. This effect has not been considered with respect to agricultural commodity markets.

In the second essay, I tested whether NM-GARCH models are appropriate for modelling and forecasting agricultural commodity price volatility. Because out-of-sample interval forecast validation, though pivotal in risk management and policy-making, is rarely applied in previous volatility modelling for agricultural commodity prices. My work also try to fill this gap by performing Value-at-Risk (VaR) validation tests. In addition, in order to capture the possible state-specific asymmetric volatility responses to negative and positive shocks, this essay followed Alexander and Lazar (2009) and considered the NM-GJR-GARCH model.

1.3   Unconditional Quantile Regression: Nonparametric Additive Model

Although much concern has been dedicated to mean regression, the conditional mean of a response variable Y given values of covariates $\mathbf{X}$. Other distributional statistics are also often of interest. For example, the sample median is more resistant to gross errors than the

sample mean of salaries, the distribution of which is typically skewed to the right because relative few people earn high salaries. In addition, conditional quantiles can be applied to the construction of laboratory reference range (e.g., Cole, 1988), and prediction intervals given past values of a time series (e.g., Koenker and Zhao, 1996; Zhou and Portnoy, 1996; Kocherginsky et al., 2005). Koenker and Bassett (1978) introduced a linear model of conditional quantile regression. In many practical applications the linear quantile regression model might not fit the available data well. In an effort to make conditional quantile regression models more flexible, there is a growing literature on nonparametric conditional quantile regression (e.g., Chaudhuri, 1991a,b; Koenker et al., 1994; He and Shi, 1994; Fan et al., 1994; He et al., 1998; Yu and Jones, 1998). Fully nonparametric estimation is usually unattractive in multivariate settings because of the curse of dimensionality. For this reason, dimension reduction modelling methods have been developed. For example, De Gooijer and Zerom (2003), Yu and Lu (2004), Horowitz and Lee (2005) and Cheng et al. (2011) have considered an additive structure of $M_\tau(\mathbf{X})$.

Conditional quantiles are with respect to the distribution of the error term ε_τ, thereby the changes of regression coefficients over different quantiles are not easily interpreted. Quantiles of the unconditional distribution of the outcome variable Y, on the other hand, may be of more general interest and easily interpreted. Firpo et al. (2009a) introduced unconditional quantile regression to estimate the impacts of changes in the distribution of covariates on *unconditional quantile* $q_\tau(Y)$. They show that the marginal effect on the unconditional quantile of a small change in the distribution of explanatory variables, defined as the *unconditional quantile partial effect* (UQPE), equals the average derivative of the probability response model divided by the probability density at the unconditional quantile point. Firpo et al. (2009a) suggested three estimation methods for the UQPE: ordinary least square (OLS), logistic regression (Logit) and a fully nonparametric method introduced by Newey (1994) (NP).

While a large body of literature has been dedicated to additive modelling of conditional quantiles, an additive modelling has not yet been applied to unconditional quantiles. In the third essay, I propose an estimator for a nonparametric additive unconditional quantile regression model. The estimator does not require d-dimensional nonparametric regression and therefore does not suffer curse of dimensionality. In addition, the estimator has an oracle property: the asymptotic distribution of the estimator of each additive component is the same as in the case when all other additive components are known. Both numerical simulations and an empirical application to Boston house price data are provided to illustrate the performance of the proposed estimator.

Chapter 2

Trade Agreements as Endogenously Incomplete Contracts: A Political Economy Approach

2.1 Introduction

The divergence of trade policy from trade theory has justifiably drawn significant attention. Trade agreements have never been easy to negotiate (e.g. World Trade Organization (WTO) Doha round) nor have they always been effectively enforced. Economists have provided frameworks/models that explain the structure of optimal trade agreements under varying assumptions. Two different avenues are prominent in the literature.

The first approach takes trade agreements as *incomplete contracts* and utilizes contract theory (see Copeland (1990); Horn (2006); Horn et al. (2010)). Because of uncertainty, a complete contract needs to be able to foresee every possible regulatory need and state-contingent. Internal measures, for example, as noted by Horn (2006), are not explicitly covered in the General Agreement on Tariffs and Trade (GATT) because writing and enforcing

9

such an agreement is impossibly complex and costly. Horn et al. (2010) endogenously included the set of policy instruments thereby modelling trade agreements as endogenously incomplete contracts. They identify a *monopoly power* effect (denoted by Johnston's optimal tariff rate), a *trade volume* effect, and an *instrument substitutability* effect as the key features of the contracting environment that determine the costs of discretion over production subsidy/consumption tax in a trade agreement.

The second approach explicitly accounts for political pressure and argues trade-related policy intervention is largely shaped in response to rent seeking demand. This approach is consistent with empirical evidence in that many free-trade-resistant industries such as the agricultural and food sector maintain significant rent seeking activities. The milestone work of Grossman and Helpman (1994, 1995a,b) "protection for sale" (PFS) model brings rent seeking behavior into the realm of trade policies and concludes tariff rates are affected both by a *political support* motive and a *terms-of-trade* motive. The central predictions emphasize determinants of cross-sectional differences in protection. First, the relationship between trade protection and import penetration depends fundamentally on whether or not an industry is politically organized. Second, protection depends inversely on import demand elasticity. Schleich and Orden (1996) extended the original PFS model by incorporating domestic production policies and conclude that production subsidies can substitute for trade policies that would have otherwise resulted from rent seeking efforts. As a result, without trade agreements, the tariff rate represents only the *terms-of-trade* motive given the presence of production subsidies.[1] Maggi and Rodriguez-Clare (1998, 2007) argued that in addition to *terms-of-trade* motive, there is *domestic-commitment* motive for trade agreements as governments use trade agreements as a credible announcement in a game with domestic lobbies.

Both approaches contribute to our understanding of the structure of trade agreements.

---

[1]Schleich and Orden (2000) incorporated both externality and political economy motive in the trade agreement framework to address the issue of environment quality.

The incomplete-contract approach assumes that production and consumption externalities give rise to policy intervention.[2] However, externalities are difficult to quantify and thus there is very little empirical evidence supporting the theory. In addition, an implicit assumption is that governments are benevolent and immune from political pressure. As noted by Bagwell and Staiger (1999), *it is thus important to consider further the rationale for a trade agreement, within a richer model in which governments may have political concerns*. Note that the political economy approach lends itself to empirical verification. The predictions from PFS model were first tested in Goldberg and Maggi (1999) and Gawande and Bandyopadhyay (2000) in a U.S. context. Recent work of McCalman (2004) has applied the PFS model to analyze ongoing trade liberalization in Australia. Mitra et al. (2002) have tested the PFS predictions in Turkey, a developing country. These analyses generally found the PFS predictions to be consistent with the empirical data. Eicher and Osang (2002) perform a comparison of predictions from the PFS model with those of Findlay and Wellisz's (1982) tariff formation function model and found both models perform well with the data. Note however that the political economy literature does not explain why trade agreements feature discretion over a production subsidy and/or an internal consumption tax.

This manuscript develop a political economy model of trade agreements which accounts for uncertainty and contracting costs.[3] Adding incomplete contract theory to the political economy framework yields new insights into the mechanisms that drive trade agreements.

---

[2] Horn et al. (2010) explored a political economy version of their model as well but utilized a reduced form for the politician's objective function and assumed that governments place a greater weight on producer surplus. However, as Grossman and Helpman (1994, pg. 834) noted in their paper, a reduced form would catch the effects of institutional changes on a government's willingness and ability to protect particular interest groups but not on the government's weighting of political contributions relative to national welfare. In addition, most of results in this manuscript could not be obtained with a reduced-form approach, for example, the political economy rationale for countervailing duty law and the implication of uncertainty about the structural change of lobbies, and governments' weight on contributions.

[3] Maggi and Rodriguez-Clare (1998, 2007) also utilized incomplete contract theory as they found that the optimal agreement that stipulates discretionary tariffs below the upper bound is identical to an incomplete contract which fails to specify the future contributions of a lobby. However, as they noted, their model does not capture other important factors such as uncertainty and contracting costs, which is the focus of the present paper.

First, if contracting costs are zero it is optimal to use a tariff to offset a foreign export subsidy and restore free trade, as domestic production subsidies can substitute for trade policies and still meet lobbyists' rent seeking demand. This provides rationale for countervailing duties (CVDs) in WTO. Second, if contracting costs are non-zero, an agreement should consider a trade-off between contracting costs and including more state contingent policy constraints. This study argues that the uncertainty during the lifetime of a trade agreement comes not only from economic conditions but also from political pressures and that the latter may dictate trade policy more than the externality framework.

This study follows Grossman and Helpman (1995a) by assuming that each lobby sets contribution schedules to maximize total net payoff of its members, and the incumbent government maximizes an expected weighted sum of aggregate social welfare and total political contributions received from the lobbies. I follow an approach similar to Battigalli and Maggi (2002) by assuming that contracting costs are increasing in the number of state variables and policies included in the agreement. Working within a competitive two-country setting, I characterize the choice of contract form endogenously as in Horn et al. (2010). However, I identify conditions that restrict the *trade volume* effect to be positive; this is in contrast to the original Horn et al. (2010) result with a benevolent government, in which an increasing *trade volume* always has positive effect on the gains of constraining production/consumption policy. In addition, I decompose the *monopoly power* effect as defined by Horn et al. (2010) into a *trade volume* effect and the effect of *price sensitivity of import demand*. This decomposition is simple yet provides new interpretation for empirical investigation.

In the following section I develop a political economy model and find the optimal trade and domestic production policies resulting from both a noncooperative equilibrium and a costless trade agreement. Section 2.3 extends the model by accounting for uncertainty and contracting costs. Section 2.4 further extends the model by characterizing the optimal

National Treatment (NT)-based trade agreement. The conclusions are outlined in section 2.5.

2.2 A Political Economy Model of Production Subsidies and Trade Policies

The present work considers trade between two countries (Home, Foreign) and denote Foreign by *. It is assumed that there is a numeraire good 0 which is not subject to any policy interventions and n other nonnumeraire goods in each country. Prior to policy intervention some of these n goods are imported while others may be exported. A representative individual of Home maximizes the following utility:

$$u = c_0 + \sum_{i=1}^{n} u_i(c_i),$$

where c_0 is the consumption of numeraire good 0 and c_i is the consumption of good i. The sub-utility functions $u_i(\cdot)$ are assumed differentiable, increasing and strictly concave. I let q_i denote the domestic consumer price of good i in Home, and $D_i(q_i)$ denote the representative individual's demand for good i, which is the inverse of $u_i'(\cdot)$. Their indirect utility is given by

$$v(q,e) = e + S(q),$$

where e is total spending, and $q = (q_1, q_2, \ldots, q_n)$ is the vector of domestic consumer prices of the nonnumeraire goods and $S(q) \equiv \sum_i u_i[D_i(q_i)] - \sum_i q_i D_i(q_i)$ is the consumer surplus associated with these goods.

The numeraire good 0 is produced using only labor, has constant returns to scale, and an input-output ratio of 1. It is assumed that the aggregate labor supply is large enough to maintain positive production of this good. The competitive wage is 1. Each of the

other goods is produced from labor and an industry-specific input. Letting p_i represent domestic producer price, the aggregate profit accruing to the specific factor used in industry i, denoted by $\Pi_i(p_i)$, is an increasing function of p_i. The aggregate supply of good i is the slope of the profit function $(X(p_i) = \Pi_i'(p_i) > 0$ for $i = 1, 2, \ldots, n.)$

In this section it is assumed that each government can intervene in any of its nonnumeraire sectors using an *ad valorem* tariff/export subsidy and a specific domestic production subsidy/tax.[4] I denote the *ad valorem* tariff or export subsidy for industry i by τ_i and thus

$$q_i \equiv \tau_i \omega_i,$$

where ω_i represents the world price. If $\tau_i > 1$ it represents the tariff on an import good or the export subsidy on an export good. Conversely, if $\tau_i < 1$ it represents an import subsidy or an export tax. I introduce a domestic production subsidy/tax for industry i and denote by s_i. The pricing relationship between the Home producer price and the Home consumer price can be expressed as

$$p_i \equiv q_i + s_i.$$

Net imports of good i in Home are $M_i = ND_i(q_i) - X_i(p_i)$, where N is the size of the population, which I henceforth normalize to 1. Similarly, net imports of good i in Foreign are $M_i^* = D_i^*(q_i^*) - X_i^*(p_i^*)$. Note that $q_i = \tau_i \omega_i$, $p_i = \tau_i \omega_i + s_i$, $q_i^* = \tau_i^* \omega_i$ and $p_i^* = \tau_i^* \omega_i + s_i^*$. Clearing of the world market requires that

$$M_i(\tau_i \omega_i, s_i) + M_i^*(\tau_i^* \omega_i, s_i^*) = 0, \qquad i = 1, 2, \ldots, n. \tag{2.1}$$

Equation (2.1) allows us to solve for ω_i, the world market clearing price of good i, as a function of τ_i, τ_i^*, s_i and s_i^*. I denote this functional relationship by $\omega_i(\tau_i, \tau_i^*, s_i, s_i^*)$.

The vector of trade policies $\tau = (\tau_1, \tau_2, \ldots, \tau_n)$, the vector of domestic production sub-

---

[4]Consumption taxes is introduced as a policy instrument in section 2.4 when analyzing the effect of an NT clause.

sidy policies $s = (s_1, s_2, ..., s_n)$, and market clearing prices $\omega = (\omega_1, \omega_2, ..., \omega_n)$ generate government revenue of

$$R(\tau, s, \omega) = \sum_i (\tau_i - 1)\omega_i [D_i(\tau_i \omega_i) - X_i(\tau_i \omega_i + s_i)] - \sum_i s_i X_i(\tau_i \omega_i + s_i).$$

A representative individual obtains income from wages, possible claims (profits) to one of the industry-specific inputs, as well as government transfers. Individuals are assumed to own at most one type of claim to the industry-specific inputs (e.g., claims to industry-specific human capital). The owners of the specific factor used in industry i, with their common interest in protection or subsidies for that industry, may choose to create a lobby or join an existing lobby in an attempt to influence government policy. However, not all owners of specific factors succeed in organizing politically (free rider problems, transaction costs, etc.) and thus some industries have no means to effectively influence policy. The set of industries, denoted by L, where specific factor owners are organized is assumed exogenous. Following Grossman and Helpman (1994) I assume that lobby groups express their policy demands by means of political contribution schedules.

Each lobby group represents a certain industry i and sets contribution schedules $C_i(\tau, s, \cdot)$ to maximize the joint welfare of its members.[5] Note that I have omitted arguments that represent Foreign policies thus allowing us to distinguish the case of a noncooperative equilibrium (where the contribution schedule depends only on the policies of the Home government) from that of cooperative equilibrium (where the contributions may also depend on policies implemented by the Foreign government). The objective of lobby group i can be expressed as

$$V_i = W_i(\tau, s, \omega) - C_i(\tau, s, \cdot),$$

---

[5]Those industries which do not organize have $C_i(\tau, s, \cdot) = 0$.

15

where

$$W_i(\tau, s, \omega) \equiv l_i + \Pi_i(p_i) + \alpha_i[R(\tau, s, \omega) + S(\tau\omega)]$$

is its gross joint welfare. Note l_i is the joint labor income of these factor owners, α_i is the fraction of the voting population that owns the specific factor used in industry i and $S(\cdot)$ is the consumer surplus as previously defined.

I assume that governments maximize their political welfare which is equal to the weighted sum of the welfare of its representative voter and total political contributions received. The Home government's objective is

$$G = \sum_{i \in L} C_i(\tau, s, \cdot) + aW(\tau, s, \omega), \qquad a \geq 0$$

where a reflects the government's weighting of aggregate social welfare relative to political contributions and W represents the aggregate social welfare which is given by

$$W(\tau, s, \omega) \equiv l + \sum_i \Pi_i(p_i) + R(\tau, s, \omega) + S(\tau\omega),$$

where l is the aggregate labor income.

The sequence of actions by the various political forces in the two-level game are as follows. First, various lobbies in each country simultaneously and noncooperatively set contribution schedules that make the amount of political contributions contingent on possible policy outcomes. Each lobby takes as given the contribution schedules of all other lobbies at home and abroad. Second, both governments weigh net gains from acting cooperatively versus noncooperatively. In either case, the contribution schedules in one country are unobservable to the other. At this stage, costs of cooperation-drafting and negotiating a detailed trade agreement-which I refer to as contracting costs, become important. An implicit assumption throughout is that trade agreements are perfectly enforceable: I abstract from issues of self-enforcement of trade agreements.

2.2.1 The Noncooperative Equilibrium

I derive the policy choices which occur in the absence of a trade agreement (i.e., a non-cooperative equilibrium). Taking Foreign government's policies (τ^*, s^*) as given, Home government's noncooperative policy vectors satisfy the following two conditions:

$$(\tau^0, s^0) = \arg\max_{(\tau, s)} G(\tau, s, \tau^*, s^*), \tag{2.2}$$

and

$$(\tau^0, s^0) = \arg\max_{(\tau, s)} [V_i(\tau, s, \tau^*, s^*) + G(\tau, s, \tau^*, s^*)] \text{ for every } i \text{ in } L. \tag{2.3}$$

The above two conditions follow directly from the proposition 1 of Grossman and Helpman (1994) by setting $P^0 = (\tau^0, s^0)$ and $P^{*0} = (\tau^{*0}, s^{*0})$, and the system of equation can be easily solved (see appendix A.1.1) to obtain an expression for Home's equilibrium policies given by

$$\tau_i^0 - 1 = -\frac{I_{iL} - \alpha_L}{a + \alpha_L} \frac{X_i}{\omega_i M_i'} + \frac{1}{e_i^*} + s_i^0 \frac{X_i'}{\omega_i M_i'}, \tag{2.4}$$

and

$$s_i^0 = \frac{I_{iL} - \alpha_L}{a + \alpha_L} \frac{X_i}{X_i'} - \frac{M_i + (\tau_i^0 - 1)\omega_i M_i^{*\prime} \tau_i^*}{D_i' \tau_i^0 + M_i^{*\prime} \tau_i^*}, \tag{2.5}$$

where I_{iL} is an indicator variable that equals 1 if industry i is represented by a lobby and 0 otherwise, $\alpha_L \equiv \sum_{j \in L} \alpha_j$ is the fraction of voters who are represented by lobbies, and $e_i^* \equiv \tau_i^* \omega_i M_i^{*\prime} / M_i^*$ is the elasticity of Foreign export supply or import demand (corresponding to M_i^* is negative or positive) in industry i.

Equation (2.4) defines the noncooperative choice of τ_i given domestic production policy s_i and Foreign policies (s_i^* and τ_i^*). The three terms on the right-hand side of equation (2.4) represent the *political support* motive, *terms-of-trade* motive, and *substitutability* of domestic production subsidies for trade policies respectively. The first two components consist of

17

the expression for noncooperative trade policies in Grossman and Helpman (1995a). Thus, the noncooperative equilibrium trade policies defined by Grossman and Helpman (1995a) is a special case when the government cannot implement production policies ($s_i^0 \equiv 0$). Equation (2.4) also shows that the *substitutability* of s_i is limited if the industry has low price sensitivity of supply (X_i' is small) or high price sensitivity of import demand ($|M_i'|$ is large). This observation suggests a possible cross-industry prediction for uses of tariff policies. That is, sectors whose market conditions may limit the use of a domestic production subsidy as a substitute for a tariff (e.g. infant industries), may be more likely to lobby for high tariffs in comparison to other sectors. As I will show later in this manuscript, it therefore saves contracting costs by leaving discretion to production subsidies while maintaining constraints on trade policies for these industries in trade agreements.

Equation (2.5) defines the noncooperative choice of s_i given Home trade policy τ_i and Foreign policies (s_i^* and τ_i^*). The two terms on the right-hand side of equation (2.5) represent the *political support* motive, and *substitutability* of trade policies for production subsidies respectively.

Solving equations (2.4) and (2.5) yields Home's noncooperative policies

$$\tau_i^0 - 1 = \frac{1}{e_i^*}, \tag{2.6}$$

and

$$\frac{s_i^0}{p_i} = \frac{I_{iL} - \alpha_L}{a + \alpha_L} \frac{1}{\eta_i}, \tag{2.7}$$

where $\eta_i \equiv p_i X_i'/X_i$ is the elasticity of supply in industry i in Home.[6] Not surprisingly, equation (2.6) illustrates that Home exploits any international markets power by exerting a tariff (or export tax) at the same level as Johnson's optimal tariff rate (the inverse of elasticity of Foreign export supply or import demand). This is because production subsidy (tax)

---

[6]This result is also identified by Schleich and Orden (1996).

replaces tariff (export tax) to compensate the lobbyists' political contributions. Equation (2.7) reflects that, in a noncooperative equilibrium, the optimal production policies for each country is to subsidize domestic production in industries represented by lobbies ($I_{iL} = 1$) at the expenses of industries not represented by lobbies ($I_{iL} = 0$).

2.2.2 The Costless Trade Agreement

Assuming there is no contracting costs, the global political welfare requires the two governments choose policy vectors to maximize the Global Policy Preference (Ω) as defined by the weighted sum

$$\Omega \equiv a^*G + aG^* = a^* \sum_{i \in L} C_i(P;P^*) + a \sum_{i \in L^*} C_i^*(P;P^*) + a^*a[W(P,P^*) + W^*(P^*,P)].$$

Note, the weight of each country's aggregate social welfare are equalized (to a^*a) while the relative weights of aggregate social welfare and political contributions within each country are identical to that of the noncooperative case. The interpretation is the Global Policy Preference represents the political welfare of a global government that consists of two countries each has its own preference for political welfare, and international lump-sum transfers are available. A close real life example is the European Union (EU).[7]

It is not difficult to get globally efficient policies following similar steps in section 2.2.1 (see appendix A.1.2):

$$\tau_i^e - \tau_i^{*e} = 0, \tag{2.8}$$

$$s_i^e = \frac{I_{iL} - \alpha_L}{a + \alpha_L} \frac{p_i}{\eta_i}, \tag{2.9}$$

---

[7]This function of global welfare follows Grossman and Helpman (1995a), further research would consider alternative forms such as using different Nash weights reflecting relative bargaining abilities of the two governments.

and

$$s_i^{*e} = \frac{I_{iL}^* - \alpha_L^*}{a^* + \alpha_L^*} \frac{p_i^*}{\eta_i^*},$$ (2.10)

where $\eta_i^* \equiv p_i^* X_i^{*\prime} / X_i^*$ is the elasticity of supply in industry i in Foreign.

Comparing equation (2.9) with equation (2.7), I find that the expression for the noncooperative and globally efficient levels of s are the same. For a given state of the world, that is, fixing I_{iL}, α_L, a, η_i, the rate of production subsidy/tax ($|s_i|/p_i$) is the same in both cases and thus global inefficiency can not be created by the lobbying over domestic production policies. Consequently, just like Horn et al. (2010), an agreement that only constrains s cannot increase global welfare relative to the noncooperative equilibrium and thus is not an optimal agreement.

On the other hand, equation (2.8) stipulates the following proposition:

Proposition 1. *In a trade agreement free of contracting costs, the tariff rate τ_i in the importing country should be equal to the export subsidy rate τ_i^* in the exporting country for any industry i.*

Proposition 1 thus provides a political economy rationale for WTO's countervailing duty law and confirms that a costless trade agreement removes inefficiency resulting from trade policies.[8] Recall the interpretation of equation (2.6) is that absent trade agreement, trade policy will not be used for redistribution despite the lobbying, when production subsidies/taxes are available. A costless trade agreement, additionally, makes distortionary trade policies offset each other so that the efficient level of trade is restored. Note this prediction differs from that in Grossman and Helpman (1995a) which rely on the assumption that

---

[8]Bagwell and Staiger (1999) also observed that even a politically motivated governments sign a trade agreement only to correct the terms of trade externalities, but through a thought experiment by hypothesizing a world where governments are not motivated by terms-of-trade effects. Their approach is quite different from my approach as they do not consider domestic production policy as a substitute for trade policy.

governments can not use domestic subsidies at their disposal, as a consequence, tariff and export subsidies in the same industry exactly offset each other only in special cases, such as when the lobby groups in the two countries are politically equally powerful.[9]

2.3 The Optimal Trade Agreement

Before characterizing the optimal agreement two important assumptions need to be introduced. First, there are four sources of uncertainty during the lifetime of the agreement that may lead to an incomplete contract: the relative weight of aggregate social welfare (a and a^*), the fraction of population that is represented by lobbies (α_L and α_L^*), whether an industry organizes or dissolves its political lobby (I_{iL} and I_{iL}^*), and the price elasticity of supply (η_i and η_i^*).[10] Second, there are two categories of contracting costs: the costs of including state variables (e.g. a, α_L, I_{iL}, M_i and their Foreign counterparts), and the costs of including policy variables (e.g. τ and s and their Foreign counterparts).[11] Following Battigalli and Maggi (2002), I assume that contracting costs are increasing in the number of state variables and policies included in the trade agreement. I use the following function to denote contracting cost:

$$c = c(n_p, n_s), c'_{n_p} > 0, c'_{n_s} > 0,$$

where n_p and n_s are the number of policy and state variables in the agreement respectively.

The optimal agreement maximizes expected Ω less contracting costs, defined as "Ex-

---

[9]In equation (25) of Grossman and Helpman (1995a), $\tau_i - \tau_i^* = \left(-\frac{I_{iL} - \alpha_L}{a + \alpha_L} \frac{X_i}{\omega_i M_i'} \right) - \left(-\frac{I_{iL}^* - \alpha_L^*}{a^* + \alpha_L^*} \frac{X_i^*}{\omega_i M_i^{*'}} \right)$

[10]Technological changes can affect the price elasticity of supply (η_i) and the number of employees remaining in a certain industry. Political circumstances can change significantly through time as different political parties may come into government, particularly in developing countries. Nordhaus (1975, pg. 188) noted the implicit weighting function on consumption has positive weight during the electoral period and zero (or small) weights in the future. In addition some industries may create or dissolve a lobby during the lifetime of the agreement.

[11]I adopt the definition of contracting costs by Horn et al. (2010), where the cost of including a variable in the agreements captures both the cost of describing the variable, the cost of verifying its value ex post, and more broadly, negotiation costs.

pected Net Global Policy Preference". An agreement of the form

$$A^0 = \left\{ \tau_i = \tau_i^*, s_i = \frac{I_{iL} - \alpha_L}{a + \alpha_L} \frac{p_i}{\eta_i}, s_i^* = \frac{I_{iL}^* - a_L^*}{a^* + \alpha_L^*} \frac{p_i^*}{\eta_i^*} \right\},$$

which imposes the first-best policies derived in section 2.2.2 has $n_p = 4n$ and $n_s = 4 + 4n$ and therefore costs $c(4n, 4 + 4n)$ and yields Expected Net Global Policy Preference equal to $\mathbb{E}(\Omega^0) - c(4n, 4 + 4n)$. It is easy to verify that if contracting costs are negligible, A^0 is the optimal trade agreement. At the other extreme, if contracting costs are prohibitively high then the noncooperative equilibrium occurs. The interesting case is where contracting costs matter but do not prohibit a trade agreement.

I have previously shown that the inefficiency in the noncooperative equilibrium results from τ, not s, and thus an optimal trade agreement should at least impose constraints on τ. The question remaining is whether the agreement should also constrain s.

Recall equation (2.5) gives the expression for $s_i^N(\tau_i, \tau_i^*)$, the noncooperative choice of s_i if τ_i and τ_i^* are constrained but s_i and s_i^* are left to discretion. That is

$$s_i^N(\tau_i, \tau_i^*) = \frac{I_{iL} - \alpha_L}{a + \alpha_L} \frac{X_i}{X_i'} - (\tau_i - 1) \omega_i \frac{M_i^{*\prime} \tau_i^*}{D_i' \tau_i + M_i^{*\prime} \tau_i^*} - \frac{M_i}{D_i' \tau_i + M_i^{*\prime} \tau_i^*}.$$

Similarly, I can get

$$s_i^{*N}(\tau_i, \tau_i^*) = \frac{I_{iL}^* - \alpha_L^*}{a^* + \alpha_L^*} \frac{X_i^*}{X_i^{*\prime}} - (\tau_i^* - 1) \omega_i \frac{M_i' \tau_i}{D_i^{*\prime} \tau_i^* + M_i' \tau_i} - \frac{M_i^*}{D_i^{*\prime} \tau_i^* + M_i' \tau_i}.$$

The efficient choices of production policies, $s_i^E(\tau_i, \tau_i^*)$ and $s_i^{*E}(\tau_i, \tau_i^*)$ solves $\nabla_{s_i} \Omega(P, P^*) = 0$, and $\nabla_{s_i^*} \Omega(P, P^*) = 0$ simultaneously.

Whether a trade agreement which binds τ_i should also constrain s_i depends on the magnitude of the gain in expected Ω implied by replacing $s_i^N(\tau, \tau^*)$ with $s_i^E(\tau, \tau^*)$. If the expected gain is less than the contracting cost incurred by negotiating on s_i, then it is better to exclude s_i from the trade agreement. Without loss of generality, assuming that

22

$s_i^N(\tau_i, \tau_i^*) > s_i^E(\tau, \tau^*)$ and $s_i^{*N}(\tau_i, \tau_i^*) > s_i^{*E}(\tau_i, \tau_i^*)$ for a given state of the world, the gain of constraining s_i and s_i^* is given by

$$\Omega(s_i^E(\tau_i, \tau_i^*), s_i^{*E}(\tau_i, \tau_i^*), \tau, \tau^*, \cdot) - \Omega(s_i^N(\tau_i, \tau_i^*), s_i^{*N}(\tau_i, \tau_i^*), \tau, \tau^*, \cdot)$$
$$= \int_{s_i^N(\tau_i, \tau_i^*)}^{s_i^E(\tau_i, \tau_i^*)} \frac{\partial \Omega}{\partial s_i}(P, P^*) \mathrm{d}s_i + \int_{s_i^{*N}(\tau, \tau^*)}^{s_i^{*E}(\tau, \tau^*)} \frac{\partial \Omega}{\partial s_i^*}(P, P^*) \mathrm{d}s_i^*. \qquad (2.11)$$

Since $\nabla_s \Omega(s^E(\tau, \tau^*), s^{*E}(\tau, \tau^*), \tau, \tau^*) = \nabla_{s^*}\Omega(s^E(\tau, \tau^*), s^{*E}(\tau, \tau^*), \tau, \tau^*) = 0$ and it is assumed that Ω is concave in s and s^*, a sufficient condition for the right-hand side in equation (2.11) to be small is that $|\frac{\partial \Omega}{\partial s_i}\left(s_i^N(\tau_i, \tau_i^*), s_i^{*N}(\tau_i, \tau_i^*), \tau, \tau^*, \cdot\right)|$ and $|\frac{\partial \Omega}{\partial s_i^*}\left(s_i^N(\tau_i, \tau_i^*), s_i^{*N}(\tau_i, \tau_i^*), \tau, \tau^*, \cdot\right)|$ are small. After manipulating it is found

$$|\frac{\partial \Omega}{\partial s_i}\left(s_i^N(\tau_i, \tau_i^*), s_i^{*N}(\tau_i, \tau_i^*), \tau, \tau^*, \cdot\right)| = \frac{aa^*X_i'}{|D_i^{*'}|\tau_i^* + |M_i'|\tau_i}|M_i - (\tau_i^* - 1)\omega_i \tau_i^* |D_i^{*'}|| \equiv B_i.$$

Due to the possible state of the world and henceforth the ambiguity of the sign of the term of $M_i - (\tau_i^* - 1)\omega_i \tau_i^* |D_i^{*'}|$ it is difficult to assess the effect of *trade volume* ($|M_i|$). Note this differs from Horn et al. (2010), where *trade volume* effect is identified as always positive, i.e., a rise in trade volume always increases the gain of constraining s and s^* in trade agreements. The difference stems from the different rationale for policy intervention, now governments' objective is a weighted sum of national welfare and political contributions. As a result, an increase in trade volume and therefore a rise in dead weight loss not necessarily lower a government's political welfare as the government is compensated from political contributions. This difference also provides scope for empirical investigation.

With some more specific assumptions, I am able to shed light on circumstances under which it is desirable to exclude s and s^* from the trade agreement. Suppose Home is the net importer in industry i and Foreign imposes an export subsidy not too high (in fact GATT/WTO prohibits export subsidies so $(\tau_i^* - 1)$ is always nonpositive) or exerts an ex-

23

port tax, then the term $M_i - (\tau_i^* - 1)\omega_i\tau_i^*|D_i^{*\prime}|$ is positive, which I refer to as an *effective constraint* on the Foreign export subsidy. It is easy to have

$$B_i = \frac{aa^*X_i'}{|D_i^{*\prime}|\tau_i^* + |M_i'|\tau_i}[M_i - (\tau_i^* - 1)\omega_i\tau_i^*|D_i^{*\prime}|]. \tag{2.12}$$

The above equation can also be written as

$$B_i = \frac{aa^*X_i'|M_i'|}{|D_i^{*\prime}|\tau_i^* + |M_i'|\tau_i}\left(\frac{M_i}{|M_i'|} - \frac{(\tau_i^* - 1)\omega_i\tau_i^*|D_i^{*\prime}|}{|M_i'|}\right),$$

where $M_i/|M_i'| = \tau_i\omega_i/|e_i^*|$ is the level of Johnson's optimal tariff and is referred to as *monopoly power effect* in Horn et al. (2010). I conclude the *monopoly power effect* can be decomposed into *trade volume* effect (denoted by $aa^*X_i'M_i$) and the effect of *price sensitivity of import demand* (denoted by $1/(|D_i^{*\prime}|\tau_i^* + |M_i'|\tau_i)$).

Similarly, if Home as an importing country exerts a tariff, or assuming an *effective constraint* on import subsidy in Home, it is easy to get

$$B_i^* = \frac{aa^*X_i^{*\prime}}{|D_i'|\tau_i + |M_i^{*\prime}|\tau_i^*}[|M_i^*| + (\tau_i - 1)\omega_i\tau_i|D_i'|]. \tag{2.13}$$

Looking closer at equations (2.12) and (2.13) leads the following proposition:

Proposition 2. *It is optimal to leave discretion over production subsidies if: (i) the* trade volume *effect on the global political welfare gains of constraining production subsidies is positive and the trade volume is sufficiently small, or (ii) the* price sensitivity of supply *is sufficiently small, or (iii) the* price sensitivity of import demand *is sufficiently large.*

Proposition 2 summarizes three circumstances under which the gains constraining s_i and s_i^* brings are so small that they may not offset the accompanying contracting costs and thus it is better to omit s_i and s_i^* from the trade agreement. First, B_i (B_i^*) will be small if M_i ($|M_i^*|$) has a positive effect on B_i (B_i^*) and is sufficiently small. This is the case when

effective constraints are set on trade-promoting border policies, Home (Foreign) has too little trade volume hence gains little to manipulate the terms of trade. Second, B_i (B_i^*) will be small if X_i' ($X_i^{*\prime}$) is sufficiently small. This indicates low price sensitivity of supply, which, as equation (2.4) predicts, is a condition where a domestic production subsidy is a poor substitute for a tariff.[12] Third, B_i (B_i^*) will be small if $|M_i'|$ ($|M_i^{*\prime}|$) is sufficiently large. This indicates high price sensitivity of import demand in Home (export supply in Foreign), also a condition where a production subsidy is a poor substitute for a tariff, as previously indicated by equation (2.4).

Proposition 2 suggests differential treatment across industries, with respect to production subsidies. Industries with small import volume, and market conditions limit the substitution of production subsidies for tariffs, are more likely to benefit from a trade agreement that does not constrain production subsidies. Nascent industries in developing countries, for example. tend to meet these conditions. The present model is therefore consistent with the *infant industry argument* and provide rationale for the WTO Agreement on Subsidies and Countervailing Measures (SCM Agreement) which offers preferential treatment to those industries in developing countries.

2.4 The Optimal Trade Agreement Based on National Treatment Principle

So far consumption tax is assumed negligible in the two countries, however, they are an important policy instrument. National Treatment (NT), which stipulates equal consumption taxes on domestically produced and imported goods, is a basic principle of GATT/WTO. Assessing the effect of the NT principle requires a broader class of trade agreements which take into account consumption tax.

---

[12]Using an externality framework, Horn et al. (2010) also identified low price sensitivity of supply as a sufficient condition for excluding domestic production subsidy in a trade agreement.

Suppose without the NT provision, each country can implement an internal tax on the consumption of the domestically produced goods and an internal tax on the consumption of the imported goods, respectively, t^h and t^f. In this setting, pricing relationships can be expressed as

$$q_i = \tau_i \omega_i + t_i^f = \left(\tau_i + \frac{t_i^f}{\omega_i} \right) \omega_i,$$

and

$$p_i = \tau_i \omega_i + t_i^f + s_i - t_i^h = \left(\tau_i + \frac{t_i^f}{\omega_i} \right) \omega_i + (s_i - t_i^h).$$

Note that the above two equations are laid out such that the term $\tau_i + t_i^f / \omega_i$ behaves like τ_i and the term $s_i - t_i^h$ behaves like s_i when no consumption taxes are present. Consequently, a non-NT agreement

$$A^1 = \left\{ \left(\tau_i + \frac{t_i^f}{\omega_i} \right) = \left(\tau_i^* + \frac{t_i^{f*}}{\omega_i} \right), s_i - t_i^h = \frac{I_{iL} - \alpha_L}{a + \alpha_L} \frac{p_i}{\eta_i}, s_i^* - t_i^{h*} = \frac{I_{iL}^* - a_L^*}{a^* + \alpha_L^*} \frac{p_i^*}{\eta_i^*} \right\}$$

has $n_p = 8n$ and $n_s = 4 + 4n$ and therefore costs $c(8n, 4 + 4n)$.

When the NT provision is included in trade agreements, however, $t_i^f = t_i^h = t_i$. So these relationships become

$$q_i = \tau_i \omega_i + t_i,$$

and

$$p_i = \tau_i \omega_i + s_i.$$

Not surprisingly, the consumption tax does not affect the relationship between the world and producer prices but does affect the relationship between world and consumer prices. Therefore, while it is possible to reduce the wedge between producer and world prices (by reducing τ and s) and leave consumption taxes to discretion in an NT-based agreement, this is not the case in the absence of the NT principle.

The question to be answered is under what circumstances is it desirable to include

the NT provision while leaving consumption taxes to discretion. First, observing that an

agreement

$$A^2 = \left\{ NT, \tau = \tau^*, s_i = \frac{I_{iL} - \alpha_L}{a + \alpha_L} \frac{p_i}{\eta}, s_i^* = \frac{I_{iL}^* - a_L^*}{a^* + \alpha_L^*} \frac{p_i^*}{\eta^*}, t = t^* \right\},$$

where NT represents the NT principle and is equivalent to using $4n$ policy instruments

$(t^h=t^f$ and $t^{h*} = t^{f*})$, has $n_p = 10n$ and $n_s = 4 + 4n$ and therefore costs $c(10n, 4 + 4n)$.

It is straightforward to see that the NT-based agreement A^2 can realize the same $\mathbb{E}(\Omega)$ as

non-NT agreement A^1, but costs more, so does not qualify as an optimal trade agreement.

Consider the following NT-based agreement

$$A^3 = \left\{ NT, \tau = \tau^*, s_i = \frac{I_{iL} - \alpha_L}{a + \alpha_L} \frac{p_i}{\eta}, s_i^* = \frac{I_{iL}^* - a_L^*}{a^* + \alpha_L^*} \frac{p_i^*}{\eta^*} \right\}.$$

A^3 saves on contracting costs as a result of excluding policy variables t and t^* but may

result in a reduction of gains from the agreement because of possible distortions caused by

leaving t and t^* unconstrained.

Again, the noncooperative choice of t conditional on P and P^* can be denoted as

$t^N(P, P^*)$ and the efficient level of t conditional on P and P^* as $t^E(P, P^*)$. The gain in

$\mathbb{E}(\Omega)$ implied by substituting $t^E(P, P^*)$ and $t^{*E}(P, P^*)$ for $t^N(P, P^*)$ and $t^{*N}(P, P^*)$ then is

the extra gain of constraining consumption taxes in an NT-based trade agreement, and can

be expressed as

$$\Omega\left(t_i^E(P, P^*), t_i^{*E}(P, P^*), P, P^*, \cdot\right) - \Omega\left(t_i^N(P, P^*), t_i^{*N}(P, P^*), P, P^*, \cdot\right)$$

$$= \int_{t_i^N(P, P^*)}^{t_i^E(P, P^*)} \frac{\partial \Omega}{\partial t_i}(t, t^*, P, P^*) \, dt_i + \int_{t_i^{*N}(P, P^*)}^{t_i^{*E}(P, P^*)} \frac{\partial \Omega}{\partial t_i^*}(t, t^*, P, P^*) \, dt_i^*.$$

Following steps similar to those in last section, I observe that a sufficient condition for this

gain in $E(\Omega)$ to be small is that $|\frac{\partial \Omega}{\partial t_i}\left(t_i^N(P, P^*), t_i^{*N}(P, P^*), P, P^*, \cdot\right)|$ and $|\frac{\partial \Omega}{\partial t_i^*}\left(t_i^N(P, P^*), t_i^{*N}(P, P^*), P, P^*, \cdot\right)|$

are small. Letting

$$|\frac{\partial \Omega}{\partial t_i}\left(t_i^N(P,P^*),t_i^{*N}(P,P^*),P,P^*,\cdot\right)|=Z_i$$

and after some manipulation it is easy to get

$$Z_i = \frac{a|D_i'|}{|M_i'|\tau_i+X_i^{*\prime}\tau_i^*}|I_{iL}^*X_i^*\tau_i^* - a^*X_i^{*\prime}\tau_i^*[(\tau_i^*-1)\omega_i+s_i^*]+a^*M_i| \qquad (2.14)$$

Based on equation (2.14), my discussion of whether t should be constrained by a NT-based trade agreement can be summarized by the following proposition:

Proposition 3. *It is optimal to include the NT clause while leaving consumption taxes to discretion if: (i) the* trade volume *effect on the global political welfare gains of constraining consumption taxes is positive and the trade volume is sufficiently small, or (ii) the* price sensitivity of demand *is sufficiently small, or (iii) the* price sensitivity of import demand *is sufficiently large.*

Firstly, if Home is the net importer in industry i, as equation (2.14) indicates, Z_i is small when $|D_i'|$ is sufficiently small, meaning low price sensitivity of demand, or when $|M_i'|$ is sufficiently large, meaning high price sensitivity of import demand.[13] In either case, t_i is a poor substitute for τ_i, and the benefits of including t_i in the NT-based agreement may be too small to offset accompanying contracting costs and thus it is optimal to exclude t_i from the NT-based trade agreement.

Secondly, once again, unlike Horn et al. (2010), it is difficult to determine the *trade volume* (M_i) effect, as the sign of the term $I_{iL}^*X_i^*\tau_i^* - a^*X_i^{*\prime}\tau_i^*[(\tau_i^*-1)\omega_i+s_i^*]+a^*M_i$ is ambiguous due to the possible state of the world. However, suppose Home is the net importer in industry i, if s_i^* and τ_i^*-1 are constrained to be negative or small enough such that $I_{iL}^*X_i^*\tau_i^* - a^*X_i^{*\prime}\tau_i^*[(\tau_i^*-1)\omega_i+s_i^*]+a^*M_i$ is positive, a situation which I refer to as an

---

[13]Using an externality framework, Horn et al. (2010) also identified low price sensitivity of demand as a sufficient condition for excluding consumption tax in a NT-based trade agreement.

effective constraint on the production and export subsidy in Foreign, then

$$Z_i = \frac{a|D_i'|}{|M_i'|\tau_i + X_i^{*\prime}\tau_i^*} \left\{ I_{iL}^* X_i^* \tau_i^* - a^* X_i^{*\prime}\tau_i^*[(\tau_i^* - 1)\omega_i + s_i^*] + a^* M_i \right\}.^{14}$$

Note, if the trade volume M_i is sufficiently small then in this situation it is optimal to exclude t_i from the NT-based trade agreement. Similarly, I can assume an *effective constraint* on the production tax ($|s_i|$) and the import subsidy in Home such that $-I_{iL}X_i\tau_i + aX_i'\tau_i[(\tau_i - 1)\omega_i + s_i] + a|M_i^*|$ is positive, then

$$Z_i^* = \frac{a^*|D_i^{*\prime}|}{|M_i^{*\prime}|\tau_i^* + X_i'\tau_i} \left\{ -I_{iL}X_i\tau_i + aX_i'\tau_i[(\tau_i - 1)\omega_i + s_i] + a|M_i^*| \right\}.$$

Again, if trade volume $|M_i^*|$ is sufficiently small then it is optimal to exclude t_i from the NT-based trade agreement.

To summarize, proposition 3 identifies sufficient conditions under which it is optimal to include an NT clause without specifying particular consumption taxes. This helps explain the existence of the NT clause in the current WTO, where significant constraints are placed on subsidies and tariffs while internal consumption taxes are largely left to discretion. For sectors where trade volumes are little and *effective constraints* are already placed on trade-promoting production and trade policies, or consumption taxes are a poor substitute policy instrument for tariffs, it is attractive to leave consumption taxes to discretion while applying NT principle.

---

[14] This equation can also be reformulated as $\frac{a|D_i'||M_i'|}{|M_i'|\tau_i + X_i^{*\prime}\tau_i^*} \left\{ \frac{I_{iL}^* X_i^* \tau_i^* - a^* X_i^{*\prime}\tau_i^*[(\tau_i^*-1)\omega_i + s_i^*]}{|M_i'|} + \frac{a^* M_i}{|M_i'|} \right\}$, where $M_i/|M_i'| = \tau_i\omega_i/|e_i^*|$ is the level of Johnson's optimal tariff and is referred to as *monopoly power effect* in Horn et al. (2010). So once again I conclude the *monopoly power effect* can be decomposed into *trade volume* effect and the effect of price sensitivity of import demand.

2.5   Conclusion

In this manuscript I have incorporated both political pressure and contracting costs in analyzing trade agreements. Like many previous studies in the political economy literature (Hillman, 1982; Snyder Jr, 1990; Grossman and Helpman, 1994, 1995a,b), I view governments as agents that maximize their own interests in response to political pressure rather than as benevolent agents that maximize aggregate social welfare. Grossman and Helpman (1994) brought a first coherent theoretical model of endogenous trade policy formation and concluded tariff rates are affected both by a *political support* motive and a *terms-of-trade* motive. However, an implicit assumption of the mainstream political economy model is that less distortionary domestic policies are not available for redistribution, which is clearly not the case in current policy mechanism. Schleich and Orden (1996) included both trade and domestic production policies in their political economy model and concluded that production subsidies substitute for trade policies that would have otherwise resulted from rent seeking efforts. As a result, when countries act noncooperatively, tariff rates are exactly Johnson's optimal tariff rates, which represents *only* the *terms-of-trade* motive. The proposed political economy model goes one step further by identifying the efficient policy choices in a cooperative equilibrium, or a costless trade agreement.

This model provides political economy rationale for countervailing duty law by showing that a costless trade agreement would lead to equal tariff rate in the importing country and export subsidy rate in the exporting country, since production subsidies can be used to redistribute and meet interest groups' demand for rent. It is also found cooperative production subsidy rates are the same as those in a noncooperative equilibrium confirming that a trade agreement which constrains production subsidies but not tariffs is not optimal, an important finding proposed by Horn et al. (2010) using externality as the rationale for trade agreements.

Horn et al. (2010) propose a *monopoly power* effect, a *trade volume* effect and an

instrument substitutability effect as the key features of the circumstances that determine the benefit of constraining domestic production subsidies and that of constraining consumption taxes in an NT-based trade agreement. They use externalities as the rationale for trade agreements and the source of uncertainty. They find that an optimal trade agreement should at least make the tariff level contingent on the consumption externality. Unfortunately, externality is difficult to measure and test.

Like Horn et al. (2010), this model predicts that uncertainty induces a trade-off between contracting costs and including more policy and state variables in a trade agreement. The uncertainty over production and consumption externalities is replaced with changing rent-seeking conditions. This replacement yields additional insights into the mechanisms that drive trade agreements. For example, if contracting is costless, an optimal trade agreement should make constraints contingent on variables representing the changing policy environment during the lifetime of the agreement, this, I think, is more convincing and empirically plausible.

I decompose the *monopoly power* effect identified in Horn et al. (2010) into a *trade volume* effect and the effect of *price sensitivity of import demand*, this decomposition is simple yet yields new interpretation for empirical investigation. In addition, the positive *trade volume* effect on the gains of constraining production policy or consumption policy proposed by Horn et al. (2010) is identified as a special case where the government assign zero to the weight of political contribution or *effective constraints* are already set on trade-promoting policies.

Chapter 3

Modelling Regime-Dependent Agricultural Commodity Price Volatilities

3.1 Introduction

Agricultural commodity price volatility has been exceptionally high during the last decade. (FAO and UNCTAD (2011)). Large and unpredictable price variations create a level of uncertainty which increases risks for producers, traders, consumers and governments. Using options and forward contracts to manage risk is more costly for producers and processors when prices are exceptionally volatile. Furthermore, large price uncertainty raises risks to investment and production decisions, particularly where the physical production cycle is long. In addition, volatile prices pose significant problems for market regulators and governments as they need greater skills to manage markets in a volatile state, this is especially the case in underdeveloped countries where households may suffer severe food scarcity and food security problems. Volatile prices also affect consumers as it becomes more diffi-

cult for them to make proper consumption decisions when the discretionary income left is uncertain.

The challenge that high commodity price volatility brings highlights the need to better understand its causes, patterns, impacts and measures available to mitigate them. Modelling commodity price volatility helps to forecast the absolute magnitude, quantiles, and in fact, the entire distribution of price changes. Such forecasts are widely used in risk management, derivative pricing and hedging, portfolio selection, among other economic activities.

A better fitting discrete time volatility model can help producers and traders manage price risk and make better production and investment decisions in a volatile market state. When budgeting for production levels, a producer of one commodity may want to know today the likelihood that the price of the product will decline in the future. To hedge a contract, an option trader will want to know the the expected volatility over the life of the contract, as well as how volatile is this option-implied volatility. To make a profit and limit their losses to a comfortable level, a futures trader may want to use a sell or buy stops based on volatility and density prediction.

Modelling commodity price volatility following price spikes is essential for market regulators and governments, particularly when prices are highly volatile. A proper volatility model help monitor and predict the movement of commodity prices. Market regulation and other policies that aims to manage the risks and mitigate the impacts of high volatility can be developed accordingly.

While there is a large body of literature on price volatility since the introduction of generalized autoregressive conditional heteroscedasticity (GARCH) framework by Engle (1982) and Bollerslev (1986), the usual assumption in the majority of the proposed volatility models applied to commodity prices is that the error process is conditionally normally distributed. However, normal GARCH models do not adequately capture heavy tails, large kurtosis, and occurrence of extreme events, which are often key features of financial

time series. As an extension, Bollerslev (1987) proposed modelling the innovations via a GARCH model with a Student's t-distribution and Fernández and Steel (1998) extended it further considering the skewed t-distribution.

A few recent studies have proposed using a mixture of two normal distributions to model volatility. Among them, Haas et al. (2004) introduced the general symmetric Normal Mixture(NM) GARCH model, and Alexander and Lazar (2006) further investigated the property of NM-GARCH(1,1) model and provided empirical evidence that the generalized two-component NM-GARCH(1,1) models perform better than both symmetric and skewed Student's t-GARCH models for modelling exchange rates. A clear advantage of the NM-GARCH model over Students' t-GARCH models is the capability to model time-varying conditional skewness and kurtosis. Another advantage of NM-GARCH model is that it accounts for multiple states, which may contribute to economic interpretation. Haas et al. (2004), for example, pointed out that an NM-GARCH model accommodates the possibility of distinct types of responses to heterogeneous market shocks. Alexander and Lazar (2009) argued that a component with relative low variance could represent a "usual" state, which generally occurs, while a component with high variance could represent a "crash" state which rarely occurs.

An important empirical regularity of equity markets is the fact that volatility increases more after price declines than after price increases by the same magnitude, such an asymmetric return-volatility relationship is documented as a financial leverage effect in early influential studies (Black, 1976; Christie, 1982; Engle and Ng, 1993; Glosten et al., 1993). Glosten et al. (1993) introduced the GJR-GARCH(1,1) model allowing unequal response weight for negative and positive shocks. In commodity markets, contrary to equity markets, an "inverse leverage effect" may exist, i.e., a rise in the price level has stronger impact on the price volatility than a drop in the price. This is understandable, as increased prices of commodities generally bring panic and give rise to higher volatility. Previous studies, such

as Geman and Shih (2009) and Chang (2012) found such an effect in energy markets. This effect has not been considered with respect to agricultural commodity markets.

One contribution of this paper is to test whether NM-GARCH models are appropriate for modelling and forecasting agricultural commodity price volatility. Out-of-sample interval forecast validation, though pivotal in risk management and policy-making, is rarely applied in previous volatility modelling for agricultural commodity prices. The present work also try to fill this gap by performing Value-at-Risk (VaR) validation tests. In addition, in order to capture the possible state-specific asymmetric volatility responses to negative and positive shocks, this manuscript followed Alexander and Lazar (2009) and considered the NM-GJR-GARCH model.

3.2 Literature Review

ARCH family, as a sophisticated group of time series volatility models, has been extensively surveyed by Bollerslev et al. (1992); Bera and Higgins (1993); Poon and Granger (2003). The seminal paper of Engle (1982) captured volatility clustering and heavy tails that are two stylized facts in financial time series data. Bollerslev (1986) introduced a generalized version of ARCH which reduces the number of parameters to be estimated by imposing autoregressive terms. Since then, numerous extensions have been made to GARCH models to capture asymmetry, long memory, structural breaks and regime switching behaviours in financial market data. Haas et al. (2004), among others, proposed extending the basic GARCH structure by assuming the conditional distribution of the error term as a mixture of normal distributions. NM-GARCH, though simple to estimate, is able to capture three regularities in financial asset returns: volatility clustering, heavy tails and time-varying skewness.

Time-varying volatility is also a stylized fact observed in agricultural commodity price data. The empirical research on agricultural price volatility has focused on the depen-

dence of price volatility across related markets (Apergis and Rezitis, 2003; Buguk et al., 2003; Rezitis and Stavropoulos, 2010; Serra et al., 2011; Serra and Gil, 2013; Serra, 2013) and determinants of price volatility (Shively, 1996; Hennessy and Wahl, 1996; Karali and Power, 2013). For example, using a multivariate GARCH model with exogenous variables incorporated in the conditional covariance model, Serra and Gil (2013) found U.S. corn price volatility could be explained by volatility clustering, the influence of biofuel prices, corn stocks and global economic conditions. Karali and Power (2013) explained price volatility in the U.S. commodity futures markets, using a spline-GARCH model of Engle and Rangel (2008) that produces estimates of low-frequency volatility. Estimates are then regressed against a series of macroeconomic variables. This empirical study is based on 11 different daily futures prices observed from April 1990 to November 2009. The U.S. Treasury interest rate spread (10-year to 2-year) is found to have negative impact on price volatility for corn, crude oil, heating oil and hopper, with the largest effect for crude oil. Working's theory of storage, whereby volatility is decreasing in inventories, is supported for corn, wheat, lean hogs, and crude oil.

The number of empirical tests of structural models in agricultural commodity prices is surprisingly limited. Hall et al. (1989) detected unconditional leptokurtic distribution in twenty daily futures price series and found support for the normal mixture distribution hypothesis relative to a stable Paretian distribution hypothesis by applying the stability-under-addition test. Yang and Brorsen (1992) were among the first to empirically test for a GARCH structure in agricultural commodity prices. They found that GARCH models with a conditional Student's t-distribution fit daily price change data better than a number of alternatives, however, both the Student's t distribution and the normal did not correctly specify the conditional distribution according to the Kolmogorov-Smirnov test. Jin and Frechette (2004) found fractionally integrated generalized autoregressive conditional heteroscedastic (FIGARCH) model performs significantly better than the basic GARCH(1,1)

models in modelling volatility of 14 agricultural futures price series, confirming long-term memory of volatility. They explained many factors can lead to long-term dependence in agricultural futures price volatility, such as supply lags, inventory holding, business cycles, agricultural policies and heterogeneity among traders.

The majority of the previous research suggest GARCH model with a conditional normal distribution or Student's t-distribution does not adequately model the agricultural commodity prices. Jin and Frechette (2004)'s finding support FIGARCH model over the basic GARCH(1,1) models in modelling volatility with long-term memory. As with other single-state models, FGARCH model can not capture state-dependent volatility dynamics and is subject to the stringent assumption of constant skewness and kurtosis. Alternatively, the persistence in commodity price volatility can also be modlled by the GARCH part of the NM-GARCH model. of In fact, the causes that lead to persistent price volatility as listed by Jin and Frechette (2004) also contribute to a multi-regime market and regime-dependent volatility dynamics. On the one hand, supply lags and business cycles may lead to incidences of different market states, on the other hand, agricultural policies, inventory holding and trade behaviours tend to be different under stable and turbulent price environments. Therefore it is interesting to access whether an NM-GARCH(1,1) model allowing for state-dependent volatility dynamics adequately captures the relevant properties of agricultural commodity prices.

3.3   Model and Data

The innovations, denoted by the error term ε_t, is assumed to follow a mixture of k Gaussian distributions with distinct component mean μ_i and component variance σ_{it}^2. That is,

$$\varepsilon_t | \Omega_{t-1} \sim \text{NM}(p_1, \ldots, p_k, \mu_1, \ldots, \mu_k, \sigma_{1t}^2, \ldots, \sigma_{kt}^2),$$

where Ω_t is the information set at time t, $p_i \in (0,1)$, $i = 1,\ldots,k$ are mixing weights, $\sum_{i=1}^{k} p_i = 1$ and $\sum_{i=1}^{k} p_i \mu_i = 0$. I consider two possibilities for the conditional variance of k components.

(i) NM(k)-GARCH(1,1):

$$\sigma_{it}^2 = \omega_i + \alpha_i \varepsilon_{t-1}^2 + \beta_i \sigma_{it-1}^2 \qquad \text{for } i = 1,\ldots,k, \tag{3.1}$$

where α_i is defined as the volatility reaction parameter, indicating the effect of market shocks on volatility, and β_i is defined as the volatility persistence parameter, referring to the extent of inertia in volatility.

(ii) NM(k)-GJR-GARCH(1,1):

$$\sigma_{it}^2 = \omega_i + \alpha_i \varepsilon_{t-1}^2 + \lambda_i d_{t-1}^- \varepsilon_{t-1}^2 + \beta_i \sigma_{it-1}^2 \qquad \text{for } i = 1,\ldots,k, \tag{3.2}$$

where $d_t^- = 1$ if $\varepsilon_t < 0$, and 0 otherwise and λ_i is the leverage parameter.

Both the NM-GARCH and NM-GJR-GARCH model allow for different non-zero component means, thus capturing overall unconditional or persistent asymmetry in the data. As the NM-GJR-GARCH model includes a leverage parameter λ_i, it is able to capture state-dependent dynamic asymmetry in the data. For example, a negative λ_i indicates the conditional variance in this regime tends to be higher following a price increase than a price decrease. In commodity markets, an "inverse leverage effect" or a negative value of the leverage parameter is expected because a rise in commodity prices generally brings panic and gives rise to higher volatility.

This study analyzes weekly cash prices of three grains, four meat and three dairy products obtained from Livestock Marketing Information Center (LMIC). Because of the data

availability, the time period across commodities are different. Specifically, I consider the following agricultural commodities:

(i) grains: corn, sorghum and wheat weekly cash price series for the January 1988 to July 2013 period (1332 observations);

(ii) meat: beef weekly cash prices for the July 1999 to July 2013 period (758 observations) , pork weekly cash prices for the January 1988 to April 2013 period (795 observations), broiler and turkey weekly cash prices for the January 1992 to December 2012 period (991 observations).

(iii) dairy products: cheddar, butter and nonfat dry milk (NFDM) for the September 1998 to February 2013 period (753 observations).

For each commodity, I fit the continuously compounded percentage changes of prices, $r_t = 100(\log P_t - \log P_{t-1})$ with an autoregressive-moving-average (ARMA(u,v)) model.

$$r_t = c + \varepsilon_t + \sum_{i=1}^{u} a_i r_{t-i} + \sum_{j=1}^{v} b_j \varepsilon_{t-j}$$

An Akaike information criterion with a correction for finite sample sizes (AICc) is used to select the appropriate values of u and v. Then I subtract the means of each series and perform estimation of the NM-GARCH models by the expectation-maximization (EM) algorithm of Dempster et al. (1977).

3.4   Estimation Results and Implications

The GARCH(1,1), the NM(2)-GARCH(1,1), and the NM(2)-GJR-GARCH(1,1) models are estimated for each of the food price series. The estimation results are given in Tables 3.1–3.3.

Table 3.1: Estimation result for grains

	p_1	μ_1	ω_1	α_1	β_1	λ_1	p_2	μ_2	ω_2	α_2	β_2	λ_2	$E\sigma_1$	$E\sigma_2$
Corn														
Normal			0.5443	0.2017	0.7828								5.9376	
			(3.3)[a]	(7.31)	(30.58)									
NM	0.6525	0.3406	0.6047	0.089	0.7731		0.3475	-0.6393	0.4147	0.4102	0.7867		4.0387	8.2998
	(14.17)	(2.94)	(2.06)	(1.89)	(9.46)		(6.21)		(0.94)	(19.56)	(18.83)			
GJR	0.6519	0.3766	0.5988	0.0949	0.7715	-0.0042	0.3481	-0.7053	0.2213	0.5447	0.801	-0.2996	4.922	10.33
	(18.72)	(2.91)	(4.38)	(1.91)	(8.87)	(-0.1)	(7.27)		(0.63)	(28.11)	(62.45)	(-2.52)		
Sorghum														
Normal			0.359	0.1374	0.8551								6.9143	
			(3.01)	(7.63)	(50.6)									
NM	0.7221	0.0245	0.2342	0.1219	0.7612		0.2779	-0.0636	0.8769	0.2485	0.8766		4.4749	9.0645
	(17.21)	(0.28)	(2.13)	(3.92)	(18.03)		(9.04)		(1.88)	(4.56)	(36.05)			
GJR	0.7123	0.015	0.2149	0.1408	0.7614	-0.038	0.2877	-0.0371	0.731	0.2932	0.8872	-0.1385	4.519	9.079
	(18.44)	(0.17)	(2.07)	(3.63)	(17.62)	(-0.99)	(11.87)		(1.98)	(6.00)	(41.48)	(-1.51)		
Wheat														
Normal			1.0543	0.1306	0.7954								3.774	
			(3.43)	(4.84)	(19.03)									
NM	0.881	-0.0425	0.046	0.0329	0.9506		0.119	0.3145	7.0006	0.6344	0.608		3.1429	6.2922
	(13.68)	(-0.62)	(1.69)	(4.37)	(79.04)		(1.69)		(2.35)	(2.25)	(5.45)			
GJR	0.8304	-0.0503	0.0257	0.0316	0.9545	-0.0024	0.1696	0.2464	5.911	1**	0.5609	-0.8082	3.031	5.564
	(5.88)	(-0.58)	(0.33)	(2.10)	(45.59)	(-0.18)	(1.21)		(1.21)	(1.11)	(2.95)	(-0.99)		

[a] Numbers in parentheses represent t-values.
[b] Boundary value (1) is reached, making estimate uninformative.

Table 3.2: Estimation result for meat

	p_1	μ_1	ω_1	α_1	β_1	λ_1	p_2	μ_2	ω_2	α_2	β_2	λ_2	$E\sigma_1$	$E\sigma_2$
Beef														
Normal			0.1134	0.0782	0.8925								1.9683	
			(2.07)[a]	(3.23)	(27.05)									
NM	0.6686	-0.5289	0.0269	0.0353	0.9205		0.3314	1.067	0.0703	0.1901	0.858		1.405	2.33
	(3.63)	(-4.57)	(0.68)	(1.76)	(21.43)		(1.77)		(0.42)	(1.47)	(10.52)			
GJR	0.6294	-0.4161	0.2792	0.0559	0.6165	0.1945	0.3706	0.7068	0.0738	0.148	0.9234	-0.1225	1.44	2.186
	(5.06)	(-3.92)	(2.02)	(1.67)	(4.53)	(1.96)	(2.95)		(0.81)	(1.60)	(16.52)	(-1.43)		
Pork														
Normal			0.7943	0.1219	0.7961								3.1116	
			(2.16)	(3.27)	(12.01)									
NM	0.5462	0.3085	0.2217	0.0523	0.8859		0.4538	-0.3713	1.6591	0.2176	0.7183		2.5229	3.6521
	(1.57)	(0.88)	(0.86)	(1.45)	(14.32)		(1.31)		(1.26)	(4.00)	(6.02)			
GJR	0.5452	0.5607	0.0972	0[b]	0.943	0.0663	0.4548	-0.6721	2.336	0.4779	0.5223	-0.2994	2.689	3.381
	(30.76)	(131.2)	(21.28)	(0.00)	(73.64)	(2.94)	(25.26)		(931.6)	(18.98)	(17.16)	(-20.85)		
Broiler														
Normal			2.142	0.1884	0.1865								1.851	
			(4.68)	(4.36)	(1.31)									
NM	0.5074	0.0285	4.0708	0.2587	0.1008		0.4926	-0.0294	0.2035	0.1093	0.5512		2.3484	1.1357
	(4.84)	(0.38)	(2.96)	(2.37)	(0.44)		(4.64)		(0.93)	(1.05)	(1.69)			
GJR	0.504	0.0389	4.746	0.1149	0	0.1679	0.496	-0.0395	0.2487	0.171	0.4676	-0.0596	2.329	1.172
	(5.27)	(0.52)	(7.31)	(0.75)	(0.00)	(0.97)	(5.13)		(1.42)	(1.94)	(1.76)	(-1.91)		
Turkey														
Normal			1.365	0.3413	0.5259								3.206	
			(4.91)	(6.08)	(8.56)									
NM	0.7158	0.2994	0.4073	0.2083	0.3938		0.2842	-0.754	3.3772	0.4487	0.6519		1.9342	4.6056
	(13.71)	(5.04)	(2.67)	(4.53)	(4.14)		(6.03)		(1.83)	(2.23)	(4.62)			
GJR	0.7099	0.3003	0.4616	0.1714	0.358	0.0884	0.2901	-0.735	4.627	0.2639	0.5185	0.7576	1.968	4.707
	(13.15)	(4.92)	(2.72)	(3.30)	(3.58)	(1.06)	(5.93)		(2.30)	(1.45)	(3.36)	(1.45)		

[a] Numbers in parentheses represent *t*-values.
[b] Boundary value (0) is reached, making estimate uninformative.

Table 3.3: Estimation result for dairy products

	p_1	μ_1	ω_1	α_1	β_1	λ_1	p_2	μ_2	ω_2	α_2	β_2	λ_2	Eσ_1	Eσ_2
Cheddar														
Normal			0.6282	0.3187	0.5335								2.0621	
			(4.13)[a]	(5.54)	(8.42)									
NM	0.5171	-0.04	1.0191	0.5581	0.5332		0.4829	0.0428	0.2302	0.072	0.4751		2.6963	1.0109
	(6.45)	(-0.60)	(2.62)	(3.38)	(5.61)		(5.99)		(2.28)	(2.14)	(3.47)			
GJR	0.518	-0.0189	1.08	0.4568	0.5273	0.1611	0.482	0.0204	0.2594	0.0408	0.4317	0.0756	2.631	1.01
	(6.34)	(-0.26)	(2.43)	(2.4)	(5.02)	(0.76)	(5.89)		(2.13)	(0.89)	(3.41)	(0.99)		
Butter														
Normal			0.5386	0.1346	0.7978								2.8241	
			(2.15)	(3.06)	(11.87)									
NM	0.7283	0.0653	1.1227	0.2744	0.2835		0.2717	-0.175	0.1425	0.104	0.9425		2.1484	4.1088
	(10.23)	(0.66)	(3.04)	(4.87)	(1.86)		(3.70)		(0.48)	(2.70)	(36.66)			
GJR	0.7487	0.0221	0.8666	0.1689	0.4055	0.2145	0.2513	-0.0659	0	0.1591	0.9681	-0.1813	2.435	4.545
	(5.74)	(0.19)	(1.82)	(2.55)	(3.38)	(2.26)	(1.68)		(0.27)	(3.33)	(73.77)	(-2.04)		
NFDM														
Normal			0.1583	0.2608	0.7584								2.8672	
			(7.10)	(6.08)	(31.34)									
NM	0.8969	0.0223	0.0198	0.1778	0.6607		0.1031	-0.1936	1.1979	0.8614	0.8221		3.5869	11.1856
	(24.45)	(1.17)	(3.98)	(3.87)	(15.06)		(8.81)		(2.87)	(11.89)	(26.47)			
GJR	0.9071	0.0239	0.0204	0.2063	0.6582	-0.0402	0.0929	-0.2337	1.349	0.3997	0.8171	1[b]	2.882	9.046
	(24.82)	(1.25)	(3.96)	(4.20)	(17.48)	(-0.79)	(25.86)		(3.01)	(0.43)	(28.15)	(0.50)		

[a] Numbers in parentheses represent t-values.
[b] Boundary value (1) is reached, making estimate uninformative.

3.4.1 Diagnostic Checks and Forecasting Performance

To assess the in-sample fit provided by the three models, I have applied several model selection criteria. First, I test the normality of the standardized residuals. As standardized residuals of GARCH-type models may not be identically distributed, I proceed with a transformation pioneered by Berkowitz (2001) and extended to NM-GARCH model testing by Haas et al. (2004) and Alexander and Lazar (2009). Specifically,

$$z_t = \Phi^{-1}\left(\hat{F}(\varepsilon_t|\Omega_{t-1})\right), \qquad (3.3)$$

where Φ^{-1} is the inverse function of standard normal cumulative distribution function, and $\hat{F}(\cdot)$ is the conditional distribution function of the error term ε_t. If the model correctly specifies the underlying data generating process (DGP), then the transformed residuals z_t's should be identically independently distributed standard normal variables. As noted by Berkowitz (2001), the transformed residuals would preserve inaccuracies in the specified density, therefore Equation (3.3) can be used to check correct specification of moment features such as skewness and kurtosis. Specifically, let T be the sample size, g_1 denotes the sample skewness of z_t and g_2 the sample kurtosis, if z_t's are normally distributed, then $m_1 = Tg_1^2/6 \overset{asy}{\sim} \chi^2(1)$ and $m_2 = T(g_2-3)^2/24 \overset{asy}{\sim} \chi^2(1)$. In addition, the following Jarque and Bera (1987) (JB) test is implemented to check the normality of the transformed series z_t.

$$JB = m_1 + m_2 \overset{asy}{\sim} \chi^2(2).$$

Table 3.4 summarizes the results for the in-sample fit. Results show that the normal GARCH model fails the skewness and/or kurtosis tests for all commodities except for pork. The JB normality test results further show that transformed residuals of the normal GARCH models for all price series except pork exhibit strong deviations from normality. However even for pork, NM-type models have smaller JB-statistics indicating a better job of mod-

elling. The performance of the NM-GARCH model and the NM-GJR models are comparable and consistently well for most commodities, indicating time-varying conditional skewness and kurtosis specification is necessary.

As volatility models are widely employed in risk management, I also assess the accuracy of Value-at-Risk (VaR) predictions. VaR is defined as the conditional τ-quantile, $\Pr(y_t \leq \mathrm{VaR}_t(\tau)|\Omega_{t-1}) = \tau$, where τ is also defined as shortfall rate or failure rate, representing the probability that the loss exceeds the VaR threshold. It is widely used to measure the downside risk on a specific portfolio of financial assets. Although many VaR backtesting criteria having been proposed, no consensus has been reached about the best method. Thus I employ two VaR backtesting methods in this manuscript.

For in-sample VaR, I follow Alexander and Lazar (2009) and use the conditional coverage test introduced by Christoffersen (1998). The hypotheses are that the realization of the variable lies outside the $(1-\tau) \times 100\%$ forecast interval $\tau \times 100\%$ of the time, and such violations should also be independent across time. In the case of VaR, the intervals are one-sided from the threshold value $\mathrm{VaR}_t(\tau)$ to infinity.

Define $I_t\{r_t < \mathrm{VaR}_t|\Omega_{t-1}\}$, $t = 1,\ldots,T$ as the indicator sequence. A conditional coverage test is a joint test of unconditional coverage test $(E(I_t) = \tau)$ and independent test $(\Pr(I_t = 1|I_{t-1} = 0) = \Pr(I_t = 1|I_{t-1} = 1))$. Because unexpected or prolonged agricultural price spikes typically raise alerts to policy makers and upstream food processors that rely on that commodity as inputs, for example, livestock enterprises are interested to know the highest levels feed prices could arise to in the future, I also assess the accuracy of the upper quantile prediction of the competitive models. The upper tail risk also represents VaR for traders in a short selling position, see Giot and Laurent (2003) for an example of application.[1]

---

[1]Securities or other financial instruments not currently owned are short-sold by traders with the intention of subsequently repurchasing them at a lower price. The short seller incurs a loss when price rises to a higher prices than the proceeds of initial sale.

The last two columns of Table 3.4 report the Christoffersen conditional coverage like-lihood ratio test statistics (LR_{CC}). It is shown that the normal GARCH model fails all VaR(5%) tests whereas the NM-GARCH model passes all the VaR tests and the NM-GJR-GARCH model only fails the test for beef, suggesting that the NM-GARCH and the NM-GJR-GARCH models are suitable for VaR calculation but the normal GARCH model is not. Test results for implied 95% quantile forecasts also confirm the conclusion that the normal GARCH model gives the worst fit. It only correctly predicts the in-sample 95% quantile for beef but fails the interval tests for all other commodities.[2] NM-GARCH and NM-GJR-GARCH models only fail one upper-tail test respectively.

Next, I use a generalized method of moments (GMM) based approach proposed by Dumitrescu et al. (2013) to test out-of-sample forecasting performance of the models with respect to VaR(1%) (in accordance with Basel II requirement) as well as 99% quantile prediction. The GMM based approach test Christoffersen's three validity hypotheses inde-pendently. It has better power and small-sample properties and can always be computed even if there is no violation in the sample, whereas the Christoffersen test requires at least one violation to compute the test statistic.

In this study, the out-of-sample forecasts of VaR's are based on a rolling window es-timation procedure. Firstly, the necessary parameters of the three models are estimated based on the latest 7 years (364 weeks)' observations. The parameters are then fixed for one month (4 weeks) to facilitate out-of-sample interval forecasting. The estimation sample is then rolled ahead in increments of 4 weeks. The estimation and prediction procedure is repeated until the end of the observations. For example, to forecast the innovation distribu-tion of the first 4 weeks of 2013, I use the data from 2006-2012 to estimate the parameters of interest, then in order to forecast the innovation distribution of the fifth-eighth weeks of

---

[2]A VaR(1%) test and 99% quantile test are also proceeded which give similar conclusion but because the normal GARCH forecast is overly cautious, for most commodities there is no violation in the sample and Christoffersen test statistics are not computable.

2013, the estimation sample period is moving forward 4 weeks, that is, from the fifth week of 2006 to the fourth week of 2013.

The results of the GMM conditional coverage test based on two moment conditions and a block size N equal to 25 are shown in Table 3.5. As expected, the normal GARCH method performs rather poorly in the VaR test at failure rate 1% as it fails 6/10 of the tests. The NM-GJR-GARCH model also fails a few tests but gives the most accurate VaR forecast at failure rate 1% for wheat, broiler and butter. The NM-GARCH model achieves the best results for downside risk forecasting. With respect to 99% quantile forecasting, the normal GARCH model only passed the test for butter and nonfat dry milk. The NM-GJR-GARCH model gives the worst 99% quantile prediction for wheat, cheddar and nonfat dry milk, possibly because the model is over-parameterized. The NM-GARCH model achieves the best results for most commodities.

Table 3.4: In-sample Fit Test

	Skewness	Kurtosis	JB	LR_{CC}	
				$\tau = .05$	$\tau = .95$
Corn					
GARCH	-0.133**	0.979***	57.1***	40.38***	47.76***
NM	0.064	0.011	0.9	0.05	4.29
GJR	0.063	-0.026	0.9	0.05	1.89
Sorghum					
GARCH	0.037	6.470***	668.2***	17.55***	20.19***
NM	0.074	0.279**	5.5**	3.96	1.72
GJR	0.056	0.218	3.3	2	0.67
Wheat					
GARCH	0.002	3.060***	519.1***	21.67***	15.11***
NM	0.031	0.146	1.4	0.66	0.7
GJR	0.023	0.179	1.9	0.77	0.67
Beef					
GARCH	0.716***	1.188***	109.4***	25.44***	0.14
NM	0.141	-0.037	2.5	0.48	0.62
GJR	0.231***	-0.167	7.6	4.67*	0.14
Pork					
GARCH	-0.089	0.234	2.9	7.67**	7.4**
NM	0.033	-0.078	0.3	1.08	2.84
GJR	0.012	0.068	0.2	1.08	1.06
Broiler					
GARCH	0.062	1.343***	83.0***	25.34***	19.37***
NM	0.017	0.043	0.1	1.41	5.31*
GJR	0.012	0.067	0.2	1.36	0.34
Turkey					
GARCH	-0.768***	4.462***	919.3***	78.46***	N/A
NM	-0.046	0.105	0.8	1.29	4.56
GJR	-0.036	0.143	1.1	1.46	7.88**
Cheddar					
GARCH	-0.399***	2.313***	187.8***	61.22***	N/A
NM	-0.189 **	0.247	6.4	3.66	3.61
GJR	-0.188 **	0.201	5.7	3.16	0.5
Butter					
GARCH	0.149*	2.823***	252.8***	6.52**	5.04*
NM	0.033	0.129	0.7	2.23	0.1
GJR	0.025	0.032	0.1	1.87	2.17
NFDM					
GARCH	-1.341***	12.81***	5370.8***	20.8***	46.56***
NM	-0.103	0.221	2.9	0.52	0.46
GJR	-0.085	0.194	2.1	0.58	0.49

*, ** and *** indicate values significant at 10%, 5% and 1% significance levels respectively.

Table 3.5: Out-of-sample forecasting test results

τ	Model	Corn	Sorghum	Wheat	Beef	Pork	Broiler	Turkey	Cheddar	Butter	NFDM
1%	GARCH	1.3	6.77**	2.74	2.14	1.86	11.91***	10.21**	48.19***	33.12**	35.36***
	NM	0.85	2.25	2.25	3.59	0.42	1.37	0.53	10.17**	9.95**	11.2 ***
	GJR	5.83*	9.07**	0.46	11.9***	3.68	0.53	1.45	66.62***	0.56	19.31 ***
99%	GARCH	4.93*	21.73***	5.98**	21.59**	8.4**	123.06***	4.72*	23.29***	1.7	1.36
	NM	14.79***	1.3	6.77**	7.61**	0.33	0.53	3.66	19.31***	3.59	0.56
	GJR	14.79***	8.93**	13.33**	19.07***	1.21	3.73	0.53	66.62***	1.36	5.5*

Notes: *, ** and *** indicate values significant at 10%, 5% and 1% significance levels respectively.

In summary, the single-state normal GARCH model performs rather poorly especially with regards to the specification of skewness and kurtosis. The NM-GJR-GARCH model that incorporates different component means and the additional leverage effect is found to fit better than the normal GARCH model but perform badly in out-of-sample forecasting, perhaps because of parameter proliferation. The NM-GARCH model with different component means achieves the best fit by all criteria.

3.4.2 State-dependent Volatility Dynamics

For most commodities, the NM-GARCH model captures a lower-volatility component that occurs with a high probability (the usual regime) and a high-volatility component that occurs with a low probability (the unusual regime). Among them, NFDM has the most unbalanced occurrence of the two market regimes, with the unusual market regime occurring 10% of the time. For broiler and cheddar, however, the two market regime occurred somewhat evenly over time, indicating two-regime model may be inappropriate for these as previously suggested by the in-sample fit diagnostics. A noticeable result regarding the NM-type GARCH models is that similar to Haas et al. (2004), Alexander and Lazar (2006, 2009), and Bauwens et al. (2007), the component that has small mixing weights may have unstable volatility dynamics in the sense that $\alpha_i + \beta_i > 1$.

The usual mean component is lowest and negative in the beef (-.53% per week) but the unconditional volatility is also low: at 1.4%, it is the lowest in the usual regime. On the other hand, corn price has the most expected increase (.34% per week) and the second highest volatility of the ten markets (around 4%, second to Sorghum (4.5%)). In the usual regime the wheat series exhibits the least reactive and most persistent volatility.

In the unusual market regime, NFDM has the highest unconditional volatility (over 10%). Most series and wheat and NFDM in particular, are highly reactive to market shocks in the unusual regime, yet because the persistence are all low, the effect of a shock decays

soon.

There is a clear-cut relationship between the component mean (μ_i) and the component volatility dynamics (reflected by α_i and β_i). For each commodity, expected negative price change corresponds to a greater volatility persistence parameter β_i, indicating volatility tends to be more persistent when shocks are negative.[3] On the other hand, expected positive price changes arise in conjunction with higher volatility reaction parameter (α_i), suggesting volatility is more reactive to price rises than price drops. This is just the opposite of the case in the equity markets, where, for example, Haas et al. (2004) found volatility is more inertia when shocks are positive, while more responsive to negative shocks. Note that the state-dependent volatility dynamics are not detectable in previous research on agricultural commodity prices as single-state GARCH models only capture an average of these effects if multiple states exist.

The NM-GJR model is found to suffer the problem of over parameterization for some commodities, on the grounds that it gives estimates that reach the boundary values in numerical optimization. For the rest of commodities, it gives similar results to the NM-GARCH model. The asymmetric parameters (λ_i) in the NM-GJR for most commodities are insignificant except on occasions when component means are negative. For example, corn has a significant inverse leverage effect in the unusual regime where the price is expected to drop. Beef series, on the other hand, has a significant leverage effect during the usual regime where price falls are expected. A possible explanation is that there are more beef producers who have long interest in their products than physical hedgers, therefore in anticipation of falling prices a realized price drop leads to panic and pushes implied volatility up. The fact that inverse leverage effect is state-dependent (only significant in a regime where negative shock are expected) also permits more refined risk management practice and market regulation in agricultural markets than those based on single-state GARCH

---

[3]A sole exception is pork, the component means and mixing weights of which are not significantly different from 0.

50

models.

3.5   Conclusion

Previous modelling of commodity price volatility assumes a single-state GARCH process and constant conditional skewness and kurtosis, and therefore is not able to detect the state dependent volatility dynamics if multiple states exist. Commodity price volatility may respond differently under different market states, for example, under the expectation of positive and negative price changes. The NM-type GARCH models allow for state-dependent volatility behavior and time-varying conditional skewness and kurtosis. Haas et al. (2004) and Alexander and Lazar (2009), among others, have applied those models in equity markets. This paper models agricultural commodity price volatility using the NM-GARCH models with the assumption of two market states.

Both in-sample and out-of-sample diagnostics are conducted to compare the fit of the NM-GARCH and the NM-GJR-GARCH models with normal GARCH specification. The overall conclusion is that the class of NM-GARCH models adequately captures relevant properties of agricultural commodity price data but the single-state normal GARCH model performs rather poorly especially regarding the specification of skewness and kurtosis. Contrary to the case in the equity market as found in Alexander and Lazar (2009), the addition of dynamic asymmetry in the NM-GJR-GARCH model is sometimes found un-necessary for a few commodities, as it disturbs the time series fit and upper tail prediction.

Empirical results on ten agricultural commodity cash prices show a clear relationship between expected price change and the volatility dynamics across regimes. For each of the ten commodities, an expected negative price change corresponds to a greater volatility persistence, while an expected positive price change arises in conjunction with an increasing responsiveness of volatility. This is just the opposite of the case in the equity market, where Haas et al. (2004) found volatility is more persistent to positive shocks and more

responsive to negative shocks.

Finally, when possible state-dependent "inverse leverage effects" are explicitly accounted for, as in the NM-GJR-GARCH model, it is found that for most commodities these effect are insignificant except on occasions when component means are negative. A significant inverse leverage effect is detected only for corn in a less frequently occurred regime where price falls are anticipated, which indicates the volatility in this regime tends to increase more following a realized price rise than a realized price drop. Conversely, beef is found to have significant leverage effects during the more frequent regime where prices are expected to fall, indicating a realized price fall would lead to higher volatility than a realized price recovery. By allowing state-dependent inverse leverage effects and volatility dynamics, two-state NM-type GARCH models could facilitate more refined risk management practice than single-state GARCH models.

Chapter 4

Nonparametric Estimation of an Additive Unconditional Quantile Regression Model

4.1 Introduction

Suppose Y denotes the response variable that depends on the vector of covariates $\mathbf{X} = (X_1, X_2, ..., X_d)^T$, $d \geq 2$, with T denoting the transpose of a matrix or a vector. Quantile regression considers the case

$$Y = M_\tau(\mathbf{X}) + \varepsilon_\tau, \tag{4.1}$$

where $M_\tau(\mathbf{X})$ is an unknown function and ε_τ is an unobserved random variable. If $M_\tau(\mathbf{X}) = Q_\tau(Y|\mathbf{X} = x) \equiv \inf\{q : F_{Y|\mathbf{X}}(q|x) \geq \tau\}$, $M_\tau(\mathbf{X})$ denotes the conditional quantile of Y given $\mathbf{X} = x$. Koenker and Bassett (1978) introduced a linear model of equation (4.1), where $M_\tau(\mathbf{X}) = \mathbf{X}^T \beta_\tau$. In many practical applications the linear quantile regression model might not fit the available data well. In an effort to make conditional quantile regression models more flexible, there is a growing literature on nonparametric conditional quantile regression

(e.g., Chaudhuri, 1991a,b; Koenker et al., 1994; He and Shi, 1994; Fan et al., 1994; He et al., 1998; Yu and Jones, 1998). Fully nonparametric estimation is usually unattractive in multivariate settings because of the curse of dimensionality. For this reason, dimension reduction modelling methods have been developed. For example, Lee (2003) considered a partially linear conditional quantile regression model and Honda (2004), Kim (2007), and Cai and Xu (2008) adopt varying coefficient models for conditional quantiles. De Gooijer and Zerom (2003), Yu and Lu (2004), Horowitz and Lee (2005) and Cheng et al. (2011) have considered an additive structure of $M_\tau(\mathbf{X})$:

$$M_\tau(X) = \beta_{0,\tau} + \sum_{j=1}^{d} m_{j,\tau}(X_j), \qquad (4.2)$$

where $\beta_{0,\tau}$ is a constant and $m_{j,\tau}(X_j), j = 1, 2, ..., d$ is an unknown function representing the τth quantile function of Y related only to X_j.

Conditional quantiles are with respect to the distribution of the error term ε_τ, thereby the changes of regression coefficients over different quantiles are not easily interpreted. Quantiles of the unconditional distribution of the outcome variable Y, on the other hand, may be of more general interest and easily interpreted. Firpo et al. (2009a) introduced unconditional quantile regression to estimate the impacts of changes in the distribution of covariates on *unconditional quantile* $q_\tau(Y)$. They show that the marginal effect on the unconditional quantile of a small change in the distribution of explanatory variables, defined as the *unconditional quantile partial effect* (UQPE), equals the average derivative of the probability response model divided by the probability density at the unconditional quantile point, that is,

$$\text{UQPE}(\tau) = \frac{1}{f_Y(q_\tau)} \int \frac{\mathrm{d}\Pr[Y > q_\tau|\mathbf{X}]}{\mathrm{d}x} f_X(x)\mathrm{d}x,$$

where $f_Y(\cdot)$ is the probability density function of the outcome variable.

Firpo et al. (2009a) suggested three estimation methods for the UQPE: ordinary least

square (OLS), logistic regression (Logit) and a fully nonparametric method introduced by Newey (1994) (NP). While a large body of literature has been dedicated to additive modelling of conditional quantiles, an additive modelling has not yet been applied to unconditional quantiles. This manuscript proposes an estimator for a nonparametric additive unconditional quantile regression model. An additive form of unconditional quantile regression model can be written as

$$\Pr\left[Y > q_\tau | \mathbf{X}\right] = \beta_{0,\tau} + \sum_{j=1}^{d} m_{j,\tau}(X_j). \tag{4.3}$$

In an additive quantile regression, the average derivative $\pi(q_\tau) = \left(\mathbf{E}(m'_{1,\tau}(x_1)), ..., \mathbf{E}(m'_{d,\tau}(x_d))\right)^T$ can be estimated by a direct plug–in method:

$$\widehat{\pi}(q_\tau) = \left(n^{-1}\sum_{i=1}^{n}\widehat{m}'_{1,\tau}(x_{1i}), ..., n^{-1}\sum_{i=1}^{n}\widehat{m}'_{d,\tau}(x_{di})\right)^T.$$

Horowitz and Mammen (2004) proposed an estimator of the additive components of a nonparametric additive mean regression model with a known link function. The estimator converges at the rate $n^{-2/5}$ when the additive components are twice differentiable. The estimator does not require d-dimensional nonparametric regression and therefore does not suffer curse of dimensionality. In addition, the estimator has an oracle property: the asymptotic distribution of the estimator of each additive component is the same as in the case when all other additive components are known. Horowitz and Lee (2005) extend Horowitz and Mammen's (2004) approach to conditional quantile regression models. This manuscript extends the approach of Horowitz and Mammen (2004) to the context of unconditional quantiles. I follow Horowitz and Mammen's (2004) two-stage estimation procedure to estimate the average derivative $\pi(q_\tau)$. In the first stage, a series approximation to each component in equation (4.3) is obtained. In the second stage, the point-wise derivative estimator of each additive component can be estimated sequentially by using

55

one-dimensional local polynomial regression in which the first-stage estimates of other components in the nuisance direction are retained.

The reminder of the manuscript is as follows. In section 4.2 the proposed estimator is described. Section 4.3 presents the asymptotic properties of the estimator. Section 4.4 reports the results of Monte Carlo experiments. Section 4.5 applies the estimator to an empirical example. Conclusions are presented in Section 4.6.

4.2 Method

For any $\mathbf{X} \in \mathbb{R}^d$, define $m(\mathbf{X}; q_\tau) \equiv \Pr[Y > q_\tau | \mathbf{X} = x] = \beta_0 + \sum_{j=1}^d m_j(X_j)$. Assume that the support of $\mathbf{X}$ is $\mathscr{X} \equiv [-1,1]^d$, $m_1, ..., m_d$ are normalized so that

$$\int_{-1}^1 m_j(v)\mathrm{d}v = 0, \qquad j = 1, ..., d.$$

Let $\{p_k : k = 1, 2, ...\}$ denote a basis for smooth functions on $[-1,1]$. Conditions that the basis functions must satisfy are given in Section 3. For any positive integer k, define

$$P_k(\mathbf{X}) = [1, p_1(X_1), ..., p_k(X_1), p_1(X_2), ..., p_k(X_2), ..., p_1(X_d), ..., p_k(X_d)]^T,$$

then for $\theta_k \in \mathbb{R}^{kd+1}$, $P_k(\mathbf{X})^T \theta_k$ is a series approximation to $m(\mathbf{X}; q_\tau)$. Section 3 gives conditions that k must satisfy. For a random sample $\{Y_i, \mathbf{X}_i : i = 1, ..., n\}$, let $\widehat{\theta}_k$ be a solution to

$$\min_\theta S_{nk}(\theta) \equiv n^{-1} \sum_{i=1}^n \left[\mathbf{I}\{Y_i > q_\tau\} - P_k(\mathbf{X}_i)^T \theta\right]. \tag{4.4}$$

The series estimator of $m(\mathbf{X}; q_\tau)$ is

$$\widetilde{m}(\mathbf{X}; q_\tau) = P_k(\mathbf{X})^T \widehat{\theta}_k.$$

56

The first–stage estimator of $m_j(X_j)$ for any $j = 1, ..., d$ and any $X_j \in [1, 1]$ is the product of $[p_1(X_j), ..., p_k(X_j)]$ with the appropriate components of $\widehat{\theta}_k$.

I illustrate the estimation of point-wise derivatives $m_1'(x_1)$. One could easily follow similar steps for $m_j'(X_j), j = 2, ..., d$. To obtain the second–stage estimator of $\mathbf{E}[m_1'(X_1)]$, let $\tilde{m}_{-1}(\widetilde{\mathbf{X}}_i) = \widetilde{m}_2(X_{2,i}) + ... + \widetilde{m}_d(X_{d,i})$, where $\widetilde{\mathbf{X}}_i = (X_{2,i}, ..., X_{d,i})$ and $\widetilde{m}_j(X_j)$ is the series estimator of $m_j(X_j)$. Assume that $m_1(X_1)$ is at least $(p+1)$-time continuously differentiable on $[-1, 1]$, then $\widehat{\mathbf{b}}_n = (\widehat{b}_0, \widehat{b}_1, ..., \widehat{b}_{q-1})^T$ is a pth local polynomial estimator which minimizes

$$S_n(b) \equiv \sum_{i=1}^{n} \left[\mathbf{I}\{Y_i > q_\tau\} - \widetilde{\beta}_0 - b_0 - \sum_{v=1}^{p} b_q(X_{1,i} - X_1)^v - \tilde{m}_{-1}(\widetilde{\mathbf{X}}_i) \right]^2 K_\lambda(X_1 - X_{1,i}), \quad (4.5)$$

where $\widetilde{\beta}_0$ is the first component of $\widehat{\theta}_k$, $K(u)$ is a kernel function, $K_\lambda(u) \equiv K(u/\lambda)/\lambda$ with λ as the bandwidth, and $\widehat{b}_1$ is the point-wise derivative estimator, that is $\widehat{m}_1'(x_1) = \widehat{b}_1$. The regularity conditions for $K(u)$ and λ are given in section 4.3. Therefore, the estimator of UQPE with respect to X_j is

$$\widehat{\delta}_j(\widehat{q}_\tau) = \frac{\sum_{i=1}^{n} \widehat{m}_j'(x_{ji}; \widehat{q}_\tau)}{n \widehat{f}_Y(\widehat{q}_\tau)},$$

where the sample quantile $\widehat{q}_\tau$ can be obtained by a check function approach (Koenker and Bassett, 1978), and $\widehat{f}_Y(\cdot)$ is a kernel density estimator. The estimator differs from Firpo et al. (2009a) in that the additive model provides a more flexible fit than the parametric estimation methods suggested by them, and a two-stage estimation procedure avoids the curse of dimensionality, where the nonparametric methods proposed by Firpo et al. (2009a) require d-dimensional nonparametric regression.

4.3 Asymptotic Properties

This section presents asymptotic properties of the estimator described in previous section. Define $m'_j(x_j;q_\tau)$ as the first derivative of m_j, the average derivative is defined as $\pi_j(q_\tau) = \mathbf{E}[m'_j(x_j;q_\tau)]$. Let $u_i = \mathbf{I}\{Y_i > q_\tau\} - m(\mathbf{X}_i;q_\tau)$. The following assumptions will be used to establish asymptotic properties of the two-stage estimator.

Assumption 1. *There is a random sample $(Y_i, \mathbf{X}_i)$ of size n from a joint distribution $F_{Y;\mathbf{X}}(\cdot)$.*

Assumption 2. *The marginal density of Y, $f_Y(\cdot)$, is positive and twice continuously differentiable in a neighborhood of a grid of selected points $q_\tau \in \mathbb{R}$, and $\int |f_Y(y)| dy < \infty$.*

Assumption 3. *The support of $\mathbf{X}$ is $\mathscr{X} \equiv [-1,1]^d$. The marginal density of X_j, $f_{X_j}(\cdot)$ is positive and twice continuously differentiable.*

Assumption 4. *The kernel function $\mathscr{K}(\cdot)$ for the density of Y is a bounded probability density function on $[-1,1]$ and symmetric about zero. To make the asymptotic squared bias of the kernel density estimator go faster to zero than the variance, the bandwidth of the kernel density estimator satisfies the following condition: $\lim_{n\to\infty} n^{1/2} h^{2/5} = 0$.*

Assumption 5. *For each j, $m_j(\cdot)$ has continuous derivatives of total order $p+1$.*

Assumption 6. *Restrictions on the basis functions $\{p_k(\cdot) : k = 1,2,...\}$ are as follows:*

(a) Each $p_k(\cdot)$ is continuous.

(b) $\int_{-1}^{1} p_k(v) dv = 0$

(c) $\int_{-1}^{1} p_j(v) p_k(v) dv = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{otherwise}. \end{cases}$

(d) $\lim_{k\to\infty} \sup_{x \in \mathscr{X}} \|P_K(\mathbf{X})\| = O(k^{1/2})$

(e) There are vectors $\theta_{k0} \in R^{d(k)}$ such that

$$\lim_{n\to\infty} \sup_{x\in\mathcal{X}} |m(\mathbf{X};q_\tau) - P_k(\mathbf{X})^T \theta_{k0}| = O(k^{-(p+1)})$$

The following assumption is established to ensure $\sqrt{n}$-consistency of the first-stage series estimator.

Assumption 7. *$\sqrt{n}k^{-(p+1)} \to 0$ and $k^2/n \to 0$.*

Next, let $\mu_v = \lambda_1^v \int u^v K(u)\mathrm{d}u$. M is the $((p+1) \times (p+1))$ matrix whose (i,j) component is μ_{i+j-2} and N_p is a $((p+1) \times (p+1))$ matrix with (i,j) component μ_{i+j-1}. Define $V(x_1) = f'_{X_1}(x_1)N_p$ and $H(x_1) = (f_{X_1}(x_1)M)^{-1}V(x_1)$. To establish the asymptotic properties of the second-stage estimator, the following additional assumption is needed.

Assumption 8. *As $n \to \infty$, the bandwidth for the local polynomial estimation satisfies, $n\lambda^{2p+2} \to 0$ and $n\lambda^3/\ln(n) \to \infty$.*

The following theorem gives the asymptotic behavior of the first-stage series estimator.

Theorem 1. *Let assumptions 1–8 hold. Then:*

(a)

$$\sqrt{n}(\widehat{\pi}_1(q_\tau) - \pi_1(q_\tau)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi_{\pi,i}(q_\tau) + o_p(1),$$

(b)

$$\sqrt{n}(\widehat{\pi}_1(q_\tau) - \pi_1(q_\tau)) \xrightarrow{d} N(0,V_\psi)$$

where $\psi_{\pi,i}(q_\tau) = m'_1(X_{1,i};q_\tau) - \pi_1(q_\tau) - u_i H(X_{1,i})_{2,1}$ and $V_\psi = \mathbf{E}\left[\psi_\pi(q_\tau)\psi_\pi(q_\tau)^T\right]$.

Define $\psi_{Q,i}(q) = \frac{\mathbf{I}\{Y_i>q\}-(1-\tau)}{f_Y(q)}$, from Theorem 2 of Firpo et al. (2009b), the asymptotic linear representation of the sample quantile is given by the following lemma.

Lemma 4.3.1. *Let assumptions 1 and 2 hold,*

$$\widehat{q}_\tau - q_\tau = \frac{1}{n}\sum_{i=1}^n \psi_{Q,i}(q_\tau) + O_p(n^{-1}).$$

Also, decompose the difference $\widehat{f_Y}(q) - f_Y(q)$ into two parts:

$$\widehat{f_Y}(q) - f_Y(q) = \frac{1}{n}\sum_{i=1}^n \psi_{f,i}(q,\mathcal{K},h) + \mathcal{B}_f(q,\mathcal{K},h),$$

where $\psi_{f,i}(q,\mathcal{K},h) = \mathcal{K}_h(Y_i - q) - \mathbf{E}\left[\widehat{f_Y}(q)\right]$ and $\mathcal{B}_f(q,\mathcal{K},h)$ is the bias of $\widehat{f_Y}(q)$,

$$\mathcal{B}_f(q,\mathcal{K},h) = \mathbf{E}\left[\widehat{f_Y}(q)\right] - f_Y(q).$$

The following lemma is from Theorem 1 of Firpo et al. (2009b).

Lemma 4.3.2. *Let assumptions 1–4 hold,*

$$\widehat{f_Y}(q) - f_Y(q) = O_p((nh)^{-1/2}) + O(h^2) + R_f,$$

where $R_f = O_p((nh)^{-1}) + O(h^4) + O_p(n^{-1/2}h^{3/2})$.

The following theorem establishes the asymptotic results of the second-stage estimator.

Theorem 2. *Let assumption 1–8 hold. Then*

(a)

$$\sqrt{nh}\left(\widehat{\delta}_1(\widehat{q}_\tau) - \delta_1(q_\tau)\right) = \frac{1}{\sqrt{nh}}\sum_{i=1}^n \psi_{2S,i}(q_\tau,h,\mathcal{K}) + r(q_\tau,h,\mathcal{K}),$$

where $\psi_{2S,i}(q_\tau,h,\mathcal{K}) = h \cdot \left(\frac{\psi_{\pi,i}(q_\tau)}{f_Y(q_\tau)} + \delta_1'(q_\tau) \cdot \psi_{Q,i}(q_\tau) - \frac{\delta_1(q_\tau)}{f_Y(q_\tau)} \cdot \psi_{f,i}(q_\tau,h,\mathcal{K})\right)$ and $r(q_\tau,h,\mathcal{K}) = -\sqrt{nh} \cdot \frac{\delta_1(q_\tau)}{f_Y(q_\tau)}\mathcal{B}_f(Q,\mathcal{K},h) + O(n^{1/2}h^{9/2}) + O_p(h^2) + O_p(n^{-1/2}h^{-1}) + O_p(n^{-1}h^{-5/2}).

(b)

$$\sqrt{nh}\left(\widehat{\delta}(\widehat{q}_\tau) - \mathrm{UQPE}(\tau)\right) \xrightarrow{d} N(0, V_{2S}(q_\tau, \mathscr{K})), \text{ where}$$

$$V_{2S}(q_\tau, \mathscr{K}) = \lim_{h \to 0} h^{-1} \cdot \mathbf{E}\left[\psi_{2S,i}(q_\tau, h, \mathscr{K}) \cdot \psi_{2S,i}(q_\tau, h, \mathscr{K})^T\right]$$

(c) If $j \neq 1$, then $(nh)^{-1/2}\left[\widehat{\delta}_1(\widehat{q}_\tau) - \delta_1(q_\tau)\right]$ and $(nh)^{-1/2}\left[\widehat{\delta}_j(\widehat{q}_\tau) - \delta_j(q_\tau)\right]$ are asymptotically independently normally distributed.

4.4 Monte Carlo Experiments

In this section the finite–sample performance of the two-stage estimator is compared with marginal integration, spline estimator and several other estimators. Experiments are carried out with the following data generating process:

$$Y = m_1(3X_1) + m_2(X_2) + \varepsilon, \tag{4.6}$$

where $m_1(v) = \Phi(v)$ is the standard normal distribution function and $m_2(v) = \sin(\pi v)$. The components of $\mathbf{X}$ are bivariate normally distributed with mean 0, unit variance and correlation ρ. The experiments are set up such that a linear model is inappropriate. Sample sizes $n = 100$ and 500 are considered and experiments are carried out with $\rho = 0$ (covariates are uncorrelated), $\rho = .2$ (low correlation between covariates) and $\rho = .8$ (high correlation between covariates) respectively.

B-splines are used for the first stage estimator. To compare the numerical performance of the two-stage estimator with that of the marginal integration estimator of Fan et al. (1998), the series term k is fixed at 4. As the marginal integration estimator requires long computation time, this Monte Carlo experiment was replicated 100 times. Next, let $I_i = \mathbf{I}\{Y_i > \widehat{q}_\tau\}$. To compare the finite-sample performance of the two-stage estimator with a spline estimator and other estimators, a generalized cross-validation described in Li and

61

Racine (2007, p. 452) is used to choose k, specifically, $k = k_{opt} + 1$, where

$$k_{opt} = \arg\min_{k} \frac{n^{-1} \sum_{i=1}^{n} (I_i - \widetilde{m}(\mathbf{X}_i; \widehat{q}_\tau))^2}{(1 - k/n)^2}.$$

For the second stage regression, local linear regressions (setting $p = 1$ in equation (4.5)) for the 15th, 50th and 85th unconditional quantiles are carried out using the quartic (biweight) kernel ($K(v) = \frac{15}{16}(1 - v^2)^2 \mathbf{I}\{|v| \leq 1\}$). The bandwidths for the regression are chosen by a simple rule of thumb described by Fan and Gijbels (1996, p. 110–12).

Neither Firpo et al. (2009a) or anyone else has provide methods to evaluate the fit of unconditional quantile regression. Considering the actual values of the dependent variable in unconditional quantile regression are dichotomous ($I_i = 1$ or 0), here I use a predictive performance (PP) test introduced by Donkers and Melenberg (2002) to compare the in-sample fit of the two-stage estimators, the three estimation methods suggested by Firpo et al. (2009a), and a number of other alternatives.

Let the predictor $\widehat{I}_i$ take value 1 if the estimated probability $\widehat{m}(\mathbf{X}_i; \widehat{q}_\tau) > .5$ and 0 otherwise. Define the hit rate as the fraction of the observations that are correctly predicted by the model under consideration, the predication performance test is based on the difference in hit rates between the model and a naive prediction method, which always predicts 1 for every outcome. Specifically, the test statistic is calculated as follows:

$$PP = \frac{1}{n} \left(\sum_{I_i = 1} \widehat{I}_i + \sum_{I_i = 0} \left(1 - \widehat{I}_i\right) \right) - \frac{1}{n} \sum_{i=1}^{n} I_i.$$

The PP test statistics are reported in Table 4.1. First note from the last two columns of Table 4.1 that the two-stage estimator has higher PP values than the marginal integration estimator for each of the 18 specifications, the comparison is least distinguishable for $\tau = .85$, where the naive approach yields least accurate prediction. The finite-sample performance of the two-stage estimator is shown to be insensitive to correlation between

covariates. De Gooijer and Zerom (2003) suggested the marginal integration estimator performs poorly in the conditional quantile regression when covariates are highly correlated according to the average absolute deviation error (AADE) criterion. It is not the case in the unconditional quantile regression regarding the predictive performance.

Table 4.1 also reports the average PP values of 400 Monte Carlo replications for the B-spline, the two-stage estimators and several other alternative models. Notice that two-stage estimator performs comparably with the spline estimator and marginally better than all other estimators. The relative better performance of the B-spline and the two-stage estimators are most apparent for unconditional median regression ($\tau = .5$), when the dichotomous variable of interest has an equal distribution. A drawback of the B-spline estimator, however, is that it is not suitable for inference.

In summary, the results of the Monte Carlo experiments suggest that the two-stage estimator has much better finite-sample performance than the marginal integration estimator, the three estimators introduced by Firpo et al. (2009a), and single index model estimator. The distinction is especially apparent when the transformed dichotomous variable has an equal distribution and therefore it might be rather difficult to predict the value correctly.

Table 4.1: Predictive Performance Test Results for Monte Carlo Experiments

							PP		PP
ρ	n	OLS	Logit	NP	SIM	BS	2S	MI	2S
					$\tau=$**.15**				
0	100	0.01616	0.01073	-3.641	0.3466	1.397	0.8853	-7.000	1.002
0	500	-0.0025	-0.1678	-10.03	0.1497	1.017	0.5952	-15.65	0.6704
0.2	100	0.0135	0.0340	-3.531	0.3219	1.401	0.934	-7.000	0.9425
0.2	500	0.0000	-0.1352	-9.828	0.1847	1.088	0.6158	-15.65	0.6851
0.8	100	0.0139	0.0074	-3.213	0.3196	1.227	0.9418	-6.998	0.9084
0.8	500	0.005	-0.1887	-9.094	0.581	1.101	0.8565	-15.65	0.8158
					$\tau=$**.5**				
0	100	1.734	1.735	1.605	2.412	4.242	3.817	0.0601	3.818
0	500	3.829	3.832	1.491	4.081	7.82	7.208	0.0085	7.095
0.2	100	1.813	1.811	1.591	2.482	4.211	3.847	0.0589	3.708
0.2	500	3.916	3.915	1.311	4.191	7.84	7.225	0.0134	7.193
0.8	100	2.239	2.241	1.555	3.241	4.275	4.053	0.146	3.872
0.8	500	4.852	4.853	1.117	7.464	8.109	7.645	0.0546	7.901
					$\tau=$**.85**				
0	100	7.005	7.018	7.266	7.116	7.58	7.219	7.022	7.227
0	500	15.65	15.65	15.71	15.67	15.94	15.72	15.66	15.74
0.2	100	7.003	7.022	7.26	7.095	7.573	7.213	7.025	7.24
0.2	500	15.65	15.65	15.71	15.68	15.95	15.73	15.66	15.75
0.8	100	7.002	7.023	7.214	7.107	7.6	7.256	7.022	7.286
0.8	500	15.65	15.66	15.71	15.74	16.11	15.78	15.66	15.76

The values given here are the predictive performance (PP) test statistics for ordinary least square (OLS), logit (Logit), nonparametric (NP), single index (SIM), marginal integration (MI), B-splines (BS) and two-stage (2S) estimators.

4.5   A Numeric Example

This section presents an empirical application to illustrate how the two-stage estimator works in practice. This application considers the Boston house price data. The data consists of 506 observations on 14 variables (see Harrison and Rubinfeld (1978) for a full description of all 14 variables). Many articles have used this data set (see Yu and Lu, 2004; Şentürk and Müller, 2004; Cai and Xu, 2008, among others.). The data set can be obtained from the library "mlbench" in R. I use a corrected version of the data set (data(BostonHousing2)) and consider the effects of four of the covariates. The response variable Y is corrected median value of owner-occupied homes in USD 1000's. The covariates include the logarithm of per capita crime rate, average number of rooms per dwelling, weighted distances of five Boston employment centers, and percentage of lower status of the population (denoted by Econstatus). I follow Cai and Xu (2008) and use the logarithm of crime rate instead of crime rate itself as the correlation between the former and Y are stronger than that between the latter and Y.

The two-stage estimator is used to estimate the additive unconditional function for τ=.15, .5, and .85. Prior to computation, covariates are standardized to have 0 mean and unit variance. In the first stage B-splines is used with $k = \widehat{k} + 1$, where $\widehat{k}$ is determined by leave-one-out cross-validation. As in the simulation illustration, local linear regression is used in the second stage. A standard normal density function with the bandwidths chosen by rule of thumb described by Fan and Gijbels (1996, p. 110–12) is used for the kernel regression.

The estimation results for unconditional quantile regression are summarized in Figures 4.1–4.3, with each panel showing the estimated additive component of interest. It can be seen that the effects of all covariates are highly nonlinear, suggesting a simple linear parametric regression model may be unsuitable.

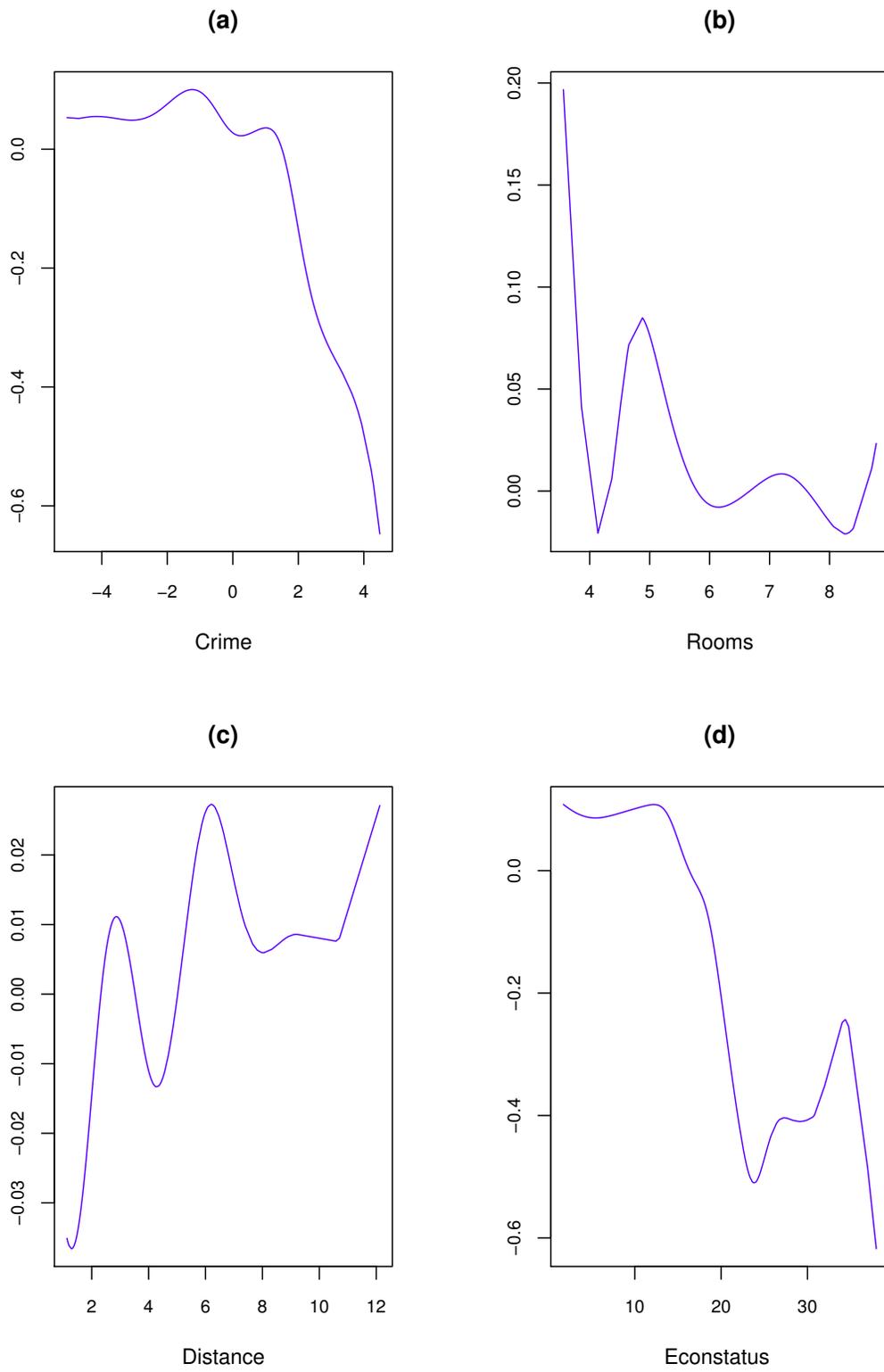Figures 4.4–4.6 further compares the fit of the two-stage model with that of OLS and

Figure 4.1: The estimated additive components of an additive unconditional quantile regression model with $\tau = .15$.
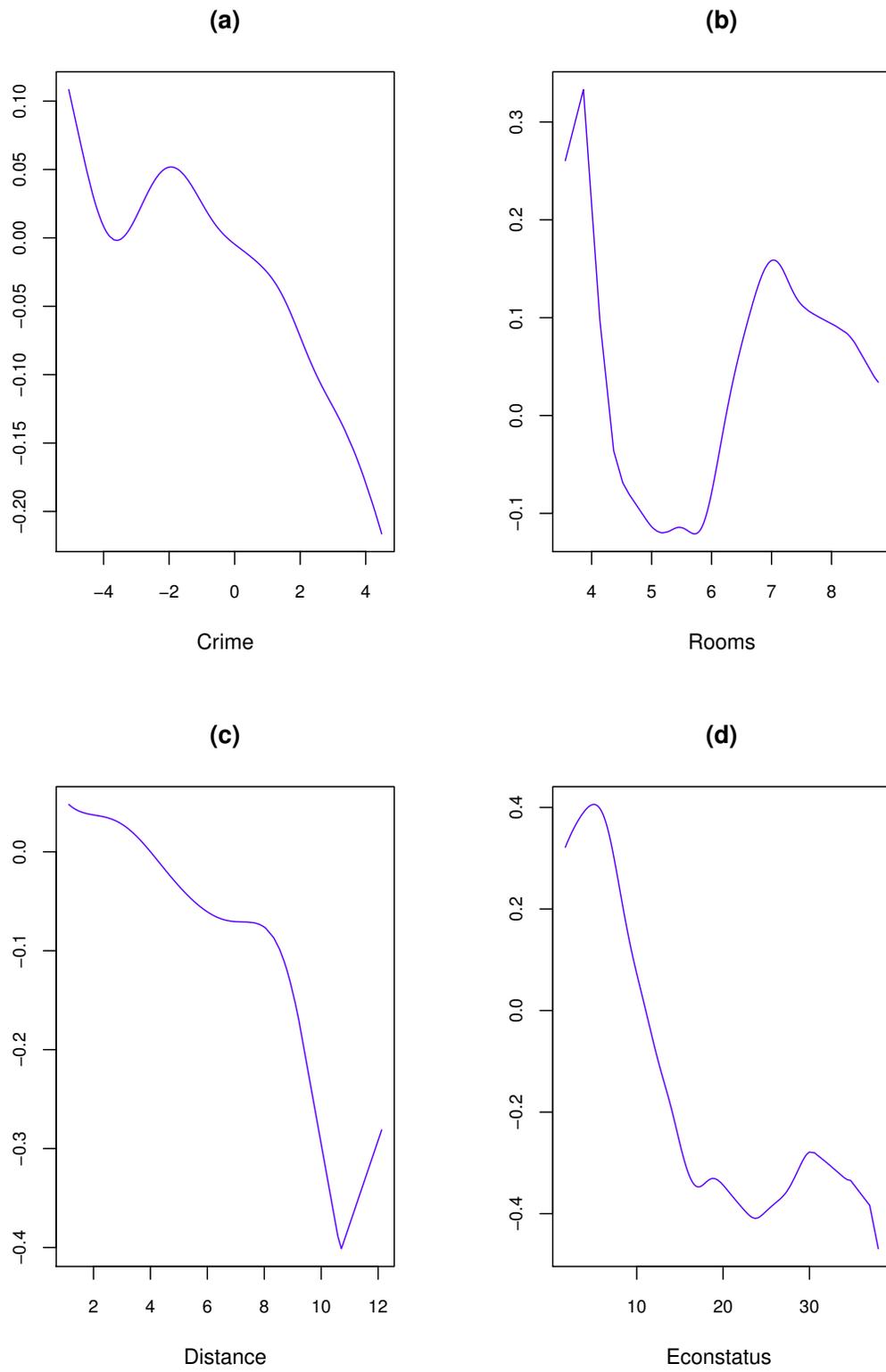
Figure 4.2: The estimated additive components of an additive unconditional quantile regression model with $\tau = .5$.
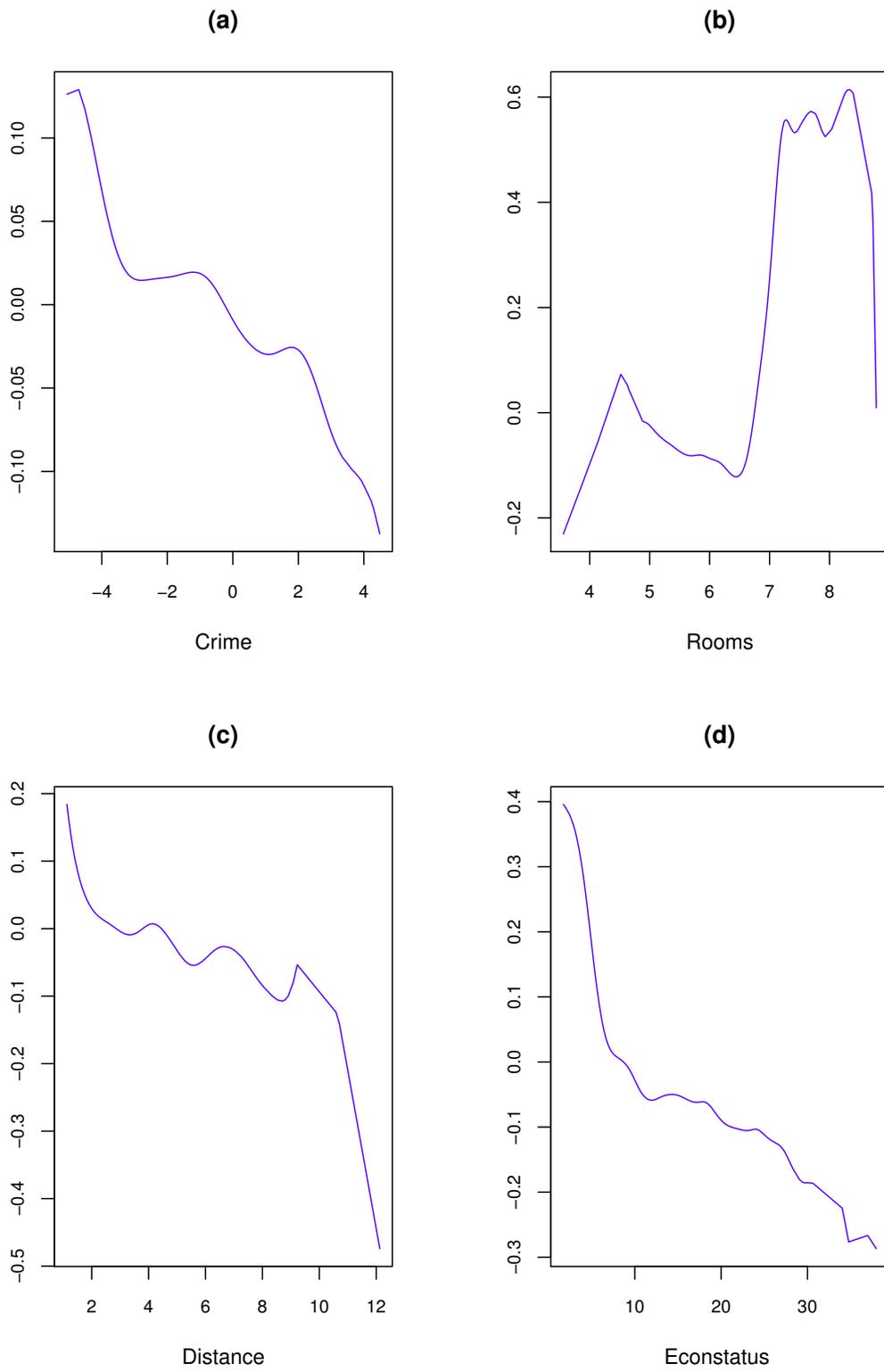
Figure 4.3: The estimated additive components of an additive unconditional quantile regression model with $\tau = .85$.

Logit models using *separation plots* introduced by Greenhill et al. (2011). Such graphs plot the estimated values of probabilities $\widehat{m}(\mathbf{X}; \widehat{q}_\tau)$ in an ascending order, and use dark and light vertical bars to represent the corresponding actual instances of the events $I_i = 1$ and nonevents $I_i = 0$ respectively. In a model with perfect predictive power, actual nonevents $I_i = 0$ would always be associated with low fitted probabilities, whereas events $I_i = 1$ would always be associated with high fitted probabilities. Therefore a perfect model would produce a complete separation of the events and nonevents, with all the events clustered on the right-hand end. Besides producing a visual display that is informative and intuitive, *separation plots* is intensive to the probability thresholds used to distinguish between $\widehat{I}_i = 1$ and $\widehat{I}_i = 0$. As those figures show, the two-stage estimator does a better job of modelling each one of the three unconditional quantiles given that less events are distributed on the left-hand side of the graphs.

Figure 4.4: Separation plots in unconditional quantile regression with $\tau = .15$.
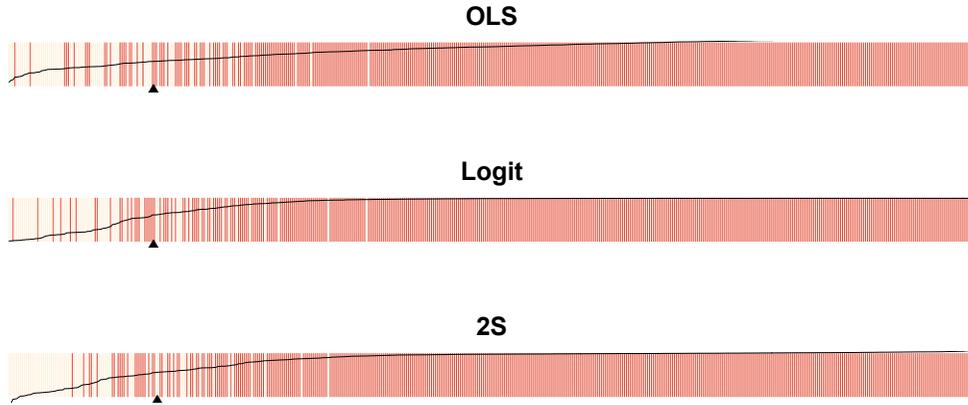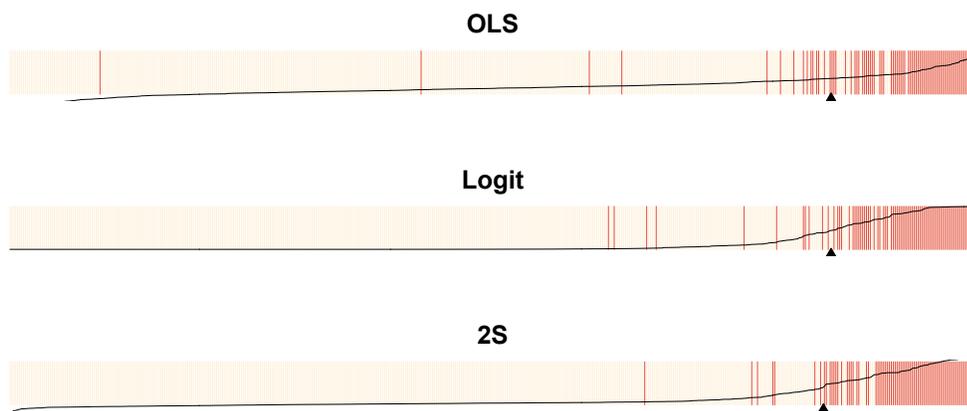
OLS

Logit

2S

Note: The curve represents the values of $\widehat{m}(\mathbf{X}_i; \widehat{q}_\tau)$ in a ascending order with y-axis ranging from 0 to 1. The dark and light panels represent the corresponding $I_i = 1$ and $I_i = 0$ respectively. The quantity above the small triangle indicates the expected number of events $\sum_{i=1}^{n} \widehat{m}(\mathbf{X}_i; \widehat{q}_\tau)$.

Figure 4.5: Separation plots in unconditional quantile regression with $\tau = .5$.
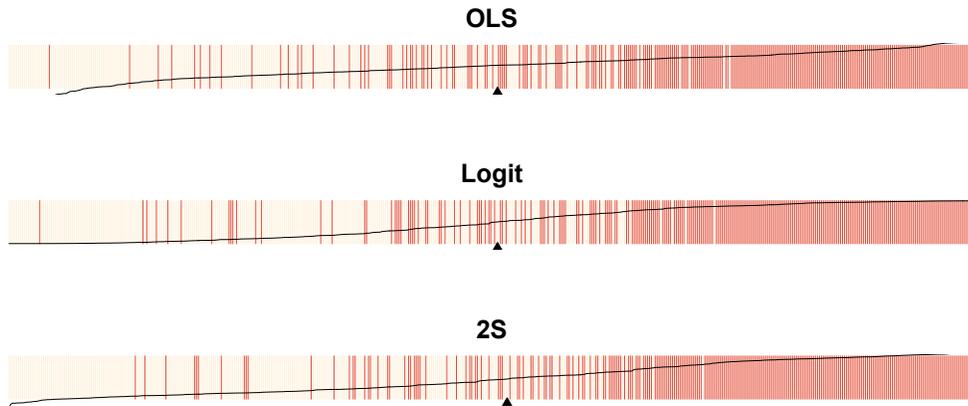
OLS

Logit

2S

Note: The curve represents the values of $\widehat{m}(\mathbf{X}_i; \widehat{q}_\tau)$ in a ascending order with y-axis ranging from 0 to 1. The dark and light panels represent the corresponding $I_i = 1$ and $I_i = 0$ respectively. The quantity above the small triangle indicates the expected number of events $\sum_{i=1}^{n} \widehat{m}(\mathbf{X}_i; \widehat{q}_\tau)$.

Figure 4.6: Separation plots in unconditional quantile regression with $\tau = .85$.

OLS

Logit

2S

Note: The curve represents the values of $\widehat{m}(\mathbf{X}_i; \widehat{q}_\tau)$ in a ascending order with *y*-axis ranging from 0 to 1. The dark and light panels represent the corresponding $I_i = 1$ and $I_i = 0$ respectively. The quantity above the small triangle indicates the expected number of events $\sum_{i=1}^{n} \widehat{m}(\mathbf{X}_i; \widehat{q}_\tau)$.

Next, the *unconditional quantile partial effect* $\delta(q_\tau)$ is estimated and the results are summarized in Table 4.2. The standard errors are obtained from the asymptotic normal approximation of Theorem 2 with kernel density estimators. As indicated in Table 4.2 , the unconditional effects of covariates are different among different quantiles. For instance, the unconditional effect of the Econstatus first decreases from -2.801 at the 15th quantile to -4.166 at the median, then decreases to a large negative effect of -12.45 at the 85th quantile, suggesting Econstatus has big effects in reducing the house prices in the top quantile of the distribution. I use expected Percentage of Correct Predictions (ePCP) proposed by Herron (1999) to assess the fit of the model, which essentially measures the average of the probabilities the hypothetical model assigns to the correct outcome category. This statistic helps avoid the problem of treating an observation with the predicted probability $\widehat{m}(\mathbf{X}_i;\widehat{q}_\tau) = 0.51$ the same as an observation with $\widehat{m}(\mathbf{X}_i;\widehat{q}_\tau) = 0.99$. The following formula is used to calculate ePCP:

$$ePCP = \frac{1}{n}\left(\sum_{I_i=1}\widehat{m}(\mathbf{X}_i;\widehat{q}_\tau) + \sum_{I_i=0}\left(1 - \widehat{m}(\mathbf{X}_i;\widehat{q}_\tau)\right)\right)$$

The last column of Table 4.2 reports the ePCP values of each regression. It can be seen that the two-stage estimate does an excellent job in the unconditional quantile regressions, assigning 90%, 78% and 92% of the probability to the correct outcome category for 15%, 50% and 85% unconditional quantile regressions respectively.

Table 4.2: Unconditional quantile partial effects

τ	Crime	Rooms	Distance	Ecostatus	ePCP
0.15	-3.464	-0.1666	1.01	-2.801	0.8997
	(0.3907)	(0.4435)	(0.5824)	(0.3397)	
0.5	-0.3501	1.95	-0.5596	-4.166	0.7814
	(0.2461)	(0.2477)	(0.2369)	(0.2279)	
0.85	-1.701	4.534	-4.913	-12.45	0.9235
	(1.114)	(1.368)	(1.125)	(0.8505)	

The standard errors are given in parentheses.

4.6　Conclusion

Unconditional quantile regression provide a workable method for estimating the effects of explanatory variables on the different unconditional quantiles of an outcome variable, defined as *unconditional quantile partial effects* (UQPE). This manuscript has developed an estimator of UQPE through a nonparametric additive regression model.

It is shown that the proposed estimator is asymptotically normally distributed with a rate of convergence in probability of $(nh)^{-1/2}$, where h is the bandwidth in the kernel density estimation of the outcome variable. This result holds regardless of the dimension of the explanatory variables. In addition, the new estimator has an oracle property, that is, the estimator of each component has the same asymptotic distribution as in the case when the other components are known. Finally Monte Carlo experiments and an empirical application show that the two-stage estimator performs better than a number of existing alternative methods.

Chapter 5

Conclusions

5.1 Main Conclusion of Essay 1

In the first essay, both political pressure and contracting costs are incorporated in analyzing trade agreements. Like many previous studies in the political economy literature (Hillman, 1982; Snyder Jr, 1990; Grossman and Helpman, 1994, 1995a,b), this study views governments as agents that maximize their own interests in response to political pressure rather than as benevolent agents that maximize aggregate social welfare. Grossman and Helpman (1994) brought a first coherent theoretical model of endogenous trade policy formation and concludes tariff rates are affected both by a *political support* motive and a *terms-of-trade* motive. However, an implicit assumption of the mainstream political economy model is that less distortionary domestic policies are not available for redistribution, which is clearly not the case in current policy mechanism. Schleich and Orden (1996) included both trade and domestic production policies in their political economy model and concluded that production subsidies substitute for trade policies that would have otherwise resulted from rent seeking efforts. As a result, when countries act noncooperatively, tariff rates are exactly Johnson's optimal tariff rates, which represents *only* the *terms-of-trade* motive. This study goes one step further by identifying the efficient policy choices in a cooperative equilib-

75

rium, or a costless trade agreement.

The model provides political economy rationale for countervailing duty law by showing that a costless trade agreement would lead to equal tariff rate in the importing country and export subsidy rate in the exporting country, since production subsidies can be used to redistribute and meet interest groups' demand for rent. It is also found that cooperative production subsidy rates are the same as those in a noncooperative equilibrium confirming that a trade agreement which constrains production subsidies but not tariffs is not optimal, an important finding proposed by Horn et al. (2010) using externality as the rationale for trade agreements.

Horn et al. (2010) propose a *monopoly power* effect, a *trade volume* effect and an *instrument substitutability* effect as the key features of the circumstances that determine the benefit of constraining domestic production subsidies and that of constraining consumption taxes in an NT-based trade agreement. They use externalities as the rationale for trade agreements and the source of uncertainty. They find that an optimal trade agreement should at least make the tariff level contingent on the consumption externality. Unfortunately, externality is difficult to measure and test.

Like Horn et al. (2010), this model predicts that uncertainty induces a trade-off between contracting costs and including more policy and state variables in a trade agreement. The uncertainty over production and consumption externalities is replaced with changing rent-seeking conditions. This replacement yields additional insights into the mechanisms that drive trade agreements. For example, if contracting is costless, an optimal trade agreement should make constraints contingent on variables representing the changing policy environment during the lifetime of the agreement, this is more convincing and empirically plausible.

This manuscript decomposes the *monopoly power* effect into a *trade volume* effect and the effect of *price sensitivity of import demand*, this decomposition is simple yet yields new

76

interpretation for empirical investigation. In addition, the positive *trade volume* effect on the gains of constraining production policy or consumption policy proposed by Horn et al. (2010) is identified as a special case where the government assign zero to the weight of political contribution or *effective constraints* are already set on trade-promoting policies.

5.2 Main Conclusion of Essay 2

Previous modelling of commodity price volatility assumes a single-state GARCH process and constant conditional skewness and kurtosis, and therefore is not able to detect the state dependent volatility dynamics if multiple states exist. Commodity price volatility may respond differently under different market states, for example, under the expectation of positive and negative price changes. The NM-type GARCH models allow for state-dependent volatility behavior and time-varying conditional skewness and kurtosis. Haas et al. (2004) and Alexander and Lazar (2009), among others, have applied those models in equity markets. In the second essay, I modelled agricultural commodity price volatility using the NM-GARCH models with the assumption of two market states.

Both in-sample and out-of-sample diagnostics are conducted to compare the fit of the NM-GARCH and the NM-GJR-GARCH models with normal GARCH specification. The overall conclusion is that the class of NM-GARCH models adequately captures relevant properties of agricultural commodity price data but the single-state normal GARCH model performs rather poorly especially regarding the specification of skewness and kurtosis. Contrary to the case in the equity market as found in Alexander and Lazar (2009), the addition of dynamic asymmetry in the NM-GJR-GARCH model is sometimes found unnecessary for a few commodities, as it disturbs the time series fit and upper tail prediction.

Empirical results on ten agricultural commodity cash prices show a clear relationship between expected price change and the volatility dynamics across regimes. For each of the ten commodities, an expected negative price change corresponds to a greater volatility

77

persistence, while an expected positive price change arises in conjunction with an increasing responsiveness of volatility. This is just the opposite of the case in the equity market, where Haas et al. (2004) found volatility is more persistent to positive shocks and more responsive to negative shocks.

Finally, when possible state-dependent "inverse leverage effects" are explicitly accounted for, as in the NM-GJR-GARCH model, we found that for most commodities these effect are insignificant except on occasions when component means are negative. A significant inverse leverage effect is detected only for corn in a less frequently occurred regime where price falls are anticipated, which indicates the volatility in this regime tends to increase more following a realized price rise than a realized price drop. Conversely, beef is found to have significant leverage effects during the more frequent regime where prices are expected to fall, indicating a realized price fall would lead to higher volatility than a realized price recovery. By allowing state-dependent inverse leverage effects and volatility dynamics, two-state NM-type GARCH models could facilitate more refined risk management practice than single-state GARCH models.

5.3   Main Conclusion of Essay 3

Unconditional quantile regression provide a workable method for estimating the effects of explanatory variables on the different unconditional quantiles of an outcome variable, defined as *unconditional quantile partial effects* (UQPE). This manuscript has developed an estimator of UQPE through a nonparametric additive regression model.

We have shown that the proposed estimator is asymptotically normally distributed with a rate of convergence in probability of $(nh)^{-1/2}$, where h is the bandwidth in the kernel density estimation of the outcome variable. This result holds regardless of the dimension of the explanatory variables. In addition, the new estimator has an oracle property, that is, the estimator of each component has the same asymptotic distribution as in the case

when the other components are known. Finally Monte Carlo experiments and an empirical application show that the two-stage estimator performs better than a number of existing alternative methods.

Bibliography

Alexander, C., Lazar, E., 2006. Normal mixture garch (1, 1): Applications to exchange rate modelling. Journal of Applied Econometrics 21 (3), 307–336.

Alexander, C., Lazar, E., 2009. Modelling regime-specific stock price volatility. Oxford Bulletin of Economics and Statistics 71 (6), 761–797.

Apergis, N., Rezitis, A., 2003. Agricultural price volatility spillover effects: the case of greece. European Review of Agricultural Economics 30 (3), 389–406.

Bagwell, K., Staiger, R., 1999. An Economic Theory of GATT. American Economic Review 89 (1), 215–248.

Battigalli, P., Maggi, G., 2002. Rigidity, discretion, and the costs of writing contracts. American Economic Review 92 (4), 798–817.

Bauwens, L., Hafner, C. M., Rombouts, J. V., 2007. Multivariate mixed normal conditional heteroskedasticity. Computational Statistics & Data Analysis 51 (7), 3551–3566.

Bera, A. K., Higgins, M. L., 1993. Arch models: properties, estimation and testing. Journal of economic surveys 7 (4), 305–366.

Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. Journal of Business & Economic Statistics 19 (4), 465–474.

Black, F., 1976. Studies of stock price volatility changes. In: Proceedings of the 1976 Meetings of the American Statistical Association,In Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Statistics Section. American Statistical Association, pp. 177–181.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of econometrics 31 (3), 307–327.

Bollerslev, T., 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. The review of economics and statistics, 542–547.

Bollerslev, T., Chou, R. Y., Kroner, K. F., 1992. Arch modeling in finance: a review of the theory and empirical evidence. Journal of econometrics 52 (1), 5–59.

Buguk, C., Hudson, D., Hanson, T., 2003. Price volatility spillover in agricultural markets: An examination of us catfish markets. Journal of Agricultural and Resource Economics, 86–99.

Cai, Z., Xu, X., 2008. Nonparametric quantile estimations for dynamic smooth coefficient models. Journal of the American Statistical Association 103 (484), 1595–1608.

Chang, K.-L., 2012. Volatility regimes, asymmetric basis effects and forecasting performance: An empirical investigation of the wti crude oil futures market. Energy Economics 34 (1), 294–306.

Chaudhuri, P., 1991a. Global nonparametric estimation of conditional quantile functions and their derivatives. Journal of Multivariate Analysis 39 (2), 246–269.

Chaudhuri, P., 1991b. Nonparametric estimates of regression quantiles and their local bahadur representation. The Annals of statistics 19 (2), 760–777.

Cheng, Y., De Gooijer, J., Zerom, D., 2011. Efficient estimation of an additive quantile regression model. Scandinavian Journal of Statistics 38, 46–62.

Christie, A. A., 1982. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. Journal of financial Economics 10 (4), 407–432.

Christoffersen, P. F., 1998. Evaluating interval forecasts. International economic review, 841–862.

Cole, T., 1988. Fitting smoothed centile curves to reference data. Journal of the Royal Statistical Society. Series A (Statistics in Society), 385–418.

Copeland, B., 1990. Strategic interaction among nations: negotiable and non-negotiable trade barriers. Canadian Journal of Economics, 84–108.

Şentürk, D., Müller, H., 2004. Covariate adjusted correlation analysis via varying coefficient models. Scandinavian Journal of Statistics 32 (3), 365–383.

De Gooijer, J., Zerom, D., 2003. On additive conditional quantiles with high-dimensional covariates. Journal of the American Statistical Association 98 (461), 135–146.

Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society. Series B (Methodological), 1–38.

Donkers, A., Melenberg, B., 2002. Testing predictive performance of binary choice models. Tech. rep., Erasmus School of Economics (ESE).

Dumitrescu, E.-I., Hurlin, C., Madkour, J., 2013. Testing interval forecasts: A gmm-based approach. Journal of Forecasting 32 (2), 97–110.

Eicher, T., Osang, T., 2002. Protection for sale: An empirical investigation: Comment. The American Economic Review 92 (5), 1702–1710.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: Journal of the Econometric Society, 987–1007.

Engle, R. F., Ng, V. K., 1993. Measuring and testing the impact of news on volatility. The Journal of Finance 48 (5), 1749–1778.

Engle, R. F., Rangel, J. G., 2008. The spline-garch model for low-frequency volatility and its global macroeconomic causes. Review of Financial Studies 21 (3), 1187–1222.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. London: Chapman & Hall.

Fan, J., Härdle, W., Mammen, E., 1998. Direct estimation of low-dimensional components in additive models. The Annals of Statistics 26 (3), 943–971.

Fan, J., Hu, T., Truong, Y., 1994. Robust non-parametric function estimation. Scandinavian Journal of Statistics, 433–446.

FAO, I., UNCTAD, W., 2011. Price volatility in food and agricultural markets: Policy responses. Food and Agricultural Organization.

Fernández, C., Steel, M. F., 1998. On bayesian modeling of fat tails and skewness. Journal of the American Statistical Association 93 (441), 359–371.

Findlay, R., Wellisz, S., 1982. Endogenous tariffs, the political economy of trade restrictions, and welfare. In: Bhagwati, J. (Ed.), Import competition and response. pp. 223–244.

Firpo, S., Fortin, N., Lemieux, T., 2009a. Unconditional quantile regressions. Econometrica 77 (3), 953–973.

Firpo, S., Fortin, N. m., Lemieux, T., 2009b. Supplement to "unconditional quantile regressions: estimation and testing. Econometrica 77 (3), 953–973.

Gawande, K., Bandyopadhyay, U., 2000. Is protection for sale? evidence on the grossman-helpman theory of endogenous protection. Review of Economics and statistics 82 (1), 139–152.

Geman, H., Shih, Y. F., 2009. Modeling commodity prices under the cev model. The Journal of Alternative Investments 11 (3), 65–84.

Giot, P., Laurent, S., 2003. Value-at-risk for long and short trading positions. Journal of Applied Econometrics 18 (6), 641–663.

Glosten, L. R., Jagannathan, R., Runkle, D. E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. The Journal of Finance 48 (5), 1779–1801.

Goldberg, P. K., Maggi, G., 1999. Protection for sale: An empirical investigation. The American Economic Review 89 (5), 1135–1155.

Greenhill, B., Ward, M. D., Sacks, A., 2011. The separation plot: a new visual method for evaluating the fit of binary models. American Journal of Political Science 55 (4), 991–1002.

Grossman, G., Helpman, E., 1994. Protection for sale. American Economie Review 84 (4), 833–850.

Grossman, G., Helpman, E., 1995a. Trade wars and trade talks. Journal of Political Economy, 675–708.

Grossman, G., Helpman, E., 1995b. The politics of free-trade agreements. American Economic Review 85 (4), 667–690.

Haas, M., Mittnik, S., Paolella, M. S., 2004. Mixed normal conditional heteroskedasticity. Journal of Financial Econometrics 2 (2), 211–250.

Hall, J. A., Brorsen, B. W., Irwin, S. H., 1989. The distribution of futures prices: A test of the stable paretian and mixture of normals hypotheses. Journal of Financial and Quantitative Analysis 24 (01), 105–116.

Harrison, D., Rubinfeld, D., 1978. Hedonic housing prices and the demand for clean air. Journal of Environmental Economics and Management 5 (1), 81–102.

He, X., Ng, P., Portnoy, S., 1998. Bivariate quantile smoothing splines. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60 (3), 537–550.

He, X., Shi, P., 1994. Convergence rate of b-spline estimators of nonparametric conditional quantile functions. Journaltitle of Nonparametric Statistics 3 (3-4), 299–308.

Hennessy, D. A., Wahl, T. I., 1996. The effects of decision making on futures price volatility. American Journal of Agricultural Economics 78 (3), 591–603.

Herron, M. C., 1999. Postestimation uncertainty in limited dependent variable models. Political Analysis 8 (1), 83–98.

Hillman, A., 1982. Declining industries and political-support protectionist motives. American Economic Review 72 (5), 1180–1187.

Honda, T., 2004. Quantile regression in varying coefficient models. Journal of statistical planning and inference 121 (1), 113–125.

Horn, H., 2006. National treatment in the GATT. American economic review 96 (1), 394–404.

Horn, H., Maggi, G., Staiger, R., 2010. Trade agreements as endogenously incomplete contracts. The American Economic Review 100 (1), 394–419.

Horowitz, J., Lee, S., 2005. Nonparametric estimation of an additive quantile regression model. Journal of the American Statistical Association 100 (472), 1238–1249.

Horowitz, J., Mammen, E., 2004. Nonparametric estimation of an additive model with a link function. The Annals of Statistics 32 (6), 2412–2443.

Jarque, C. M., Bera, A. K., 1987. A test for normality of observations and regression residuals. International Statistical Review/Revue Internationale de Statistique, 163–172.

Jin, H. J., Frechette, D. L., 2004. Fractional integration in agricultural futures price volatilities. American Journal of Agricultural Economics 86 (2), 432–443.

Karali, B., Power, G. J., 2013. Short-and long-run determinants of commodity price volatility. American Journal of Agricultural Economics 95 (3), 724–738.

Kim, M., 2007. Quantile regression with varying coefficients. The Annals of Statistics 35 (1), 92–108.

Kocherginsky, M., He, X., Mu, Y., 2005. Practical confidence intervals for regression quantiles. Journal of Computational and Graphical Statistics 14 (1), 41–55.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46 (1), 33–50.

Koenker, R., Ng, P., Portnoy, S., 1994. Quantile smoothing splines. Biometrika 81 (4), 673.

Koenker, R., Zhao, Q., 1996. Conditional quantile estimation and inference for arch models. Econometric Theory 12, 793–813.

Lee, S., 2003. Efficient semiparametric estimation of a partially linear quantile regression model. Econometric theory 19 (1), 1–31.

Li, Q., Lu, X., Ullah, A., 2003. Multivariate local polynomial regression for estimating average derivatives. Journal of Nonparametric Statistics 15 (4-5), 607–624.

Li, Q., Racine, J. S., 2007. Nonparametric econometrics: Theory and practice. Princeton University Press.

Maggi, G., Rodriguez-Clare, A., 1998. The value of trade agreements in the presence of political pressures. Journal of Political Economy, 574–601.

Maggi, G., Rodriguez-Clare, A., 2007. A political-economy theory of trade agreements. American economic review 97 (4), 1374–1406.

McCalman, P., 2004. Protection for sale and trade liberalization: an empirical investigation. Review of International Economics 12 (1), 81–94.

Mitra, D., Thomakos, D. D., Ulubasoglu, M. A., 2002. protection for sale in a developing country: democracy vs. dictatorship. Review of Economics and Statistics 84 (3), 497–508.

Newey, W. K., 1994. The asymptotic variance of semiparametric estimators. Econometrica, 1349–1382.

Newey, W. K., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79 (1), 147–168.

Nordhaus, W., 1975. The political business cycle. Review of Economic Studies 42 (2), 169–190.

Poon, S.-H., Granger, C. W., 2003. Forecasting volatility in financial markets: A review. Journal of Economic Literature 41 (2), 478–539.

Rezitis, A. N., Stavropoulos, K. S., 2010. Modeling beef supply response and price volatility under cap reforms: the case of greece. Food Policy 35 (2), 163–174.

Schleich, J., Orden, D., 1996. Efficient choice among domestic and trade policies in the grossman-helpman interest-group model. Working Paper 7458, University of Minnesota, Economic Development Center.

Schleich, J., Orden, D., 2000. Environmental quality and industry protection with noncooperative versus cooperative domestic and trade policies. Review of International Economics 8 (4), 681–697.

Serra, T., 2013. Time-series econometric analyses of biofuel-related price volatility. Agricultural Economics.

Serra, T., Gil, J. M., 2013. Price volatility in food markets: can stock building mitigate price fluctuations? European Review of Agricultural Economics 40 (3), 507–528.

Serra, T., Zilberman, D., Gil, J., 2011. Price volatility in ethanol markets. European Review of Agricultural Economics 38 (2), 259–280.

Shively, G. E., 1996. Food price variability and economic reform: An arch approach for ghana. American Journal of Agricultural Economics 78 (1), 126–136.

Snyder Jr, J., 1990. Campaign contributions as investments: The US house of representatives, 1980-1986. Journal of Political Economy, 1195–1227.

Yang, S.-R., Brorsen, B. W., 1992. Nonlinear dynamics of daily cash prices. American Journal of Agricultural Economics 74 (3), 706–715.

Yu, K., Jones, M., 1998. Local linear quantile regression. Journal of the American Statistical Association 93 (441), 228–237.

Yu, K., Lu, Z., 2004. Local linear additive quantile regression. Scandinavian Journal of Statistics 31 (3), 333–346.

Zhou, K., Portnoy, S., 1996. Direct use of regression quantiles to construct confidence sets in linear models. The Annals of Statistics 24 (1), 287–306.

Appendix

A.1 Appendix of Essay 1

A.1.1 Derivation of Home's noncooperative policies defined by equations(2.4) and (2.5)

Let $P^0 = (\tau^0, s^0)$ and $P^{*0} = (\tau^{*0}, s^{*0})$ and assume that contribution schedules are differentiable around the equilibrium point. The first order conditions (FOC) of equations (2.2) and (2.3) give

$$\sum_{j \in L} \nabla_P C_j^0(P^0, P^*) + a \nabla_P W(P^0, P^*) = 0, \tag{A-1}$$

and

$$\nabla_P W_i(P^0, P^*) - \nabla_P C_i^0(P^0, P^*)$$
$$+ \sum_{j \in L} \nabla_P C_j^0(P^0, P^*) + a \nabla_P W(P^0, P^*) = 0 \text{ for all } i \text{ in } L. \tag{A-2}$$

The system above implies

$$\nabla_P C_i^0(P^0, P^*) = \nabla_P W_i(P^0, P^*) \text{ for all } i \text{ in } L. \tag{A-3}$$

Summing equation (A-3) over all i and substituting into equation (A-1) give

$$\sum_{i \in L} \nabla_P W_i(P^0, P^*) + a \nabla_P W(P^0, P^*) = 0. \tag{A-4}$$

This equation gives the equilibrium Home policy choices conditional on Foreign policy vector P^*. Similarly, the following equilibrium Foreign policy vectors can be obtained

$$\sum_{i \in L^*} \nabla_{P^*} W_i^*(P^{*0}, P) + a^* \nabla_{P^*} W^*(P^{*0}, P) = 0. \tag{A-5}$$

The noncooperative equilibrium policy vectors are characterized tby substituting P^{*0} for P^* in equation (A-4) and P^0 for P in equation (A-5) and treating these as a system of simultaneous equations.

Substituting $P^0 = (\tau^0, s^0)$ into equation (A-4) and taking derivatives give

$$(I_{iL} - \alpha_L)(\omega_i + \tau_i^0 \omega_{i1})X_i + (a + \alpha_L)[(\tau_i^0 - 1)\omega_i M_i'(\omega_i + \tau_i^0 \omega_{i1})$$
$$- \omega_{i1}M_i - s_i^0 X_i'(\omega_i + \tau_i^0 \omega_{i1})] = 0, \tag{A-6}$$

and

$$(I_{iL} - \alpha_L)(\tau_i^0 \omega_{i2} + 1)X_i + (a + \alpha_L)\{(\tau_i^0 - 1)\omega_i[D_i'\tau_i^0 \omega_{i2} - X_i'(\tau_i^0 \omega_{i2} + 1)]$$
$$- \omega_{i2}M_i - s_i^0 X_i'(\tau_i^0 \omega_{i2} + 1)\} = 0. \tag{A-7}$$

From equation (2.1), the partial derivatives of the world price functions are derived, $\omega_{i1} = \partial \omega_i / \partial \tau_i = -M_i' \omega_i / (M_i' \tau_i + M_i^{*'} \tau_i^*)$, $\omega_{i2} = \partial \omega_i / \partial s_i = X_i' / (M_i' \tau_i + M_i^{*'} \tau_i^*)$. Substituting them into equations (A-6) and (A-7) yields equation (2.4) and (2.5).

A.1.2 Derivation of globally efficient policies defined by equations (2.8) through (2.10)

Following the same derivations as the noncooperative case, the following two conditions are satisfied:

$$a^* \sum_{i \in L} \nabla_P W_i(P^0, P^{*0}) + a \sum_{i \in L^*} \nabla_P W_i^*(P^{*0}, P^0) + a^* a [\nabla_P W(P^0, P^{*0})$$

$$+ \nabla_P W^*(P^{*0}, P^0)] = 0, \tag{A-8}$$

and

$$a^* \sum_{i \in L} \nabla_{P^*} W_i(P^0, P^{*0}) + a \sum_{i \in L^*} \nabla_{P^*} W_i^*(P^{*0}, P^0) + a^* a [\nabla_{P^*} W(P^0, P^{*0})$$

$$+ \nabla_{P^*} W^*(P^{*0}, P^0)] = 0. \tag{A-9}$$

It is convenient to begin with the case in which factor owners represented by lobby groups comprise a negligible fraction of the voters in each country, i.e., $a_L = a_L^* = 0$. Substituting $P^0 = (\tau^0, s^0)$ into equation (A-8) and solving yield the globally efficient policies defined by

$$\tau_i^0 - \tau_i^{*0} = \left(-\frac{I_{iL}}{a} \frac{X_i}{\omega_i M_i'} + s_i^0 \frac{X_i'}{\omega_i M_i'} \right) - \left(-\frac{I_{iL}^*}{a^*} \frac{X_i^*}{\omega_i M_i^{*\prime}} + s_i^{*0} \frac{X_i^{*\prime}}{\omega_i M_i^{*\prime}} \right), \tag{A-10}$$

and

$$a^* a X_i' [s_i^0 (D_i' \tau_i^0 + M_i^{*\prime} \tau_i^{*0}) + s_i^{*0} X_i^{*\prime} \tau_i^{*0}] = a^* (D_i' \tau_i^0 + M_i^{*\prime} \tau_i^{*0}) I_{iL} X_i + a X_i' \tau_i^{*0} I_{iL}^* X_i^*$$

$$- a^* a X_i' (\tau_i^0 - \tau_i^{*0}) \omega_i M_i^{*\prime} \tau_i^{*0}. \tag{A-11}$$

89

Solving the system of equations gives

$$s_i^e = \frac{I_{iL}}{a} \frac{X_i}{X_i'}. \tag{A-12}$$

Similarly, from equation (A-9),

$$s_i^{*e} = \frac{I_{iL}^*}{a^*} \frac{X_i^*}{X_i^{*\prime}}. \tag{A-13}$$

Substituting equations (A-12) and (A-13) into equation (A-10) yields equation (2.8).

The analysis can be extended to the more general case. When $a_L > 0$ and $a_L^* > 0$, equation (2.8) still holds. With respect to production policies, it is straightforward to replace a^*, a, I_{iL} and I_{iL}^* with $a^* + \alpha_L^*$, $a + \alpha_L$, $I_{iL} - \alpha_L$ and $I_{iL}^* - \alpha_L^*$ in equations (A-10) and (A-11) and follow the process as the derivation of equation (A-12) to obtain equation (2.9) and (2.10).

A.2 Appendix of Essay 2

A.2.1 Proof of Theorem 1

Let $\widehat{\delta}(q) = \frac{\widehat{\pi}(q)}{\widehat{f}_Y(q)}$ and $\delta(q) = \frac{\pi(q)}{f_Y(q)}$, The difference can be decomposed into two parts: $\widehat{\delta}(\widehat{q}_\tau) - \delta(q_\tau) = \widehat{\delta}(\widehat{q}_\tau) - \widehat{\delta}(q_\tau) + \widehat{\delta}(q_\tau) - \delta(q_\tau)$. Define $\beta(x_1) = [m(x_1), m'(x_1), \cdots, m^{(p)}(x_1)/p!]^T$. Also let Z be a $(n \times (p+1))$ matrix whose ith row vector is defined as $Z_i(x_1) = [1, (X_{1,i} - x_1), \cdots, (X_{1,i} - x_1)^p]$, and define B as a vector with ith component $B_i(x_1) = m_1(X_{1,i}) - m_1(x_1) - \sum_{v=1}^{p} \frac{1}{v!} m^{(v)}(x_1)(X_{1,i} - x_1)$ for $i = 1, \ldots, n$. Define W as a diagonal matrix with entries $W_i(x_1) = K_h(X_{1,i} - x_1)$, and let $\mathbf{e}_v \in \mathbb{R}^{p+1}$ be a unit vector of zero with the $(v+1)$th element replaced by 1. Also define $\bar{b}_k(\tilde{x}) = \beta_0 + m_{-1}(\tilde{x}) - \bar{P}_k(\tilde{x})^T \theta_k$, where $\bar{P}_k(\tilde{x}) = \left[1, \underbrace{0, \cdots, 0}_{k}, p_1(x_2), \cdots, p_k(x_2), \right.$ then $\widetilde{\beta}_0 + \tilde{m}_{-1}(\tilde{\mathbf{X}}_i) = \bar{P}_k(\tilde{\mathbf{X}}_i)^T \widehat{\theta}_k$, where $\widetilde{\beta}_0$ is the first component of $\widehat{\theta}_k$. And $\left(\widetilde{\beta}_0 - \beta_0 \right) +$

90

$\left(\tilde{m}_{-1}(\tilde{\mathbf{X}}_i) - m_{-1}(\tilde{\mathbf{X}}_i) \right) = \bar{P}_k(\tilde{\mathbf{X}}_i)^T \left(\hat{\theta}_k - \theta_k \right) - \bar{b}_k(\tilde{\mathbf{X}}_i).$ Define

$$G(\hat{\beta}, x_1) = Z^T W \left[Z \left(\hat{\beta} - \beta \right) - B + \bar{P}_k(\tilde{\mathbf{X}}^T \left(\hat{\theta}_k - \theta_k \right) - \bar{b}_k(\tilde{\mathbf{X}}) \right] \qquad \text{(A-14)}$$

Equation A-14 leads to

$$\hat{\beta} - \beta$$
$$= \left(Z^T W Z \right)^{-1} G(\hat{\beta}, x_1) + \left(Z^T W Z \right)^{-1} Z^T W B - \left(Z^T W Z \right)^{-1} Z^T W$$
$$\cdot \left[\bar{P}_k(\tilde{\mathbf{X}})^T \left(\hat{\theta}_k - \theta_k \right) - \bar{b}_k(\tilde{\mathbf{X}}) \right]$$

It is easy to verify that $\|B\| = o_p(\lambda_1^p)$. And from Theorem 8 of Newey (1997) $\|\bar{P}_k(\tilde{\mathbf{X}})^T \left(\hat{\theta}_k - \theta_k \right) - \bar{b}_k(\tilde{\mathbf{X}}_i)\| = O(k/n^{1/2} + k^{-p-1/2}) = o_p(n^{-1/2})$

Let $S = Z^T W Z$, The sample average of $\hat{\beta} - \beta$ is

$$A_n \equiv \frac{1}{n} \sum_{l=1}^{n} \left(\hat{\beta}(X_{1,l}) - \beta(X_{1,l}) \right)$$
$$= \frac{1}{n} \sum_{l=1}^{n} \left(S^{-1}(X_{1,l}) \right)^{-1} G(\hat{\beta}, X_{1,l}) + o_p(\lambda_1^p) + o_p(n^{-1/2}).$$

Lemma A.2.1 to A.2.3 follows from Lemma A.1, A.3 and A.4 of Li et al. (2003) by letting $d = 1$.

Lemma A.2.1.

$$\|S^{-1} - \left[(f_{X_1}(x_1)M)^{-1} - \frac{M^{-1}V(x_1)M^{-1}}{f_{X_1}^2(x_1)} \right] \| = o(1).$$

Lemma A.2.2.

$$\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{e}_1\left(f_{X_1}(X_{1,i})M\right)^{-1}G(\widehat{\beta},x_1)\| = O_p(n^{-1}\lambda_1^{-3/2}).$$

Let

$$J_1 = \frac{1}{n}\sum_{i=1}^{n}\mathbf{e}_1\frac{M^{-1}V(X_{1,i})M^{-1}}{f_{X_1}^2(X_{1,i})}G(\widehat{\beta},X_{1,i}),$$

Lemma A.2.3.

$$J_1 = \frac{1}{n}\sum_{i=1}^{n}u_iH(X_{1,i})_{2,1} + o_p(n^{-1/2}),$$

where $H(x_1) = (f_{X_1}(x_1)M)^{-1}V(x_1)$. Let $\tilde{\pi}_1 = \frac{1}{n}\sum_{i=1}^{n}m_1'(X_{1,i};q_\tau)$. By the results of Lemmas A.2.1–A.2.3, I have

$$
\begin{aligned}
\sqrt{n}\left(\widehat{\pi}_1(q_\tau) - \tilde{\pi}_1(q_\tau)\right) &= \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\beta}_2 - \beta_2\right)\right] \\
&= \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{e}_1\left(f(X_{1,i})M\right)^{-1}G(\widehat{\beta},x_1) - J_1 + o_p(\lambda_1^p) + o_p(n^{-1/2})\right] \\
&= O_p\left((n^{-1/2}\lambda_1^{-3/2}\right) - n^{1/2}J_1 + o_p(n^{1/2}\lambda_1^p) + o_p(1) \\
&= -n^{1/2}J_1 + o_p(1) \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}u_iW(X_{1,i})_{2,1} + o_p(1) \\
&\xrightarrow{d} N(0,V_\psi)
\end{aligned}
$$

Therefore,

$$\sqrt{n}\left(\widehat{\pi}_1(q_\tau) - \pi_1(q_\tau)\right)$$

$$= \sqrt{n}\left(\widehat{\pi}_1(q_\tau) - \tilde{\pi}_1(q_\tau)\right) + \sqrt{n}\left(\tilde{\pi}_1(q_\tau) - \pi_1(q_\tau)\right)$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(m'_1(X_{1,i;q_\tau}) - \pi(q_\tau) - u_i W(X_{1,i})_{2,1}\right) + o_p(1),$$

which proves part (a) of Theorem 1. Part (b) follows from a central limit theorem.

A.2.2 Proof of Theorem 2

Define $\widehat{f'_Y}(q)$ as the derivative of $\widehat{f_Y}$ with respect to q,

$$\widehat{f'_Y}(q) = -\frac{1}{nh^2}\sum_{i=1}^{n}K'(\frac{Y_i - q}{h}).$$

Lemma A.2.4. *If the assumptions of Theorem 1 hold, then:*

$$\widehat{\delta}_1(\widehat{q}_\tau) - \widehat{\delta}_1(q_\tau) = \left(\frac{\widehat{\pi}_1^{\prime *}(q_\tau)}{\widehat{f_Y}(q_\tau)} - \frac{\widehat{\pi}_1(q_\tau)}{\widehat{f_Y}(q_\tau)} \cdot \frac{\widehat{f'_Y}(q_\tau)}{\widehat{f_Y}(q_\tau)}\right) \cdot (\widehat{q}_\tau - q_\tau) + \mathbf{R}_2,$$

where

$$\widehat{\pi}_1^{\prime *}(q_\tau) = -\frac{1}{n}\sum_{l=1}^{n}\mathbf{e}_1^T S^{-1}(X_{1,l}) \cdot \sum_{i=1}^{n}Z_i^T(X_{1,l})W_i(X_{1,l}) \cdot f_{Y|X}(q|\mathbf{X}_i)$$

and the remainder term $\mathbf{R}_2 = O_p(n^{-1})$

Proof.

$$\widehat{\delta}_1(\widehat{q}_\tau) - \widehat{\delta}_1(q_\tau) = \frac{\widehat{\pi}_1(\widehat{q}_\tau)}{\widehat{f_Y}(\widehat{q}_\tau)} - \frac{\widehat{\pi}_1(q_\tau)}{\widehat{f_Y}(q_\tau)}$$

$$= \frac{\widehat{f_Y}(q_\tau) \cdot (\widehat{\pi}(\widehat{q}_\tau) - \widehat{\pi}(q_\tau)) - \widehat{\pi}_1(q_\tau) \cdot (\widehat{f_Y}(\widehat{q}_\tau) - \widehat{f_Y}(q_\tau))}{\widehat{f_Y}(\widehat{q}_\tau) \cdot \widehat{f_Y}(q_\tau)}$$

$$= \frac{\widehat{\pi}_1(\widehat{q}_\tau) - \widehat{\pi}_1(q_\tau)}{\widehat{f_Y}(q_\tau)} - \frac{\widehat{\pi}_1(q_\tau)}{\widehat{f_Y}(q_\tau)} \cdot \left(\frac{\widehat{f_Y}(\widehat{q}_\tau) - \widehat{f_Y}(q_\tau)}{\widehat{f_Y}(q_\tau)}\right) + \mathbf{R}_A,$$

93

where

$$\mathbf{R}_A = -\left(\widehat{f}_Y(q_\tau) \cdot (\widehat{\pi}_1(\widehat{q}_\tau) - \widehat{\pi}_1(q_\tau)) - \widehat{\pi}_1(\widehat{q}_\tau) \cdot (\widehat{f}_Y(\widehat{q}_\tau) - \widehat{f}_Y(q_\tau))\right) \cdot \left(\frac{\widehat{f}_Y(\widehat{q}_\tau) - \widehat{f}_Y(q_\tau)}{\widehat{f}_Y(\widehat{q}_\tau) \cdot \widehat{f}_Y^2(q_\tau)}\right)$$

$$= O_p\left(\|\widehat{\pi}_1(\widehat{q}_\tau) - \widehat{\pi}_1(q_\tau)\| \cdot |\widehat{f}_Y(\widehat{q}_\tau) - \widehat{f}_Y(q_\tau)|\right) + O_p\left(|\widehat{f}_Y(\widehat{q}_\tau) - \widehat{f}_Y(q_\tau)|^2\right)$$

$$= O_p(n^{-1})$$

$$\widehat{\pi}_1(\widehat{q}_\tau) - \widehat{\pi}_1(q_\tau) = \frac{1}{n}\sum_{l=1}^n \mathbf{e}_1^T S^{-1}(X_{1,l}) \cdot \sum_{i=1}^n Z_i^T(X_{1,l}) W_i(X_{1,l}) \cdot (\mathbf{I}\{Y_i - (\widehat{q}_\tau - q_\tau) \leq q_\tau\} - \mathbf{I}\{Y_i \leq q_\tau\})$$

$$= \frac{1}{n}\sum_{l=1}^n \mathbf{e}_1^T S^{-1}(X_{1,l}) \cdot \sum_{i=1}^n Z_i^T(X_{1,l}) W_i(X_{1,l})$$

$$\cdot \mathbf{E}\left[(\mathbf{I}\{Y_i - (\widehat{q}_\tau - q_\tau) \leq q_\tau\} - \mathbf{I}\{Y_i \leq q_\tau\})|\mathbf{X}_i\right] + \mathbf{R}_B,$$

where

$$\mathbf{R}_B = \frac{1}{n}\sum_{l=1}^n \mathbf{e}_1^T S^{-1}(X_{1,l}) \cdot \sum_{i=1}^n Z_i^T(X_{1,l}) W_i(X_{1,l})$$

$$\cdot \{\mathbf{I}\{Y_i - (\widehat{q}_\tau - q_\tau) \leq q_\tau\} - \mathbf{E}\left[\mathbf{I}\{Y_i - (\widehat{q}_\tau - q_\tau) \leq q_\tau\}|\mathbf{X}_i\right] - \mathbf{I}\{Y_i \leq q_\tau\} + \mathbf{E}\left[\mathbf{I}\{Y \leq q_\tau\}|\mathbf{X}_i\right]\}$$

$$= O_p(n^{-1/2}) \cdot (\widehat{q}_\tau - q_\tau)$$

$$= O_p(n^{-1})$$

One can also write

$$\widehat{\pi}_1(\widehat{q}_\tau) - \widehat{\pi}_1(q_\tau) = \mathbf{E}\left[\frac{d}{dX}\{\mathbf{E}\left[(\mathbf{I}\{Y - (\widehat{q}_\tau - q_\tau) \leq q_\tau\} - \mathbf{I}\{Y \leq q_\tau\})|\mathbf{X}\right]\}\right] + \mathbf{R}_C,$$

where

$$\mathbf{R}_C = \frac{1}{n}\sum_{l=1}^{n}\mathbf{e}_1^T S^{-1}(X_{1,l}) \cdot \sum_{i=1}^{n} Z_i^T(X_{1,l})W_i(X_{1,l}) \cdot \mathbf{E}\left[\left(\mathbf{I}\{Y_i - (\widehat{q}_\tau - q_\tau) \le q_\tau\} - \mathbf{I}\{Y_i \le q_\tau\}\right)|\mathbf{X}_i\right]$$

$$- \mathbf{E}\left[\frac{d}{dX}\left\{\mathbf{E}\left[\left(\mathbf{I}\{Y - (\widehat{q}_\tau - q_\tau) \le q_\tau\} - \mathbf{I}\{Y \le q_\tau\}\right)|\mathbf{X}\right]\right\}\right]$$

$$= \frac{1}{n}\sum_{l=1}^{n}\mathbf{e}_1^T S^{-1}(X_{1,l}) \cdot \sum_{i=1}^{n} Z_i^T(X_{1,l})W_i(X_{1,l}) \cdot \mathbf{E}\left[\mathbf{I}\{Y - (\widehat{q}_\tau - q_\tau) \le q_\tau\}|\mathbf{X}_i\right]$$

$$- \mathbf{E}\left[\frac{d}{dX}\left\{\mathbf{E}\left[\mathbf{I}\{Y - (\widehat{q}_\tau - q_\tau) \le q_\tau\}|\mathbf{X}\right]\right\}\right]$$

$$- \frac{1}{n}\sum_{l=1}^{n}\mathbf{e}_1^T S^{-1}(X_{1,l}) \cdot \sum_{i=1}^{n} Z_i^T(X_{1,l})W_i(X_{1,l}) \cdot \mathbf{E}\left[\mathbf{I}\{Y \le q_\tau\}|\mathbf{X}_i\right] + \mathbf{E}\left[\frac{d}{dX}\left\{\mathbf{E}\left[\mathbf{I}\{Y \le q_\tau\}|\mathbf{X}\right]\right\}\right]$$

$$= O_p(n^{-1/2}) \cdot (\widehat{q}_\tau - q_\tau)$$

$$= O_p(n^{-1}),$$

since the average derivative estimator should converge at the parametric rate. Thus

$$\widehat{\pi}_1(\widehat{q}_\tau) - \widehat{\pi}_1(q_\tau) = \mathbf{E}\left[\frac{d}{dX}\left\{\Pr\left[Y - (\widehat{q}_\tau - q_\tau) \le q_\tau|\mathbf{X}\right] - \Pr\left[Y \le q_\tau|\mathbf{X}\right]\right\}\right] + O_p(n^{-1}).$$

Now define $U_n/\sqrt{n} = \widehat{q}_\tau - q_\tau$. For any $\widetilde{q}_\tau$ between $\widehat{q}_\tau$ and q_τ I can write $\widetilde{q}_\tau - q_\tau = C \cdot U_n/\sqrt{n} = O_p(n^{-1/2})$ for some constant C. Therefore,

$$\left|\Pr\left[Y \le \widehat{q}_\tau|\mathbf{X}\right] - \Pr\left[Y \le q_\tau|\mathbf{X}\right] - f_{Y|\mathbf{X}}(q_\tau|\mathbf{X}) \cdot (\widehat{q}_\tau - q_\tau)\right|$$

$$= \left|\mathbf{E}\left[\Pr\left[Y \le q_\tau + U_n/\sqrt{n}|\mathbf{X}, U_n\right] - \Pr\left[Y \le q_\tau|\mathbf{X}, U_n\right] - f_{Y|\mathbf{X},U_n}(q_\tau|\mathbf{X}, U_n) \cdot U_n/\sqrt{n}|\mathbf{X}\right]\right|$$

$$= \left|\mathbf{E}\left[\left(f_{Y|\mathbf{X},U_n}(\widetilde{q}_\tau|\mathbf{X}, U_n) - f_{Y|\mathbf{X},U_n}(q_\tau|\mathbf{X}, U_n)\right) \cdot U_n/\sqrt{n}|\mathbf{X}\right]\right|$$

$$\le \frac{C}{n}\left|\mathbf{E}\left[f'_{Y|\mathbf{X},U_n}(\widetilde{q}_\tau^*|\mathbf{X}, U_n) \cdot U_n^2|\mathbf{X}\right]\right|$$

$$= O_p(n^{-1}).$$

Therefore,

$$\widehat{\pi}_1(\widehat{q}_\tau) - \widehat{\pi}_1(q_\tau) = \mathbf{E}\left[\frac{\mathrm{d}}{\mathrm{d}X} f_{Y|X}(q_\tau|\mathbf{X})\right] \cdot (\widehat{q}_\tau - q_\tau) + O_p(n^{-1}),$$

And

$$\widehat{\pi}_1(\widehat{q}_\tau) - \widehat{\pi}_1(q_\tau) = \widehat{\pi}_1'^*(q_\tau) \cdot (\widehat{q}_\tau - q_\tau) + \mathbf{R}_D,$$

where

$$\mathbf{R}_D = \left(\widehat{\pi}_1'^*(q_\tau) - \widehat{\pi}_1'(q_\tau)\right) \cdot (\widehat{q}_\tau - q_\tau) + O_p(n^{-1})$$

According to theorem 1, the parametric rate of convergence of average derivative estimator would prevail, thus

$$\mathbf{R}_D = O_p(n^{-1})$$

Follow the same line of reasoning used before,

$$
\begin{aligned}
&|f_Y(\widehat{q}_\tau) - f_Y(q_\tau) - f_Y'(q_\tau) \cdot (\widehat{q}_\tau - q_\tau)| \\
&= |\frac{1}{nh}\sum_{l=1}^{n} K\left(\frac{Y_i - \widehat{q}_\tau}{h}\right) - K\left(\frac{Y_i - q_\tau}{h}\right) + \frac{1}{nh^2}\sum_{l=1}^{n} K'\left(\frac{Y_i - q_\tau}{h}\right) \cdot (\widehat{q}_\tau - q_\tau)| \\
&= |-\frac{1}{nh^2}\sum_{l=1}^{n} \left(K'\left(\frac{Y_i - \widetilde{q}_\tau}{h}\right) - K'\left(\frac{Y_i - q_\tau}{h}\right)\right) \cdot (\widehat{q}_\tau - q_\tau)| \\
&\leq C \cdot (\widehat{q}_\tau - q_\tau)^2 \\
&= O_p(n^{-1}).
\end{aligned}
$$

Therefore,

$$\widehat{\delta}_1(\widehat{q}_\tau) - \widehat{\delta}_1(q_\tau) = \left(\frac{\widehat{\pi}_1'^*(q_\tau)}{\widehat{f}_Y(q_\tau)} - \frac{\widehat{\pi}_1(q_\tau)}{\widehat{f}_Y(q_\tau)} \cdot \frac{\widehat{f}_Y'(q_\tau)}{\widehat{f}_Y(q_\tau)}\right) \cdot (\widehat{q}_\tau - q_\tau) + \mathbf{R}_2,$$

where

$$\mathbf{R}_2 = \mathbf{R}_A + \frac{1}{\widehat{f}_Y(q_\tau)} \cdot \mathbf{R}_D - \frac{\widehat{\pi}_1(q_\tau)}{\widehat{f}_Y^2(q_\tau)} \cdot \left(\widehat{f}_Y(\widehat{q}_\tau) - \widehat{f}_Y(q_\tau) - \widehat{f}'_Y(q_\tau) \cdot (\widehat{q}_\tau - q_\tau) \right)$$

$$= O_p(n^{-1}).$$

$\square$

Lemma A.2.5. *If the assumptions of Theorem 1 holds, then:*

$$\delta_1(\widehat{q}_\tau) - \delta_1(q_\tau) = \left(\frac{\pi'_1(q_\tau)}{f_Y(q_\tau)} - \frac{\pi_1(q_\tau)}{f_Y(q_\tau)} \cdot \frac{f'_Y(q_\tau)}{f_Y(q_\tau)} \right) \cdot (\widehat{q}_\tau - q_\tau) + \mathbf{R}_3,$$

where

$$\pi'_1(q_\tau) = E\left[\mathrm{d}\frac{f_{Y|X}(q|\mathbf{X})}{\mathrm{d}X_1} \right],$$

and the remainder term $\mathbf{R}_3 = O_p(n^{-1/2}h^2) + O_p(n^{-1}h^{-3/2}) + O_p(n^{-3/2}h^{-3}).$

Proof. It is straightforward to have

$$\mathbf{R}_3 = \left(\left(\frac{\widehat{\pi}'^*_1(q_\tau)}{\widehat{f}_Y(q_\tau)} - \frac{\widehat{\pi}_1(q_\tau)}{\widehat{f}_Y(q_\tau)} \cdot \frac{\widehat{f}'_Y(q_\tau)}{\widehat{f}_Y(q_\tau)} \right) - \left(\frac{\pi'_1(q_\tau)}{f_Y(q_\tau)} - \frac{\pi_1(q_\tau)}{f_Y(q_\tau)} \cdot \frac{f'_Y(q_\tau)}{f_Y(q_\tau)} \right) \right) \cdot (\widehat{q}_\tau - q_\tau) + O_p(n^{-1})$$

$$= \left(\frac{\widehat{\pi}'^*_1(q_\tau)}{\widehat{f}_Y(q_\tau)} - \frac{\pi'_1(q_\tau)}{f_Y(q_\tau)} - \left[\left(\frac{\widehat{\pi}_1(q_\tau)}{\widehat{f}_Y(q_\tau)} \cdot \frac{\widehat{f}'_Y(q_\tau)}{\widehat{f}_Y(q_\tau)} \right) - \left(\frac{\pi_1(q_\tau)}{f_Y(q_\tau)} \cdot \frac{f'_Y(q_\tau)}{f_Y(q_\tau)} \right) \right] \right) \cdot (\widehat{q}_\tau - q_\tau) + O_p(n^{-1})$$

Note the terms multiplying $\widehat{q}_\tau - q_\tau$:

$$\frac{\widehat{\pi}'^*_1(q_\tau)}{\widehat{f}_Y(q_\tau)} - \frac{\pi'_1(q_\tau)}{f_Y(q_\tau)} = \frac{1}{f_Y(q_\tau)} \cdot (\widehat{\pi}'^*_1(\widehat{q}_\tau) - \pi'_1(q_\tau)) - \frac{\pi'_1(q_\tau)}{f_Y^2(q_\tau)} \cdot \left(\widehat{f}_Y(q_\tau) - f_Y(q_\tau) \right) + \mathbf{R}_E$$

$$= O_p(n^{-1/2}) + O_p\left((nh)^{-1/2} + O(h^2) \right) + \mathbf{R}_E,$$

97

where

$$\mathbf{R}_E = - \left(\frac{\widehat{\pi}_1'^*(q_\tau) - \pi_1'(q_\tau)}{f_Y(q_\tau)} - \frac{\pi_1'(q_\tau)}{f_Y(q_\tau)} \cdot \frac{\widehat{f}_Y(q_\tau) - f_Y(q_\tau)}{f_Y(q_\tau)} \right) \cdot \left(\frac{\widehat{f}_Y(q_\tau) - f_Y(q_\tau)}{\widehat{f}_Y(q_\tau)} \right)$$

$$= O_p((nh)^{-1}) + O_p(h^4) + O_p(n^{-1/2}h^{3/2}),$$

because $\widehat{\pi}_1'^*(q_\tau) - \pi_1'(q_\tau) = O_p(n^{-1/2})$. The second term is

$$\left(\frac{\widehat{\pi}_1(q_\tau)}{\widehat{f}_Y(q_\tau)} \cdot \frac{\widehat{f}_Y'(q_\tau)}{\widehat{f}_Y(q_\tau)} \right) - \left(\frac{\pi_1(q_\tau)}{f_Y(q_\tau)} \cdot \frac{f_Y'(q_\tau)}{f_Y(q_\tau)} \right)$$

$$= \frac{f_Y^2(q_\tau)f_Y'(q_\tau)}{f_Y^4(q_\tau)} \cdot (\widehat{\pi}_1(q_\tau) - \pi_1(q_\tau)) - 2\frac{\pi_1(q_\tau)f_Y(q_\tau)f_Y'(q_\tau)}{f_Y^4(q_\tau)} \cdot (\widehat{f}_Y(q_\tau) - f_Y(q_\tau))$$

$$+ \frac{\pi_1'(q_\tau)f_Y^2(q_\tau)}{f_Y^4(q_\tau)} \cdot (\widehat{f}_Y'(q_\tau) - f_Y'(q_\tau)) + \mathbf{R}_F$$

$$= O_p(n^{-1/2}) + O_p\left((nh)^{-1/2} + O(h^2) \right) + O_p\left((nh^3)^{-1/2} + O(h^2) \right) + \mathbf{R}_F,$$

where

$$R_F = f_Y^{-4}(q_\tau) \left[f_Y^2(q_\tau) \left(\widehat{\pi}_1(q_\tau) - \pi_1(q_\tau) \right) \left(\widehat{f}_Y'(q_\tau) - f_Y'(q_\tau) \right) \right.$$

$$\left. - \pi_1(q_\tau)f_Y'(q_\tau) \left(\widehat{f}_Y(q_\tau) - f_Y(q_\tau) \right) \left(\widehat{f}_Y(q_\tau) - f_Y(q_\tau) \right) \right] \cdot \left(\frac{\widehat{f}_Y^2(q_\tau) + f_Y^2(q_\tau)}{\widehat{f}_Y^2(q_\tau)} \right)$$

$$- f_Y^{-4}(q_\tau) \cdot \left\{ f_Y^2(q_\tau) \left[\pi_1(q_\tau) \left(\widehat{f}_Y'(q_\tau) - f_Y'(q_\tau) \right) + f_Y'(q_\tau) \left(\widehat{\pi}_1(q_\tau) - \pi_1(q_\tau) \right) \right] \right.$$

$$\left. - 2\pi_1(q_\tau)f_Y(q_\tau)f_Y'(q_\tau) \left(\widehat{f}_Y(q_\tau) - f_Y(q_\tau) \right) \right\} \cdot \left(\frac{\widehat{f}_Y^2(q_\tau) - f_Y^2(q_\tau)}{\widehat{f}_Y^2(q_\tau)} \right)$$

$$= O_p\left(|\widehat{\pi}_1(q_\tau) - \pi_1(q_\tau)| \cdot |\widehat{f}_Y'(q_\tau) - f_Y'(q_\tau)| \right) + O_p\left(|\widehat{f}_Y(q_\tau) - f_Y(q_\tau)|^2 \right)$$

$$+ O_p\left(|\widehat{f}_Y(q_\tau) - f_Y(q_\tau)||\widehat{f}_Y'(q_\tau) - f_Y'(q_\tau)| \right)$$

$$= O(h^4) + O_p((nh^3)^{-1}) + O_p(n^{-1/2}h^{1/2}).$$

because $\widehat{f}'_Y(q_\tau) - f'_Y(q_\tau) = O(h^2) + O_p((nh^3)^{-1/2})$. Therefore,

$$
\begin{aligned}
\mathbf{R}_3 &= \left[\frac{1}{f_Y(q_\tau)} \cdot (\widehat{\pi}_1'^*(\widehat{q}_\tau) - \pi_1'(q_\tau)) - \frac{\pi_1'(q_\tau)}{f_Y^2(q_\tau)} \cdot \left(\widehat{f}_Y(q_\tau) - f_Y(q_\tau) \right) + \mathbf{R}_E \right. \\
&\quad - \frac{f_Y^2(q_\tau) f_Y'(q_\tau)}{f_Y^4(q_\tau)} \cdot (\widehat{\pi}_1(q_\tau) - \pi_1(q_\tau)) + 2 \frac{\pi_1(q_\tau) f_Y(q_\tau) f_Y'(q_\tau)}{f_Y^4(q_\tau)} \cdot (\widehat{f}_Y(q_\tau) - f_Y(q_\tau)) \\
&\quad \left. - \frac{\pi_1'(q_\tau) f_Y^2(q_\tau)}{f_Y^4(q_\tau)} \cdot (\widehat{f}_Y(q_\tau) - f_Y'(q_\tau)) + \mathbf{R}_F \right] \cdot (\widehat{q}_\tau - q_\tau) + O_p(N^{-1}) \\
&= O_p(n^{-1/2}h^2) + O_p(n^{-1}h^{-3/2}) + O_p(n^{-3/2}h^{-3}).
\end{aligned}
$$

$\square$

Part (a) of Theorem 2 follows from a combination of previous results. Part (b) follows from a central limit theorem and substitute h for $o(n^{-1/5})$. Part (c) follows from arguments identical to those used to establish asymptotic normality of local polynomial estimators.