Modeling Quantile Dependence

Author: Nicholas Sim

Boston College

The Graduate School of Arts and Sciences

Department of Economics

MODELING QUANTILE DEPENDENCE

a dissertation

by

NICHOLAS C.S. SIM

submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

June 2009

MODELING QUANTILE DEPENDENCE

by

NICHOLAS C.S. SIM

Dissertation Committee:

Christopher F. Baum

Fabio Schiantarelli

Zhijie Xiao (Chair)

ABSTRACT

In recent years, quantile regression has achieved increasing prominence as a quantitative method of choice in applied econometric research. The methodology focuses on how the quantile of the dependent variable is influenced by the regressors, thus providing the researcher with much information about variations in the relationship between the covariates. In this dissertation, I consider two quantile regression models where the information set may contain quantiles of the regressors. Such frameworks thus capture the dependence between quantiles - the quantile of the dependent variable and the quantile of the regressors - which I call models of quantile dependence. These models are very useful from the applied researcher's perspective as they are able to further uncover complex dependence behavior and can be easily implemented using statistical packages meant for standard quantile regressions.

The first chapter considers an application of the quantile dependence model in empirical finance. One of the most important parameter of interest in risk management is the correlation coefficient between stock returns. Knowing how correlation behaves is especially important in bear markets as correlations become unstable and increase quickly so that the benefits of diversification are diminished especially when they are needed most.

In this chapter, I argue that it remains a challenge to estimate variations in correla-

tions. In the literature, either a regime-switching model is used, which can only estimate correlation in a finite number of states, or a model based on extreme-value theory is used, which can only estimate correlation between the tails of the returns series. Interpreting the quantile of the stock return as having information about the state of the financial market, this chapter proposes to model the correlation between quantiles of stock returns. For instance, the correlation between the 10th percentiles of stock returns, say the U.S. and the U.K. returns, reflects correlation when the U.S. and U.K. are in the bearish state. One can also model the correlation between the 60th percentile of one series and the 40th percentile of another, which is not possible using existing tools in the literature.

For this purpose, I propose a nonlinear quantile regression where the regressor is a conditional quantile itself, so that the left-hand-side variable is a quantile of one stock return and the regressor is a quantile of the other return. The conditional quantile regressor is an unknown object, hence feasible estimation entails replacing it with the fitted counterpart, which then gives rise to problems in inference. In particular, inference in the presence of generated quantile regressors will be invalid when conventional standard errors are used. However, validity is restored when a correction term is introduced into the regression model.

In the empirical section, I investigate the dependence between the quantile of U.S. MSCI returns and the quantile of MSCI returns to eight other countries including Canada and major equity markets in Europe and Asia. Using regression models based on the Gaussian and Student-t copula, I construct correlation surfaces that reflect how the correlations between quantiles of these market returns behave. Generally, the correlations tend to rise gradually when the markets are increasingly bearish, as reflected by the fact that the returns are jointly declining. In addition, correlations tend to rise when markets are increasingly bullish, although the magnitude is smaller than the increase associated with bear markets.

The second chapter considers an application of the quantile dependence model in empirical macroeconomics examining the money-output relationship. One area in this line of research focuses on the asymmetric effects of monetary policy on output growth. In particular, letting the negative residuals estimated from a money equation represent contractionary monetary policy shocks and the positive residuals represent expansionary shocks, it has been widely established that output growth declines more following a contractionary shock than it increases following an expansionary shock of the same magnitude. However, correctly identifying episodes of contraction and expansion in this manner presupposes that the true monetary innovation has a zero population mean, which is not verifiable.

Therefore, I propose interpreting the quantiles of the monetary shocks as having information about the monetary policy stance. For instance, the 10th percentile shock reflects a restrictive stance relative to the 90th percentile shock, and the ranking of shocks is preserved regardless of shifts in the shock's distribution. This idea motivates modeling output growth as a function of the quantiles of monetary shocks. In addition, I consider modeling the quantile of output growth, which will enable policymakers to ascertain whether certain monetary policy objectives, as indexed by quantiles of monetary shocks, will be more effective in particular economic states, as indexed by quantiles of output growth. Therefore, this calls for a unified framework that models the relationship between the quantile of output growth and the quantile of monetary shocks.

This framework employs a power series method to estimate quantile dependence. Monte Carlo experiments demonstrate that regressions based on cubic or quartic expansions are able to estimate the quantile dependence relationships well with reasonable bias properties and root-mean-squared errors. Hence, using the cubic and quartic regression models with M1 or M2 money supply growth as monetary instruments, I show that the right tail of the output growth distribution is generally more sensitive to M1 money supply shocks, while both tails of output growth distribution are more sensitive than the center is to M2 money supply shocks, implying that monetary policy is more effective in periods of very low and very high growth rates. In addition, when non-neutral, the influence of monetary policy on output growth is stronger when it is restrictive than expansive, which is consistent with previous findings on the asymmetric effects of monetary policy on output.

ACKNOWLEDGEMENTS

I am especially grateful to my dissertation chair, Professor Zhijie Xiao, for his generously sharing his wisdom on quantile regression, and to Professors Fabio Schiantarelli and Christopher F. Baum, readers of my dissertation, for giving me much support and help in writing my dissertation. I cannot thank them enough for their generosity.

I would like to thank Professor Fabio Ghironi for giving me the opportunity to work and learn from him as a summer research assistant and Professor Francis McLaughlin from whom I have learned many interesting facts both in economics and outside of it. I would also like to Professor Susanto Basu for his enlightening research conversations and advice. My special thanks also go to my fellow graduate students at Boston College for their warm and supportive friendship.

Finally, I would like to express my debt to my parents John and June, my brothers Stephen and Edward, and especially to my wife Sabrina for their constant encouragement.

CHAPTER 1

Modeling Quantile Dependence: Estimating the Correlations of International Stock Returns†

Nicholas C.S. Sim*

ABSTRACT

We propose a quantile regression method to estimate the correlation between quantiles of international stock returns where the regressor is itself a conditional quantile. Since the true conditional quantile regressor is unknown, feasible estimation entails using generated values, leading us to examine the implications for inference in quantile regression with generated regressors. Using the Gaussian and Student-t copula, we investigate the dependence between quantiles of U.S. MSCI returns with quantiles of returns to eight other countries, and find that correlations increase gradually when returns are jointly declining or increasing, although the increase is larger in bear than in bull markets.

JEL Classification: G15, C01, E30.
Key Words: Quantile Regression, Quantile Dependence, Generated Regressors, Copula, Stock Returns Correlation.

†I am indebted to Professors Christopher Baum, Fabio Schiantarelli and Zhijie Xiao for their invaluable discussions and immense support. I am also grateful to Professor Shuxia Sun for many enlightening exchanges, as well as Professors Arthur Lewbel, Shannon Seitz, Matteo Iacoviello and Susanto Basu and students in the Boston College dissertation workshop for their helpful comments.

*Corresponding address: Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467 USA. Tel: +1-617-552-6347, Fax: +1-617-552-2308, Email: nicholas.sim@bc.edu.

1 Introduction

By now, it is well-known that the correlations of international stock returns tend to be larger in bear than in bull markets (Longin and Solnik, 2001; Hu, 2006).[1] Recognizing that correlation is asymmetric is important for risk management. For the investor who diversifies across international stock markets, the increase in correlation during bear markets implies that the benefits of diversification are diminished especially when they are most needed. The welfare losses are also significant for the investor who ignores how dependence between stock markets changes across different financial regimes (Ang and Bekaert, 2001).

To estimate changes in the correlations, the stock returns data may be fitted with a mixture of joint distributions, such as a mixture of bivariate Gaussian distributions (Kim and Finger, 2000) or more sophisticatedly in a regime switching framework (Ang and Bekaert, 2001) where the correlation may differ across a finite number of states. The alternative to fitting correlations using mixture models is to estimate how the stock returns series are correlated in the tails. This is in line with the contribution of Longin and Solnik (2001), who employed Extreme Value Theory, utilizing the result that the only non-degenerate distribution in the tail is generalized Pareto regardless of the data generating process. Following their methodology, one first models the univariate tail distribution as an extreme value distribution, then the dependence between the tails by specifying a dependence function. While Longin and Solnik's method provides a "smoother" picture of correlation as opposed to regime-switching models, it only estimates correlations between the tails of the returns series. Moreover, the extreme value distribution, i.e. generalized Pareto, may not be a sufficiently good approximation for distribution of observations that are located away from the tails.

Understanding the limitations of both methods, how can correlations be modeled with-

[1] See Ang and Chen (2002) for an example of asymmetric correlation between the U.S. portfolios' returns and the U.S. aggregate returns.

out restricting our focus on a small number of states as typically required by regime-switching frameworks, where at the same time these correlations may also be modeled for returns located in any point on their distributions other than the extremes? This question is addressed using the insight that the quantiles of the stock returns may be informative about the states of the financial markets. For instance, the 10th percentile of a monthly stock return may be observed when the stock market is bearish for that month. Conversely, the 90th percentile of return may be observed when the market is bullish. Therefore, if the correlation between the quantiles of international stock returns may be estimated, then this correlation may vary continuously across the different states of the financial markets so that the data may reveal to us any potentially interesting dependence behavior.

This motivates a statistical framework that can estimate the dependence between quantiles. To do so, we employ quantile regression, which has been featured prominently in recent applied econometric research. Unlike ordinary least squares regression that examines how a set of regressors influence the conditional mean of the dependent variable, quantile regression examines how these regressors may influence the conditional quantile of the dependent variable. For our empirical purpose, we will extend the standard quantile regression framework. In particular, we will construct a model that captures the association between the τ_X^{th} conditional quantile of X, which we will call the quantile regressor, with the τ_Y^{th} conditional quantile of Y, the dependent variable. This relationship can be written as $Q_Y(\tau_Y|Q_X(\tau_X|Z))$, where the quantile of X is in turn modeled using a set of exogenous variables Z. We approach this problem by first considering a framework that expresses the dependence of Y on $Q_X(\tau_X|Z)$, an unknown variable which can be consistently estimated. To feasibly estimate $Q_Y(\tau_Y|Q_X(\tau_X|Z))$, we will first obtain the fitted quantiles of X, then employ the fitted values in place of the true conditional quantiles to estimate the dependence between the quantiles.

By employing the fitted quantiles of X, we are confronted with the issue of using a generated regressor. The presence and implications of generated regressors in two-step ordinary least squares or maximum likelihood estimation have been considered by Pagan

(1984) and Murphy and Topel (1985). Typically, generated regressors do not cause problems for consistency, although some adjustments must be made to the standard errors to conduct inference correctly. Similarly in the quantile regression context, substituting the unknown quantiles with their generated counterparts introduces an error term that asymptotically converges to zero. The fact that this term converges to zero ensures that the estimated second-step parameters are consistent. However, without controlling for this asymptotically negligible component in the actual estimation itself, we show that conventional standard errors of quantile regressions will be incorrect as the true asymptotic covariance matrix will contain an additional term that is related to the covariance of the generated regressors. While one may estimate the model by simply replacing the unknown quantiles with the fitted ones and then fix standard errors for inference, a better alternative is to also fix the model itself so that the asymptotic covariance matrix reduces to the conventional case in quantile regression.

For the latter, we consider Taylor expansion of the above-mentioned error term and show that a first order Taylor expansion is sufficient for the covariance matrix of the generated regressors model to have the same properties as that of the standard case.[2] In addition to efficiency considerations, we also derive the linear (Bahadur) representation under fairly general conditions, for instance, for a nonlinear model with weakly dependent errors, and demonstrate that this linear representation generalizes the one that is derived under the assumption of independently distributed errors as in Koenker and Zhao (1994) and Zhou and Portnoy (1996). To the best of our knowledge, the linear representation given in this paper is a new result, which complements the asymptotic analysis of Oberhofer and Haupt (2006) for nonlinear quantile regression under the same weak dependence consideration in the form of α-mixing. In general, even though this paper is concerned about estimating the dependence between the quantiles of stock returns, it also provides some useful findings on the implications for inference and the method of efficient estimation in the presence of the generated regressors in quantile regression, which have yet to be discussed in the literature.

[2]This is similar to the strategy of McKenzie and McAleer (1997) in addressing the issue of generated regressors for nonlinear ordinary least squares regressions.

For the empirical work, we estimate the correlations between eight pairs of country-level returns based on the MSCI: the correlations between the quantiles of U.S. returns and the quantiles of returns in Australia, Canada, France, Germany, U.K., Japan, Hong Kong and Singapore. To this end, we employ a regression model that is derived from the *copula* in the same way as Bouyè and Salmon (2003) did in their copula quantile regression framework. Copulas provide a convenient way to model joint dependence, as every joint distribution can be written as a copula function over the marginal distributions. Hence, the task of modeling complicated relationships can be flexibly decomposed into specifying the copula and the marginal distributions separately. Importantly, using the copula to model dependence is advantageous because the copula parameter maps into a rank correlation measure known as Kendall's Tau. In the special case of a Gaussian or Student-t copula, the copula parameter is just the correlation coefficient. Therefore, formulating models of quantile dependence using a copula-based approach will give us the ability to detect changes in correlation across different states of the financial markets, captured by the quantiles of the stock returns.

After estimating the Gaussian and Student-t copula regression model, we construct correlation surfaces that demonstrate how correlation varies depending on where the returns are located in their respective distributions. Generally, correlations tend to be increasing gradually as markets are becoming more bearish, reflected by the case where the stock returns are jointly declining. At the center of the distributions, correlations are typically the lowest, but begin to rise as the quantiles of returns move further into the right tails so that larger correlations are also found in bullish markets. Therefore, unlike previous findings, this implies that both bear and bull market correlations are greater than those in the typical quiet environments. Nevertheless, correlation asymmetry still exists as correlations in bear markets remain larger than those associated with bull markets.

The rest of the paper is organized as follows. Section 2 discusses the methodology proposed by this paper. Section 3 presents the asymptotic analysis of the proposed estimator, where the asymptotic distribution of the estimator will be used for inference. Section

4 contains a brief discussion on the copula methodology, compares the new methodology with previous related work, and documents our empirical findings on the correlations of the international stock returns. Section 5 concludes.

2 The Model

Notations used frequently in the paper are defined as follows. We let n be the observation index and N be the sample size. For variables X and Y, let their distribution functions be represented by F_X and F_Y, their density functions by f_X and f_Y, and their quantiles by Q_X and Q_Y. In addition, let $\|\cdot\|$ be the Euclidean norm, i.e. $\|A\| = tr(A'A)$. The prototypical model examined in this paper has the form of

$$Q_Y(\tau_Y|Q_X(\tau_X|Z_n)) = h(Q_X(\tau_X|Z_n), \beta_{\tau_X}(u_n)) \tag{1}$$

$$Q_X(\tau_X|Z_n) = Z_n'\gamma(\tau_X) \tag{2}$$

where (2) can be obtained from

$$X_n = Z_n'\gamma(w_n) \tag{3}$$

by setting w_n to $F_w^{-1}(\tau_X)$. The innovation terms w_n and u_n are assumed to be independent of each other and are each weakly dependent across n as defined in Section 3. Furthermore, w_n and u_n are distributed with distribution functions $F_{w,n}$ and $F_{u,n}$ respectively. When the innovation terms are homoskedastic, the distribution functions will be denoted as F_w and F_u. In our empirical work in Section 4, Y and X are $N \times 1$ vectors representing stock returns. Z is a $N \times k$ matrix of exogenous variables, Z_n is a $k \times 1$ vector and γ is a $k \times 1$ parameter vector. X_n is assumed to be independent of u_n while Z_n is independent of both w_n and u_n for all n. The nonlinear function h is assumed to be strongly monotonic in u_n, nondecreasing in $Q_X(\tau_X|Z_n)$, and twice differentiable in both $Q_X(\tau_X|Z_n)$ and β. The parameter of interest is $\beta_{\tau_X}(u_n)$, which captures the dependence of Y on the τ_X conditional quantile of X. From now on, the τ_X subscript on β will be omitted to simplify notation.

Estimating (1) will be our main focus while the auxiliary equation (2) serves the purpose of obtaining estimates for $Q_X(\tau_X|Z_n)$. $\beta(\tau_Y)$ is estimated in two steps: the first step estimates (2) to obtain $\hat{Q}_X(\tau_X|Z_n)$ and the second step uses $\hat{Q}_X(\tau_X|Z_n)$ to estimate $\beta(\tau_Y)$ in (1).

For the reader familiar with Ma and Koenker's (2006) work in examining the relationships between quantiles, our setup appears to be very similar to theirs with one key difference. We implicitly assume in (1) that Y depends on the conditional quantile of X whereas in their model Y depends on X and not its conditional quantile. In the former setup, no attempt is made to model causal or structural relationships, which is suitable for our empirical objective of modeling the correlation between quantiles of stock returns since the correlation merely captures non-causal dependence behavior.

Note that conditional quantile function of Y can be estimated from (1) using standard quantile regression models if $Q_X(\tau_X|Z_n)$ is known. As $Q_X(\tau_X|Z_n)$ is unobservable, (1) can only be feasibly estimated using $\hat{Q}_X(\tau_X|Z_n)$ first obtained from an auxiliary equation. Replacing $Q_X(\tau_X|Z_n)$ with $\hat{Q}_X(\tau_X|Z_n)$ introduces a generated regressor problem, which has been explored for mean regression by Pagan (1984), Murphy and Topel (1985) and McAleer and McKenzie (1997) but has yet to be addressed for quantile regression. Generally, the estimator of $\beta(\tau_Y)$ is consistent even under the presence of generated regressors. However, one has to make adjustments to the conventional standard errors for quantile regression, as the generated regressors will introduce an error term that will invalidate inference that uses these conventional standard errors. The effect of using generated quantiles in (1) can be seen by writing the conditional quantile of Y as

$$
\begin{aligned}
&Q_Y(\tau_Y|Q_X(\tau_X|Z_n)) \\
=&h(\hat{Q}_X(\tau_X|Z_n), \beta(\tau_Y)) + h(Q_X(\tau_X|Z_n), \beta(\tau_Y)) - h(\hat{Q}_X(\tau_X|Z_n), \beta(\tau_Y)) \\
=&h(\hat{Q}_X(\tau_X|Z_n), \beta(\tau_Y)) + \omega_n \qquad (4)
\end{aligned}
$$

The difference between using actual quantiles and the fitted ones is the error term ω_n. In turn, how large ω_n is depends on the difference between $\hat{Q}_X(\tau_X|Z_n)$ and $Q_X(\tau_X|Z_n)$, which converges to zero if $\hat{Q}_X(\tau_X|Z_n)$ is a consistent estimator of the true quantile. To deal with ω_n, we expand $h(Q_X(\tau_X|Z_n), \beta(\tau_Y))$ around $\hat{Q}_X(\tau_X|Z_n)$ yielding

$$\begin{aligned}\omega_n &= -h_X(\hat{Q}_X(\tau_X|Z_n), \beta(\tau_Y))Z_n'(\hat{\gamma}_N(\tau_X) - \gamma(\tau_X)) + O_p(\|\hat{\gamma}_N(\tau_X) - \gamma(\tau_X)\|^2)\\ &= -h_X(\hat{Q}_X(\tau_Y|Z_n), \beta(\tau_X))Z_n'\Delta_N + O_p(\|\Delta_N\|^2)\end{aligned}$$

where $\Delta_N = \hat{\gamma}_N(\tau_X) - \gamma(\tau_X)$, the linear representation based on (3), is

$$\Delta_N = \left(\frac{(N^{-1}\sum_{n=1}^{N} Z_n Z_n')^{-1}}{f_w(F_w^{-1}(\tau_X))} N^{-1}\sum_{n=1}^{N} Z_n \psi_{\tau_X}(w_n)\right) + o_p(N^{-1/2}) \tag{5}$$

with $\psi_{\tau_X}(w_n) = \mathbb{I}(X - Z_n'\gamma(\tau_X) < 0) - \tau_X$ and $\mathbb{I}(\cdot)$ is an indicator function. It should be noted that while (3) is specified as a linear model, it is also generalizable as a nonlinear one since we may easily obtain an analog of Δ_N for a nonlinear auxiliary equation. With Δ_N, we may control for ω by introducing a correction term as follows:

$$\begin{aligned}&Q_Y(\tau_Y|Q_X(\tau_X|Z_n))\\ =&h(\hat{Q}_X(\tau_X|Z_n), \beta(\tau_Y)) - h_X(\hat{Q}_X(\tau_X|Z_n), \beta(\tau_Y))Z_n'\Delta_N + O_p(\|\Delta_N\|^2)\\ =&\breve{h}(\hat{Q}_X(\tau_X|Z_n), \beta(\tau_Y)) + O_p(\|\Delta_N\|^2)\end{aligned} \tag{6}$$

As it will be shown, although including $h_X(\hat{Q}_X(\tau_X|Z_n), \beta(\tau_Y))Z_n'\Delta_N$ in the regression only controls for the first order component of $h(Q_X(\tau_X|Z_n), \beta(\tau_Y)) - h(\hat{Q}_X(\tau_X|Z_n), \beta(\tau_Y))$, the higher order term $\|\Delta_N\|^2$ converges quickly enough so that the effect of the higher order expansions on the asymptotic standard errors for $\hat{\beta}_N(\tau_Y)$ is negligible.

Similar to ordinary least squares, it is easy to observe that consistency is preserved whether or not the correction term is included. For instance, when the correction term is taken into account, the gradient function, G_c, from minimizing the quantile regression objective function becomes

$$G_c(\beta) = \frac{1}{N}\sum_{n=1}^{N} \psi_\tau \left(Y_n - h(\hat{Q}_X(\tau_X|Z_n), \beta) + h_X(\hat{Q}_X(\tau_X|Z_n), \beta) Z_n' \Delta_N\right) \times \left(h_\beta(\hat{Q}_X(\tau_X|Z_n), \beta) - h_{X\beta}(\hat{Q}_X(\tau_X|Z_n), \beta) Z_n' \Delta_N\right)$$

In the case where the correction term is omitted, the gradient function, G_o, becomes:

$$G_o(\beta(\tau)) = \frac{1}{N}\sum_{n=1}^{N} \psi_\tau \left(Y_n - h(\hat{Q}_X(\tau_X|Z_n), \beta)\right) \left(h_\beta(\hat{Q}_X(\tau_X|Z_n), \beta)\right)$$

The estimate of $\beta(\tau_Y)$ is found by locating the zero of the respective gradient functions. It can be argued that if $\max_n |h_X(\hat{Q}_X(\tau_X|Z_n), \beta)|$ satisfies an appropriate regularity condition that holds uniformly in β, the gradient conditions G_c and G_o will be asymptotically identical, since both $\max_n |h_X(\hat{Q}_X(\tau_X|Z_n), \beta) Z_n' \Delta_N|$ and $\max_n \|h_{X\beta}(\hat{Q}_X(\tau_X|Z_n), \beta) Z_n' \Delta_N\|$ in G_c are $o_p(1)$. Therefore, ignoring the correction term will not affect consistency in the second step regression. This conclusion is also typically observed by Pagan (1984) and Murphy and Topel (1985), as the error component introduced by the generated quantiles should disappear asymptotically.[3]

Remark: Our methodology is similar to the double-stage quantile regression model of Kim and Muller (2003) as well as the median regression framework of Amemiya (1982) and Powell (1983). These papers address the issue of endogeneity in quantile or median regressions. Despite the close resemblance between our model and the previous frameworks, the difference stems from the fact that our model is used to investigate the relationship between quantiles rather than to address the issue of endogeneity.

[3] For two-step estimators in ordinary least squares, consistency may not be achieved if the model is a function of the variances of the unknown quantities. See Model 6 of Pagan (1984).

2.1 Including an Information Set in the Second Step Regression

Consider extending (1) to include an information set $Z_{1,n}$ in h. For instance, one may wish to examine how certain macroeconomic factors such as industrial production growth and money supply growth may affect the dependence between quantiles of stock returns by modeling β as a function of $Z_{1,n}$. Generally, the conditional quantile of Y can be written as $Q_Y(\tau_Y|Q_X(\tau_X|Z_n), Z_{1,n})$. Therefore, the extension of (1) to include additional exogenous variables $Z_{1,n}$ becomes

$$Y_n = h(Q_X(\tau_X|Z_n), Z_{1,n}, \beta(u_n)) \tag{7}$$

To identify the conditional quantile of X, we require Z_n to contain at least one variable excluded from $Z_{1,n}$. The conditional quantile to be estimated becomes

$$\begin{aligned}
&Q_Y(\tau_Y|Q_X(\tau_X|Z_n), Z_{1,n}) \\
=&h(\hat{Q}_X(\tau_X|Z_n), Z_{1,n}, \beta(\tau_Y)) + h(\hat{Q}_X(\tau_X|Z_n), Z_{1,n}, \beta(\tau_Y)) - h(\hat{Q}_X(\tau_X|Z_n), Z_{1,n}, \beta(\tau_Y)) \\
=&h(\hat{Q}_X(\tau_X|Z_n), Z_{1,n}, \beta(\tau_Y)) + \omega_n
\end{aligned} \tag{8}$$

As before, we use a first order Taylor expansion of ω_n around $\hat{Q}_X(\tau_X|Z_n)$ to control for the preliminary estimation effect. Hence, this extension does not affect the asymptotic analysis for the baseline case where $Z_{1,n}$ is absent in the second step regression.

3 Inference

Inference in Section 4 will be carried out using asymptotic standard errors. Therefore, this section focuses on obtaining the asymptotic distribution of the point estimator for $\beta(\tau_Y)$. Previously, Jurečková and Procházka (1994) derived the linear representation for the nonlinear quantile regression model. Here, we provide the asymptotic analysis for nonlinear quantile regression with α-mixing innovation terms. We first obtain the linear representation for the estimator where the correction term is not included, then derive the linear representation for the correction model. We impose some fairly high level assumptions, including consistency of $\hat{\gamma}_N(\tau_X)$ as well as $\hat{\beta}_N(\tau_Y)$ or $\hat{\hat{\beta}}_N(\tau_Y)$ depending on whether the correction term is included. For consistency in nonlinear quantile regression, the reader is referred to Oberhofer and Haupt (2006). The assumptions below are required for root-N consistency as well as for deriving the linear representation for $\hat{\beta}_N(\tau)$ or $\hat{\beta}_N(\tau)$:

A1. The regression model consists of

$$Y_n = h(Q_X(\tau_X|Z_n), \beta(u_n))$$

and

$$X_n = Z_n'\gamma(w_n)$$

such that $h : \Re \times \Re^p \to \Re$ and $Q_X(\tau_X|Z_n) = Z_n'\gamma(\tau_X)$, where Z_n is a $q \times 1$ vector with its l^{th} element represented by $Z_n^(l)$. The innovation terms u_n and w_n are independent of each other and are weakly dependent across n with mixing coefficients $\alpha_u(N) = O(N^{-\zeta})$ and $\alpha_w(N) = O(N^{-\zeta})$ for $\zeta > (4.5 + 2p)/(1 - \nu)$ for some $\nu > 0$.[4]. Z_n is mixing random variable of size $-\zeta$.

A2. Let Θ and Γ be compact sets where $\beta(u_n) \in \Theta$ and $\gamma(w_n) \in \Gamma$ are $p \times 1$ and $q \times 1$ vectors, strictly monotonic in u_n and w_n, respectively.

The weak dependence assumption in A1 is adopted by Sun (2006) while A2 is a standard condition. For the next assumption, let $\Delta < \infty$ represent some generic constant. We define $h_\beta^{(j)}$ as the first partial derivative of h with respect to the j^{th} element of β, $h_{\beta\beta}^{(jk)}$ as the second partial derivative of h with respect to the j^{th} and k^{th} elements of β. Equivalently, $h_\beta^{(j)}$ is the j element of the vector h_β and $h_{\beta\beta}^{(jk)}$ is the jk element of the matrix $h_{\beta\beta}$. In addition, expressing $Q_X(\tau|Z)$ as the $N \times 1$ vector of conditional quantiles, we may likewise define $h_X^{(n)}$ as the partial derivative of h with respect to the n^{th} element of $Q_X(\tau|Z)$, and

[4] The case where u_n is independent across n is reflected by $\zeta \to \infty$. The same is true also for w_n.

$h_{XX}^{(nm)}$ as the second partial derivative of h with respect to the n^{th} and m^{th} elements of $Q_X(\tau|Z)$. However, there will no confusion if we drop the n superscript as h_X and h_{XX} may be interpreted as a scalar.

For brevity, even though the following conditions are stated for h only, we here assume that A3 to A7 are also imposed with $\breve{h}$ replacing h, where $\breve{h}$ is the function used in the error-correction model.

A3. The moments $E|h_\beta^{(k)}(Z_n'\gamma, \beta(\tau))|^4$, $E|h_X(Z_n'\gamma, \beta(\tau)|^4$, $E|h_{XX}(Z_n'\gamma, \beta(\tau)|^4$, $E|h_{\beta\beta}^{(kk)}(Z_n'\gamma, \beta(\tau)|^4$, $E|h_{\beta X}^{(j)}(Q_{X,n,\tau}, \beta(\tau))Z_n^{(l)}|^4$, and $E|h_{\beta\beta X}^{(jk)}(\tilde{Q}_{X,n,\tau}, \tilde{\beta}(\tau))Z_n^{(l)}|^4$ are bounded above by Δ for all n uniformly in $\beta \in \Theta$ and $\gamma \in \Gamma$.

Assumption A3 are bounds on the moments, which are needed since the regressors are stochastic. Define $\tilde{u}_n = Y_n - h(Q_X(\tau_X|Z_n), \beta(\tau_Y))$ and $\tilde{w}_n = X_n - Z_n'\gamma(\tau_X)$. With this normalization, the τ_Y^{th} and τ_X^{th} quantile of $\tilde{u}_n$ and $\tilde{w}_n$ are respectively set to zero.

A4. Let C be some generic constant. The conditional distribution functions of $\tilde{u}_n$, $F_{\tilde{u},n}(\cdot)$, are absolutely continuous with continuously differentiable densities $f_{\tilde{u},n}(\cdot)$ that are bounded away from zero at all points $Q_Y(\tau_Y|Q_X(\tau_X|Z_n))$ and above by C together with their first derivatives $f'_{\tilde{u},n}(\cdot)$. Similarly, the cumulative distribution functions of $\tilde{w}_n$, $F_{\tilde{w},n}(\cdot)$, are absolutely continuous with continuously differentiable densities $f_{\tilde{w},n}$ that are bounded away from zero at all points $Q_X(\tau_Y|Z_n)$ and above by C together with their first derivatives $f'_{\tilde{w},n}(\cdot)$.

In A4, that the density functions are bounded away from zero at $\tilde{u}_n = 0$ is required for the existence of the linear representation, as the inversion of these densities are required at these points. A similar assumption to A4 is to express the distribution and density functions for Y_n and X_n instead. For instance, consider $F_{\tilde{u},n}(0) = P[\tilde{u}_n \leq 0] = P[Y_n \leq h(Q_X(\tau_X|Z_n))] = F_{Y,n}(h(Q_X(\tau_X|Z_n))$. Hence, for some constant a, the relationship $\int_{-\infty}^{a} f_{\tilde{u},n}(\tilde{u})d\tilde{u} = F_{\tilde{u},n}(a) = F_{Y,n}(a + h(Q_X(\tau_X|Z_n)) = \int_{-\infty}^{a+h(Q_X(\tau_X|Z_n)} f_{Y,n}(Y)dY$ holds as an identity. Differentiating and applying Leibnitz rule, $\partial F_{\tilde{u},n}(a)/\partial a = f_{\tilde{u},n}(a) = f_{Y,n}(a + h(Q_X(\tau_X|Z_n)) = \partial F_{Y,n}(a + h(Q_X(\tau_X|Z_n))/\partial a$. Therefore, we may consider the conditional densities in terms of Y_n and X_n as well.

Next, let $r = \zeta/(\zeta - 1)$ be a constant related to the size of mixing so that weaker dependence is characterized by an r closer to one. Also, let $\delta > 0$ be some constant. Given that the densities function are assumed to be bounded above, the following conditions are required to ensure law of large numbers for α-mixing sequences.

A5. i) $E\|h_\beta(Z_n'\gamma,\beta)h_\beta(Z_n'\gamma,\beta)'\|^{r+\delta} \leq \Delta < \infty$ for all n and uniformly in $\gamma \in \Gamma$ and $\beta \in \Theta$, and ii) $\bar{Q}_N = E[N^{-1}\sum_{n=1}^{N} f_{Y,n}(Q_{Y,n}(\tau_Y|X_n))h_\beta(\cdot,\beta)h_\beta(\cdot,\beta)']$ is uniformly positive definite.

A6. i) $E\|Z_nZ_n'\|^{r+\delta} \leq \Delta < \infty$ for all n and uniformly in $\gamma \in \Gamma$ and $\beta \in \Theta$, and ii) $\bar{Q}_{1,N} = E[N^{-1}\sum_{n=1}^{N} f_{X,n}(Q_{\tilde{w},n}(\tau_X|Z_n))Z_nZ_n']$ is uniformly positive definite.

A7. $E\|h_\beta(Z_n'\gamma,\beta)\|^{\frac{r+\delta}{2}} < \Delta < \infty$ and $E\|h_X(Z_n'\gamma,\beta)\|^{\frac{r+\delta}{2}}\Delta < \infty$ for all n and uniformly in $\gamma \in \Gamma$ and $\beta \in \Theta$.

Assumptions A5, A6 and A7 are needed for law of large numbers under α-mixing regressors. A5 ensures that the $p\times p$ positive definite matrix $Q_N = N^{-1}\sum_{n=1}^{N} f_{Y,n}(Q_{Y,n}(\tau_Y|X_n))h_\beta(\cdot,\beta)h_\beta(\cdot,\beta)'$ converges to $\bar{Q}_N$, the $q \times q$ positive matrix $Q_{N,1} = N^{-1}\sum_{n=1}^{N} f_{X,n}(Q_{X,n}(\tau_X|Z_n))Z_nZ_n'$ converges to $\bar{Q}_{1,N}$, and $N^{-1}\sum_{n=1}^{N}\|h_\beta(Z_n'\gamma,\beta)Z_n'\|^{1/2}$ and $N^{-1}\sum_{n=1}^{N}\|h_X(Z_n'\gamma,\beta)Z_n'\|^{1/2}$ are convergent as required in Lemma 5.

It is well-known that consistency is typically preserved in ordinary least squares or maximum likelihood estimation when a generated regressor is used. This observation also carries over for quantile regression. In particular, given consistency of $\hat{\gamma}_N(\tau_X)$ as well as $\hat{\beta}_N(\tau_Y)$ or $\hat{\hat{\beta}}_N(\tau_Y)$, we may establish the root-N rate of convergence for the second step estimator. Proposition 1 states the linear representation when the correction term is not included. Derivations for all of the asymptotic results in this section, including root-N convergence, are relegated to the Appendix.

Proposition 1. *Suppose a second step non-corrected model is used, that is, we estimate (1) replacing the true conditional quantiles with the fitted quantiles of X, and let $\hat{\hat{\beta}}_N(\tau_Y)$ be the point estimator of $\beta(\tau_Y)$. Under A1-A7,*

$$
\begin{aligned}
&\sqrt{N}(\hat{\hat{\beta}}_N(\tau_Y) - \beta(\tau_Y)) \\
=& - Q_N^{-1}\frac{1}{\sqrt{N}}\sum_{n=1}^{N} h_\beta(Q_X(\tau_X|Z_n), \beta(\tau_Y))\psi_{\tau_Y}(\tilde{u}_n) \\
&- Q_N{}^{-1}\frac{1}{\sqrt{N}}\sum_{n=1}^{N} f_{\tilde{u},n}(0)(\tau_Y))h_\beta(Q_X(\tau_X|Z_n), \beta(\tau_Y))h_X(Q_X(\tau_X|Z_n), \beta(\tau_Y))Z_n'(\hat{\gamma}_N(\tau_X) - \gamma(\tau_X)) \\
&+ o_p(1) \qquad (9)
\end{aligned}
$$

Remark: It may be of independent interest to provide a sharper expression for the $o_p(1)$ term in (9). By using Lemmas 3 and 4 together with Koenker and Zhao's (1994) Lemmas 6 to 8, we can write the $o_p(1)$ term as $O_p(N^{-3/4+\lambda}\log N)$, where $\lambda \in ((11/4+p)/(1+\zeta(1-\nu)), 1/2)$ with $\zeta > (4.5+2p)/(1-\nu)$, $\nu > 0$ is some constant. Rate of convergence expressed in this form follows from Raghu Raj Bahadur in his seminal work on the linear representation for sample quantiles (see Bahadur, 1966).[5] For regression quantiles, Koenker and Zhao (1994) and Zhou and Portnoy (1996) established a rate of $O_p(N^{-3/4}\log N)$ for the $o_p(1)$ term under the independence assumption. Oberhofer and Haupt (2006) examined the consistency and asymptotic normality of nonlinear quantile regression under α-mixing, but did not provide a specific rate. Therefore, the rate $O_p(N^{-3/4+\lambda}\log N)$ includes independence as a special case when $\zeta \to \infty$ so that $\lambda \to 0$.

When deriving Proposition 1, we have shown that $\hat{\hat{\beta}}_N(\tau_Y)$ is root-N consistent, since

$$N^{-1}\sum_{n=1}^{N} h(Q_X(\tau_X|Z_n), \beta(\tau_Y))\psi_\tau(\tilde{u}_n)$$

[5]See, *inter alia*, Kiefer, (1967), Sen, (1972), Babu and Singh, (1978), Yoshihara, (1995), Sun (2006).

and

$$N^{-1}\sum_{n=1}^{N} h_\beta(Q_X(\tau_X|Z_n),\beta(\tau_Y))h_X(Q_X(\tau_X|Z_n),\beta(\tau_Y))Z_n'$$

in (9) are $o_p(1)$ and $\sqrt{N}(\hat{\gamma}_N(\tau_X)-\gamma(\tau_X))$ is $O_p(1)$. Consistency of the first step estimator, $\hat{\gamma}_N(\tau_X)$, is necessary for consistency in the second step, $\hat{\hat{\beta}}_N(\tau_Y)$, a result that is familiar in the generated regressors literature. However, inference based on conventional standard errors, say assuming homoskedasticity, of $\tau_Y(1-\tau_Y)(f_u(F_u^{-1}(\tau_Y))Q_N)^{-1}$ will be incorrect, as the presence of the generated regressor will introduce an additional term in the linear representation which must be taken into account. For instance, assuming that the errors u_n and w_n are *i.i.d.* and using the asymptotic representation for $\sqrt{N}(\hat{\gamma}_N(\tau_X)-\gamma(\tau_X))$, the covariance matrix is

$$Var(\hat{\hat{\beta}}_N(\tau_Y)) = \frac{\tau_Y(1-\tau_Y)}{f_u(Q_u(\tau_Y))}Q_N^{-1} + \frac{\tau_X(1-\tau_X)}{f_w(Q_w(\tau_X))}Q_N^{-1}Q_{N,2}Q_{N,1}^{-1}Q_{N,2}'Q_N^{-1}$$

where $Q_{N,2} = N^{-}\sum_{n=1}^{N} h_\beta(Q_{X,n,\tau},\beta(\tau))h_X(Q_{X,n,\tau},\beta(\tau))Z_n'$ is a $p\times k$ matrix that converges to Q_2 and Q_N and $Q_{N,1}$ are already defined. Hence, failure to take into account of the additional term in the variance will cause of the null hypothesis to be rejected more often than it should.

Instead of correcting the standard errors, we may include the correction term in the regression model as discussed. This is a better alternative, as the estimator will be more efficient by doing so. The linear representation under the correction model is:

Proposition 2. *Suppose a second step correction model is used and let $\hat{\beta}_N(\tau_Y)$ be the point estimator of $\beta(\tau_Y)$. Under A1-A7,*

$$\begin{aligned}&\sqrt{N}(\hat{\beta}_N(\tau_Y)-\beta(\tau_Y))\\ &= -Q_N^{-1}\frac{1}{\sqrt{N}}\sum_{n=1}^{N} h_\beta(Q_X(\tau_X|Z_n),\beta(\tau_Y))\psi_{\tau_Y}(\tilde{u}_n)+o_p(1)\end{aligned} \quad (10)$$

Since the additional term in (10) disappears by adding the correction term, the asymptotic covariance is smaller and hence the estimator $\hat{\beta}_N(\tau_Y)$ is more efficient than $\hat{\hat{\beta}}_N(\tau_Y)$. From

(9), we can understand intuitively why adding the correction term will lead to the linear representation in (10). Comparing (9) and (10), only (9) contains an additional term related to $\hat{\gamma}_N(\tau_X) - \gamma(\tau_X)$. Hence, by including the correction term derived from the first order Taylor expansion of ω_n, we control for $\hat{Q}_X(\tau_X|Z_n) - Q_X(\tau_X|Z_n) = O_p(\|\hat{\gamma}_N(\tau_X) - \gamma(\tau_X)\|)$. The neglected higher order terms are $o_p(\|\hat{\gamma}_N(\tau_X) - \gamma(\tau_X)\|)$, which are then absorbed in the $o_p(1)$ term in the linear representation. Hence, a first order Taylor expansion suffices to ensure that conventional standard errors may be used for inference.

3.1 Computation of the Standard Errors

In the empirical application discussed in Section 4, we will estimate correlations using the correction model. Hence, we will focus on constructing standard errors from the linear representation shown in (10). First, consider the case where u is homoskedastic. Then, the asymptotic variance of $\hat{\beta}_N(\tau_Y) - \beta(\tau_Y)$ from the correction model simplifies to

$$Var(\hat{\beta}_N(\tau_Y) - \beta(\tau_Y)) = \frac{\tau_Y(1-\tau_Y)}{f_u(Q_u(\tau_Y))} \frac{1}{N} \sum_{n=1}^{N} h_\beta(Q_X(\tau_X|Z_n), \beta(\tau_Y)) h_\beta(Q_X(\tau_X|Z_n), \beta(\tau_Y))' \tag{11}$$

To construct the standard errors, we need to estimate the quantile density function $s(\tau_Y) := 1/f_u(Q_u(\tau_Y))$, which is typical for inference involving sample quantiles and regression quantiles. There exists a large literature addressing this issue (e.g. Bloch and Gastwirth, 1968; Bofinger, 1975; Hall and Sheather, 1986; Goh and Knight, 2007). Typically, the starting point of estimating the quantile density comes from the classical method of Siddiqui (1960), which is based on the insight that the density of the quantile can be written as

$$s(\tau) = \lim_{k\to\infty} \frac{u_{\tau+b_k} - u_{\tau-b_k}}{2b_k}$$

where b_k is a bandwidth that approaches zero as k becomes arbitrarily large. In the sample analog, we may estimate $s(\tau)$ nonparametrically using

$$\hat{s}(\tau) = \frac{N^{-1}\sum_{n=1}^{N}(u_{n,\tau+b_k} - u_{n,\tau-b_k})}{2b_k}$$

where $u_{n,\tau}$ is the τ^{th} sample quantile of u. Extending this idea by replacing u_{n,τ_Y+b_k} and u_{n,τ_Y-b_k} with $\hat{u}_{n,\tau_Y+b_k}$ and $\hat{u}_{n,\tau_Y-b_k}$, the residuals from the τ_Y+b_k and τ_Y-b_k quantile regressions, the quantile density estimator becomes

$$\hat{s}_{u,N}(\tau_Y) = \frac{N^{-1}\sum_{n=1}^{N}\left(\breve{h}(\hat{Q}_X(\tau_X|Z_n), \hat{\beta}(\tau_Y+b_k)) - \breve{h}(\hat{Q}_X(\tau_X|Z_n), \hat{\beta}(\tau_Y-b_k))\right)}{2b_k} \tag{12}$$

recalling that $\breve{h}$ is the sum of h and the correction term. When constructing the quantile density estimate, $\hat{s}_{u,N}(\tau)$, one has to worry that this quantity may be potentially negative. While regression quantiles are not usually monotonic in τ, they are monotonic in τ at the centroid of the design matrix (Koenker, 2005). Extending this idea, the proposed estimator (12) ensures that $\hat{s}_{u,N}(\tau_Y)$ is non-negative, as it can be easily shown that the property of monotonicity at the centroid of the design matrix extends to the nonlinear model as well.

Like kernel density estimation, the estimator $\hat{s}_{u,N}(\tau_Y)$ suffers from finite sample bias, resulting from the fact that $\hat{s}_{u,N}(\tau_Y)$ is a biased estimator of $s_{u,N}(\tau_Y)$ up to the first order in a finite sample. This can be seen by considering a general nonlinear quantile regression model, taking a first order expansion of $h(\hat{Q}_X(\tau_X|Z_n), \hat{\beta}(\tau_Y))$ around $\beta(\tau_Y)$ and $Q_X(\tau_X|Z_n)$ and using the linear representation of $\hat{\beta}_N(\tau_Y) - \beta(\tau_Y)$ and $\hat{\gamma}_N(\tau_X) - \gamma(\tau_X)$ to obtain

$$\hat{s}_{u,N}(\tau_Y) = s_{u,N}(\tau_Y) + \frac{1}{2b_k\sqrt{N}}O_p(1)$$

where $s_{u,N}(\tau_Y)$ is evaluated at $\beta(\tau_Y)$ and $Q_X(\tau_X|Z_n)$, while the $O_p(1)$ term comes from applying the Central Limit Theorem. This implies that $|\hat{s}_{u,N}(\tau_Y) - s_{u,N}(\tau)| = O_p\left(\frac{1}{b_k\sqrt{N}}\right)$. Hence to ensure that $\hat{s}_{u,N}(\tau)$ is consistent, the bandwidth must converge at a slower rate than $N^{-1/2}$. This motivates our choice of using the Hall and Sheather (1988) bandwidth over the Bofinger (1975) bandwidth. Having an order of $N^{-1/3}$, the Hall and Sheather bandwidth will lead to faster convergence of the quantile density estimator than if the Bofinger bandwidth, with an order of $N^{-2/5}$, is adopted. For the moment, we will simplify notation by defining $Q_{X,n,\tau} := Q_X(\tau|Z_n)$ so that the dependence on Z_n is suppressed. In practice, we compute the bandwidth, $b_k = \frac{\hat{m}}{N}$, via three equations:

$$\hat{m} = (1.5\tilde{s}_N/|V_N|)^{1/3}(z_{\alpha/2})^{2/3}N^{2/3}$$

$$\tilde{s}_N = \frac{\bar{h}(\hat{Q}_{X,n,\tau_X}, \hat{\beta}(\tau_Y + b_{1,k})) - \bar{h}(\hat{Q}_{X,n,\tau_X}, \hat{\beta}(\tau_Y - b_{1,k}))}{2b_{1,k}}$$

$$\begin{aligned} V_N =& \frac{\bar{h}(\hat{Q}_{X,n,\tau_X}, \hat{\beta}(\tau_Y + 2b_{2,k})) - \bar{h}(\hat{Q}_{X,n,\tau_X}, \hat{\beta}(\tau_Y - 2b_{2,k})}{2b_{2,k}^3} \\ &+ \frac{2\bar{h}(\hat{Q}_{X,n,\tau_X}, \hat{\beta}(\tau_Y - b_{2,k})) - 2\bar{h}(\hat{Q}_{X,n,\tau_X}, \hat{\beta}(\tau_Y + b_{2,n}))}{2b_{2,k}^3} \end{aligned}$$

where $\bar{h}(\hat{Q}_X(\tau_X|Z_n), \hat{\beta}(\tau_Y + b_{1,k})) = N^{-1}\sum_{n=1}^{N} \breve{h}(\hat{Q}_X(\tau_X|Z_n), \hat{\beta}(\tau_Y + b_k))$. These bandwidths $b_{1,k} = \min(\tau_Y, 1-\tau_Y)N^{-0.2}$ and $b_{2,k} = 0.5\min(\tau_Y, 1-\tau_Y)N^{-1/9}$ modify Hall and Sheather's recommendation of $b_{1,k} = 0.5N^{-0.2}$ and $b_{2,k} = 0.25N^{-1/9}$ for the sample median. It is clear that our modified bandwidths are the same as those in Hall and Sheather for median regression. This modification is necessary because at the tails, for instance at $\tau = 0.05$, the bandwidths recommended by Hall and Sheather may result in negative values of $\tau_Y - b_{1,k}$ or $\tau_Y - b_{2,k}$ which are not admissible.

Under the assumption that u_n is heteroskedastic, we must estimate the density for each $\hat{u}_n$. Using the Hall and Sheather bandwidth, we compute

$$\hat{f}_{\tilde{u},n} = \frac{2b_k}{h(\hat{Q}_X(\tau_X|Z_n), \hat{\beta}_N(\tau_Y + b_k)) - h(\hat{Q}_X(\tau_X|Z_n), \hat{\beta}_N(\tau_Y - b_k)))} \tag{13}$$

This method is similar to the Hendricks-Koenker sandwich method discussed in Koenker (2005, p.80). As Koenker pointed out, a potential problem in using the sandwich method is that $h(\cdot, \beta(\tau))$ may not be monotonic except at the centroid of the design matrix. This implies that for some observations, the denominator in (13) may be negative. In the actual implementation, the Hendricks-Koenker method employs

$$\hat{f}^{+}_{\tilde{u},n} = \max\left\{0, \frac{2b_k}{h(\hat{Q}_X(\tau_X|Z_n), \hat{\beta}_N(\tau_Y + b_k)) - h(\hat{Q}_X(\tau_X|Z_n), \hat{\beta}_N(\tau_Y - b_k))) - e}\right\} \tag{14}$$

where e is a small value to prevent division by zero.

3.2 Estimation

We now describe the actual implementation that is used in Section 4. The first-step objective function is

$$\sum_{n=1}^{N} \rho_{\tau_X}(X_n - Z_n'\gamma) \tag{15}$$

where $\rho_\tau(w) = (\tau - \mathbb{I}(w < 0))w$ is a "check" function proposed by Koenker and Bassett (1978). From the first step regression, we want to form Δ_N expressed in (5), which will be used in the second step. To do so, residuals from the first step regression $\hat{w}_n$ are used to define

$$\hat{\Delta}_N = \hat{s}_w(\tau_X)(N^{-1}\sum_{n=1}^{N} Z_n Z_n')^{-1} N^{-1} \sum_{n=1}^{N} Z_n \psi_\tau(\hat{w}_n) \tag{16}$$

where $\hat{s}_w(\tau_X)$ is the estimated quantile density function computed as discussed in Section 3.1. Note that (16) is based on the assumption that w_n is homoskedastic. Alternatively, one may estimate an heteroskedastic version yielding

$$\hat{\Delta}_N = (N^{-1}\sum_{n=1}^{N} \hat{f}_{w,n} Z_n Z_n')^{-1} N^{-1} \sum_{n=1}^{N} Z_n \psi_\tau(\hat{w}_n) \tag{17}$$

where $\hat{f}_{w,n}$ is the estimated density of $f_{w,n}$. For the actual computation, we obtain $\hat{\Delta}_N$ using the assumption that w is homoskedastic, as there are cases where the estimated densities are very large when the denominator in (13) is close to zero. Then, $\hat{\Delta}_N$ is used in the correction model, and with it, we minimize

$$\sum_{n=1}^{N} \rho_{\tau_Y}(Y_n - \breve{h}(\hat{Q}_X(\tau_X|Z_n), \beta)) \tag{18}$$

where $\breve{h}$ is the original regression function plus the correction term.[6]

4 Empirical Application

In this section, we estimate the correlations between the quantile of U.S. MSCI returns and the quantile of MSCI returns of Canada, France, Germany, U.K., Australia, Japan, Hong Kong and Singapore. These countries reflect the major stock markets in North America, Europe and Australasia. We focus on eight pairwise relationships: the correlation between U.S. returns and the returns to each of the other eight countries. To obtain a correlation measure, we employ a copula-based model due to Bouyè and Salmon (2003) to derive the nonlinear regression function h.

To get a sense of how copulas work, suppose we are interested in investigating how X and Y are dependent. Due to Sklar's theorem (see Bouyè and Salmon, 2003), there exists a unique copula function C with copula parameter ρ for every joint distribution $F_{X,Y}$ over X and Y that satisfies

$$F_{X,Y}(X, Y) = C(F_X(X), F_Y(Y); \rho)$$

Hence, every joint distribution can be expressed as a copula function over the marginal distributions and vice versa. Therefore, by using Sklar's Theorem, the task of modeling even the most complex relationships can be simplified by specifying the copula function and the marginal distributions separately. The main advantage of using the copula function is that the copula parameter ρ maps into a measure of rank correlation known as Kendall's Tau.[7] For the Gaussian or Student-t copula, the copula parameter is simply the correlation

[6]While we have estimated an unweighted version of the objective function, one may use the weighted quantile regression approach described by Ma and Koenker (2006). The optimal weights are the estimated densities, meaning that we first weight the summands in the objective function by $\hat{f}_{\tilde{u},n}$ obtained from estimating a unweighted version.

[7]To illustrate the concept of a rank correlation, consider N observation pairs of X and Y. From these observations, one may construct $\binom{n}{2}$ rankings of each observation pairs. The Kendall S calculates the difference between the number of concordant and discordant pairs, denoted by C and D respectively. Then, Kendall's Tau computes the rank correlation through the measure $S/\binom{n}{2}$.

coefficient itself. To derive the dependence function, we observe that

$$F_{Y|X}(Y|X_n) = C_1(F_X(X), F_Y(Y); \rho).$$

where $F_{Y|X}(Y|X_n)$ is the marginal distribution of Y conditioned on $X = X_n$ and $C_1(u, v)$ is the partial derivative of the copula with respect to the first argument. By inverting C_1 with respect to the second argument, we obtain the required regression framework as

$$F_Y(Y) = C_1^{-1}(F_X(X_n); \rho) \tag{19}$$

Equation (19) provides the basis for writing down the nonlinear quantile regression model that we represent as h. We consider a parametric specification for both the copula function and the marginal distributions, while Chen and Fan (2006) considered combining a parametric copula with nonparametric marginal distributions. To analyze the dependence between quantiles, we consider a second step model based on (19) as

$$Q_{F_Y}(\tau_Y|Q_X(\tau_X|Z_n)) = h(Q_X(\tau_X|Z_n), \rho(\tau_Y)) \tag{20}$$

where h is derived from the right-hand-side term of (19). Note that by inverting F_Y in (20), we may also estimate ρ from an equation that expresses the left-hand-side variable as $Q_Y(\tau_Y|Q_X(\tau_X|Z_n))$ instead of $Q_{F_Y}(\tau_Y|Q_X(\tau_X|Z_n))$. However, this distinction is not crucial since F_Y^{-1} is a monotone transformation. Hence, we will proceed with estimating ρ based on (20). As explained earlier, we first estimate the auxiliary regression of

$$Q_X(\tau_X|Z_n) = Z_n'\gamma(\tau_X) \tag{21}$$

and obtain the fitted values $\hat{Q}_X(\tau_X|Z_n)$ for the estimation of (20).

For our empirical application, we use monthly time series that spans from March 1971 to December 2008. The dataset is obtained from Datastream and the returns are expressed in the U.S. currency. All quantile regression estimations reported here are implemented in MATLAB using the interior point algorithm of Koenker and Park (1996). The stock returns series are obtained from log-differencing the MSCI index.

Our sample contains data for the U.S. and eight other countries that we call "foreign" in this paper. For presentation purpose, we group the foreign countries into two: 1) Canada, France, Germany and the U.K. and 2) Australia, Hong Kong, Japan, Singapore. The U.S. returns will be used as the anchor, meaning that histograms and correlations will be computed for the pair of countries involving the U.S. and another foreign country.

To get a sense of how the returns series are distributed, bivariate histograms for the U.S. and foreign returns are plotted. The histogram shows the number of observations such that values of the bivariate returns series of interest fall into bins whose boundaries are determined by the returns quantiles.[8]

Figure 1 plots the histograms for Canada and the European countries while Figure 2 does the same for Australia and Asian countries. All the histograms show that most of the data are concentrated around the main diagonal, meaning it is less likely to observe the simultaneous realization of U.S. and foreign returns belonging to the opposite tails of the returns distributions. In addition, the modes of the histograms are located at the extreme left and right tails. This implies that one is likely to see a large drop in the foreign stock return when the same is also observed for the U.S. stock return. The converse is true, although the observations are more highly concentrated in the left than in the right tails.

The histograms tend be less dispersed for Canada and Europe as opposed to those for Australia and Asia. Moreover, the frequency in the extreme tails tend to be larger for the first than for the second group of countries. This implies that the stock markets of Canada and Europe tend to move more closely to the U.S. market in extreme events than the stock markets of Australia and Asia do.

To estimate the correlation between the quantiles of returns, a preliminary estimation

[8]Specifically, the histogram plots the number of observations of X and Y that jointly satisfy $Q_X(\tau_i) < X \leq Q_X(\tau_i)$ and $Q_Y(\tau_j) < Y \leq Q_Y(\tau_j)$, for $\tau_i, \tau_j = 0, 0.1, ..., 1$. The axis labels indicates the upper bound of the interval. For $\tau = 0$, we set $Q_X(0) = \min(X) - 1$ and $Q_Y(0) = \min(Y) - 1$ for computational purposes.

step is needed to obtain the fitted conditional quantiles of the U.S. returns. The fitted quantile will then be used as a regressor in the second estimation step whose model will be specified later. Letting X represent the U.S. returns and Y represent the foreign counterpart, the auxiliary model employed by the preliminary estimation step is specified as

$$\begin{aligned} X_n =& \gamma_0 + \sum_{i=1}^{12} \gamma_{X,i} X_{n-i} + \sum_{i=1}^{12} \gamma_{P,i} USProd_{n-i} + \sum_{i=1}^{12} \gamma_{M2,i} USM2_{n-i} + \sum_{i=1}^{12} \gamma_{I,i} USPPI_{n-i} \\ &+ \sum_{i=1}^{12} \gamma_{S,i} Spread_{n-i} + \sum_{i=1}^{12} \gamma_{F,i} Fed_{n-i} + \sum_{i=1}^{12} \gamma_{Gi} Gold_{n-i} + \sum_{i=1}^{12} \gamma_{C,i} Copper_{n-i} \\ &+ \sum_{i=1}^{12} \gamma_{H,i} House_{n-i} + w_n \end{aligned} \tag{22}$$

where $USProd$ is the growth in U.S. industrial production index, USM2 is the growth in U.S. M2 money, $USPPI$ is U.S. producer price index inflation, *Spread* is the difference between the yields on 10-year and 3-month treasury bills, *Gold* is gold's monthly returns, *Copper* is copper's monthly returns and $House$ is the number of housing starts denominated in millions. The growth rates and returns are computed by log-differencing the variables. We find that for the second step regression to work well, in the informal sense that the estimated correlation in the second step is less attenuated towards zero, the first step design matrix should include twelve lags of each exogenous variable. To avoid mistakenly employing regressors that may be endogenous, we do not include contemporaneous variables in (22).

For the second step regression, the regression function is derived based on the assumption that the copula is either Gaussian or Student-t. These copula models are chosen for their simplicity although other copulas may also be considered. Detailed discussions on the types of copulas suitable for this quantile-based framework can be found in Bouyè and Salmon (2003). Based on the Gaussian copula, the second step regression model can be expressed as

$$\Phi(Y_n) = \Phi\left(\Upsilon_\Phi\right) - \rho(u_n)\phi\left(\Upsilon_\Phi\right) Z'_{n-1}\Delta_N$$

where

$$\Upsilon_{\Phi} = \rho(u_n)\hat{Q}_X(\tau_X|Z_{n-1}) + \sqrt{1-\rho(u_n)^2}\Phi^{-1}(\tau_Y),$$

Φ is the standard normal cumulative distribution function and ϕ is the standard normal density function. The second term is the correction term that approaches zero asymptotically as Δ_N is $o_p(1)$. When the Student-t copula is used, the second step regression model becomes

$$T_\nu(Y_n) = T_\nu\left(\Upsilon_{T_\nu}\right) - \rho(u_n)t_\nu\left(\Upsilon_{T_\nu}\right)Z'_{n-1}\Delta_N \tag{23}$$

where

$$\Upsilon_{T_\nu} = \rho(u_n)\hat{Q}_X(\tau_X|Z_{n-1}) + \Big((\nu+\hat{Q}_X(\tau_X|Z_{n-1})^2)(\nu+1)^{-1}(1-\rho(u_n)^2)\Big)^{1/2}T_\nu^{-1}(\tau_Y),$$

T_ν and t_ν represent the Student-t cumulative distribution and density function with ν degrees of freedom. As expected, the Student-t model reduces to the Gaussian model as ν goes to infinity. In the actual estimation, we will only report the estimation results based on the Student-t model with ten degrees of freedom as choosing other degrees of freedom can be shown to produce similar outcomes. The presence of τ_Y, seen in (23) and (24), comes from the partial derivative of the copula function with respect the marginal distribution of X and this is typical in copula quantile regressions. Note that simplification is achieved by combining a Gaussian copula with standard normal marginal distributions or by combining a Student-t copula that shares the same degrees of freedom with the univariate t-distributions. This is solely for convenience and if necessary, one may use marginal distributions belonging to a different family as the copula function.

As mentioned in the introduction, Longin and Solnik (2001) and Ang and Chen (2002) are among the groundbreaking works that found correlation asymmetries in stock returns. Longin and Solnik focused on the correlation of international markets while Ang and Chen examined the correlation between U.S. portfolio returns and U.S. aggregate returns. While both demonstrated that correlation generally rises in bear markets, the results should be interpreted with some caution. This is because the exceedance-based method of the papers estimates the correlation between the tails of the stock returns, whose distributions are approximated by the generalized Pareto distribution. This approximation becomes more imprecise when correlations are calculated for observations located nearer the center of the returns' distributions and further away from the tails.

The Longin and Solnik exceedance correlation involving the U.S. and Canada, France, Germany and the U.K. is plotted in Figure 3. The same involving the U.S. and Australia, Hong Kong, Japan and Germany is plotted in Figure 4. For $\tau \leq 0.5$, the figures report the correlation when returns are jointly *less than or equal* to their respective $100\tau^{\text{th}}$ percentiles. For instance, the correlation at $\tau = 0.2$ is the correlation for returns when both markets are less than or equal to their respective 20th percentiles. For $\tau \geq 0.5$, the reported correlations are those when returns are jointly *greater than or equal* to their respective $100\tau^{\text{th}}$ percentiles. This implies that two correlation points are plotted at τ equal 0.5, which explains the discontinuity observed at the median.

While bearing in mind that exceedance correlations involving the extremes are likely to have a smaller bias, Figures 3 and 4 show that these correlations typically display an asymmetric shape. There is a discontinuity at $\tau = 0.5$, reflecting the fact that exceedance correlations are correlations between the tails where in our case, the left and right tails are defined for observations up to the median. The U.S.-Canada correlation appears to be the least asymmetric compared to the rest of the country pairs. The U.S.-Australia correlation has the sharpest drop in correlation in the right tails. The figures, however, show that the correlation generally does not increase monotonically from the median to the 10th percentile, suggesting that extreme negative events do not necessarily lead to tighter comovements with the U.S. market. In fact, except for Germany and Japan, the negative semicorrelation at the median is larger than the correlation at the 10th percentile.[9] Thus, while the Longin and Solnik method produces correlation with an asymmetric pattern, it

[9]The negative semicorrelation at the median refers to the correlation when returns that are less than or equal to the median.

also broadly suggests that extreme bear events do not increase the correlation between stock markets. This observation, however, does not conform with the typical beliefs of researchers and market observers alike.

Using our quantile dependence framework that measures the correlation between quantiles rather than between tail observations, Figures 5 plots the correlation of U.S. returns quantiles with returns quantiles of Canada, France, Germany and U.K. Figure 6 plots the same when the foreign countries are Australia, Hong Kong, Japan and Singapore. The correlations are plotted against the U.S. returns quantiles and the foreign returns quantiles, resulting in a three-dimensional surface. For future reference, the main diagonal of the surface refers to the case where the same τ indexes the returns quantiles for both U.S. and the foreign country. Correlations in Figures 5 and 6 are obtained from the Gaussian copula model.

The correlations exhibit several broadly similar characteristics. First, the typical correlation surface has a saddle shape exhibiting higher elevations along the main diagonal. Correlations located nearer to the extreme ends of the main diagonal display even larger increases, implying correlations between markets that are jointly bullish or bearish should rise relative to correlations between markets when returns are typically seen. Concerning the fact that correlations increase when both markets are bullish, this has not been found previously although it is somewhat appealing, based on cursory observation, that foreign markets should typically be bullish when the same happens to the U.S. market also. When moving away from the main diagonal, the surface tends to decline and this implies that markets experiencing increasingly dissimilar environments should become less correlated. For example, one would expect that correlation of the markets should fall when one market is bullish while the other is bearish.

Despite the similarities, asymmetric correlation in the conventional sense still exists. This can be clearly seen by plotting the correlation along the main diagonal, as Figure 7 does so for Canada, France, Germany and U.K. and Figure 8 for Australia, Hong Kong, Japan and U.K. Given the saddle shape of the correlation surfaces, it is not surprising that the main-diagonal correlation has an inverted-U or cup shape. The asymmetry manifests when we observe in all cases that the correlation between the 10th percentiles is larger than that between the 90th percentiles. The difference between the 10th and the 90th percentile correlation are among the largest for Hong Kong (0.295), Australia (0.237), the U.K. (0.220) and Germany (0.221). The difference is the smallest for Canada (0.054), which is also the case when the difference is computed using the exceedance correlations estimated for Figure 3.

Given the cup shape of the main diagonal correlation, we attempt to measure the depth of the "cup" by defining

$$Depth = |\rho_{0.1} - \rho_{0.5}| + |\rho_{0.9} - \rho_{0.5}| \tag{24}$$

where ρ_τ is the $100\tau^{\text{th}}$ percentile correlation between the U.S. and the foreign country. *Depth* measures the sensitivity of the correlation to extreme events, where large magnitudes of *Depth* reflect larger increases in the correlation when markets are bearish or bullish. The results for the eight country pairs are reported in Table 1, which shows that *Depth* is among the largest for Singapore (0.653), Hong Kong (0.637) and the U.K. (0.594) and is the smallest for Canada (0.278). The fact that *Depth* is large for Singapore and Hong Kong, being small economies that are also among the most open, suggests that openness and size of the countries could account for the sensitivity of correlations to extreme events.

The results reported so far are estimated from the Gaussian copula model based on (23). Here, we conduct the same analysis using the Student-t copula model based on (24). Figure 9 plots the correlation surface for Canada, France, Germany and the U.K. and Figure 10 does the same for Australia, Hong Kong, Japan and Singapore.

We can see that correlations behave very similarly when comparing Figures 5 and 9 for Canada and Europe and between Figures 6 and 10 for Australia and Asia, suggesting that the difference from using the Gaussian versus Student-t model is small. That said, correlations estimated from the Student-t model tend to have larger values and this is especially the case in the tails.

This can be seen from the main-diagonal plots of Figure 11 for Canada and Europe and Figure 12 for Australia and Asia. We can see from the figure that the main-diagonal correlations exhibit very similar patterns as those in Figures 7 and 8 based on the Gaussian model. Clearly, correlations at the tails are larger using the Student-t model, with the increase in correlations between the 10th percentiles ranging from 0.03616 for the U.K. to 0.0977 for Canada. For the 90th percentiles, the increase in correlation ranges from 0.0366 for Germany to 0.1102 for Hong Kong. However, the difference between the medians estimated from both copula functions remains very close to zero.

4.1 Extensions

Before we conclude, it is useful to consider two extensions of the copula-based model that may be useful for future empirical work.

Extension 1: Relaxing the Marginal Distribution Assumptions

In the previous section, the marginal distributions are restricted to be the univariate counterparts of the copula. For instance, we combine the Gaussian copula with standard normal marginal distributions, or the Student-t copula with univariate t-distributions all having the same degrees of freedom. Here, we relax the restriction on the marginal distributions. Using the Gaussian copula as illustration sake, the method proceeds from Chen et al. (2008) by modeling the joint distribution as

$$F_{X,Y}(X,Y) = \Phi_\rho(\Phi^{-1}(F_X(X)), \Phi^{-1}(F_Y(Y)))$$

where Φ_ρ is the bivariate Gaussian distribution and F_X and F_Y are any continuously differentiable distribution functions. The second step model becomes

$$
\begin{aligned}
&F_Y(Y_n)\\
=&\Phi\left(\rho(u_n)\Phi^{-1}(F_X(\hat{Q}_X(\tau_X|Z_{n-1}))) + \sqrt{1-\rho(u_n)^2}\Phi^{-1}(\tau_Y)\right)\\
&-\rho(u_n)\phi(\rho(u_n)\Phi^{-1}(F_X(\hat{Q}_X(\tau_X|Z_{n-1}))) + \sqrt{1-\rho(u_n)^2}\Phi^{-1}(\tau_Y))\\
&\times\frac{f_X(\hat{Q}_X(\tau_X|Z_{n-1}))}{\phi(\Phi^{-1}(F_X(\hat{Q}_X(\tau_X|Z_{n-1})))}Z'_{n-1}\Delta_N
\end{aligned}
$$

so that the marginal distributions other than the standard normal may be combined with the Gaussian copula.

4.2 Extension 2: Time-Varying Correlation

Time-varying correlations may also be estimated as an extension. This is especially convenient for the copula-based model, since the copula parameter may be modeled as a function of an information set as

$$\rho_n(Z_{1,n-1}) = \Lambda(Z'_{1,n-1}\alpha)$$

where $Z_{1,n-1}$ represents the information set and $\Lambda(x) = (\exp(2x)-1)/(\exp(2x)+1)$ is the inverse Fisher transformation that maps a real number into the $[-1,1]$ interval. To identify the conditional correlation in the second step regression, $Z_{1,n-1}$ must be a strict subset of Z_{n-1}, the information set in the first step, so that there must at least be one regressor in Z_{n-1} that is excluded from $Z_{1,n-1}$ which follows from our discussion in Section 2.1. Then, estimating time-varying correlation becomes a problem of estimating α and this can be carried out in a straightforward manner using the methods discussed earlier.

5 Conclusion

Quantile regression is a useful tool for investigating the regressors' influence on the quantiles of the dependent variable. This paper discusses two contributions. First, we have proposed to model the statistical relationships of quantiles using a generated regressors framework. In doing so, we have addressed the issue of generated regressors and examined their asymptotic implications in a nonlinear quantile regressions. Second, we have constructed correlation surfaces that show how correlations between quantiles of returns to the stock markets behave. These surfaces reveal that the tails are typically more strongly dependent than is true at the center of the distributions. In addition, our copula methodology flexibly allows the quantiles of the stock returns to be generated by different joint distributions belonging to the same family. Our estimation results provide evidence that the documented asymmetric correlation of international stock returns is related to changes in the correlation coefficient across different states of the economy, thus supporting the idea that correlation breakdowns have taken place.

There are several ways to extend the current paper. Theoretically, we may further relax the assumption about parametric marginal distributions by investigating a semiparametric model with a parametric copula and nonparametric marginals as in Chen and Fan (2006). The issue of bootstrapping standard errors in quantile regressions with generated regressors is also useful from the applied perspective. Empirically, we have considered the dependence of nine international stock markets, hence it would be interesting to extend the study to include other international stock markets. In addition, further research on examining the effects of monetary policy and the business cycle on stock market correlations can be carried out with our framework.

References

Amemiya, T. (1982) Two stage least absolute deviations estimators, *Econometrica*, 50(3), pp. 689-711.

Ang, A. and Chen, J. (2002) Asymmetric correlations of equity portfolios, *Journal of Financial Economics*, 63(3), pp. 443-494.

Babu, G. J. and Singh, K. (1978) On deviations between empirical and quantile processes for mixing random variables, *Journal of Multivariate Analysis*, 8(4), pp. 532-549.

Ball, L. and Cecchetti, S.G. (1991) Inflation and uncertainty at long and short horizons, *Brookings Papers on Economic Activity*, 1, pp. 215-254.

Bahadur, R. R. (1966) A note on quantiles in large samples, *Annals of Mathematical Statistics*, 37(3), pp. 577-80.

Bloch, D. A. and Gastwirth, J. L. (1968) On a simple estimate of the reciprocal of the density function, *Annals of Mathematical Statistics*, 39, pp. 1803-1805.

Bofinger, E. (1975) Estimation of a density function using order statistics, *Australian Journal of Statistics*, 17, pp. 1-7.

Bosq, D. (1998) *Nonparametric statistics for stochastic processes*, Springer, New York (1998).

Bouyè, E. and Salmon, M. (2003) Dynamic copula quantile regressions and tail area dynamic dependence in forex markets, Mimeo.

Boyer, B. H., Gibson, M. S. and Loretan, M. (1999) Pitfalls in testing for changes in correlation, Mimeo.

Chen, X. and Fan, Y. (2006) Estimation of copula-based semiparametric time series models, *Journal of Econometrics*, 130(2), pp. 307-335.

Chen, X., Koenker, R. and Xiao, Z. (2008) Copula-based nonlinear quantile regression autoregression, *Cowles Foundation Discussion Paper* No. 1679.

Doukhan, P. (1994) *Mixing: Properties and examples*, Springer, New York.

Goh, C. and Knight, K. (2007) Nonstandard quantile-regression inference, Mimeo.

Guo, H. (2004) Stock prices, firm size, and changes in the federal funds rate target, *The Quarterly Review of Economic and Finance*, 44(4), pp. 487-507.

Hall, P. and Sheather, S. (1987) On the distribution of a studentized quantile, *Journal of Royal Statistical Society B.*, 50(3), pp. 381-391.

He, X. and Shao, Q.M. (2000) On parameters of increasing dimensions, *Journal of Multivariate Analysis*, 73(1), pp. 120-135.

Hu, L. (2006) Dependence patterns across financial markets: a mixed copula approach, *Applied Financial Economics*, 16(10), pp. 717-729.

Jurečková, J. and Procházka, B. (1994) Regression quantiles and trimmed least squares estimators in nonlinear regression models, *Journal of Nonparametric Statistics*, 3(3-4), pp.

201-222.

Kiefer, J. (1967) On Bahadur's representation of sample quantiles, *Annals of Mathematical Statistics*, 38(5), pp. 1323-1342.

Kim, J. and Finger, C. (2000) A stress test to incorporate correlation breakdown, *Journal of Risk*, 2(3), pp. 6-19.

Koenker, R. and Bassett, G. (1978) Regression quantiles, *Econometrica*, 46(1), pp. 33-50.

Koenker, R. and Park, B.J. (1996) An interior point algorithm for nonlinear quantile regression, *Journal of Econometrics*, 71(1-2), pp. 265-283.

Koenker, R. and Zhao, Q. (1994) L-estimation for linear heteroscedastic models, *Journal of Nonparametric Statistics*, 3(3-4), pp. 223-235.

Koenker, R. (2005) *Quantile Regression*, Econometric Society Monographs 38, Cambridge University Press.

Longin, F. and Solnik, B. (1995) Is the correlation in international equity returns constant: 1960-1990?, *Journal of International Money and Finance*, 14(1), pp. 3-26.

Longin, F. and Solnik, B. (2001) Extreme correlations in international equity markets, *Journal of Finance*, 56(2), pp. 649-676.

Ma, L. and Koenker, R. (2006) Quantile regression methods for recursive structural equation models, *Journal of Econometrics*, 134(2), pp. 471-506.

McKenzie, C. and McAleer, M. (1997) On efficient estimation and correct inference in models with generated regressors: a general approach, *Japanese Economic Review*, 48(4), pp. 368-389.

Murphy, K. M. and Topel, R. H. (1986) Estimation and inference in two-step econometric models, *Journal of Business and Economic Statistics*, 3(4), pp. 370-379.

Oberhofer, W. and Haupt, H. (2006) Asymptotic theory for nonlinear quantile regression under weak dependence, Mimeo.

Pagan, A. (1984) Econometric issues in the analysis of regressions with generated regressors, *International Economic Review*, 25(1), pp. 221-247.

Powell, J. (1983) The asymptotic normality of two-stage least absolute deviations estimators, *Econometrica*, 51(5), pp. 1569-1575.

Sen, P. K. (1972) On Bahadur representation of sample quantile for sequences of ϕ-mixing random variables, *Journal of Multivariate Analysis*, 2(1), pp. 77-95.

Siddiqui, M. M. (1960) Distribution of quantiles in samples from a bivariate population, *Journal of Research of the National Bureau Standards B*, 64, pp. 145-150.

Sun, S. (2006) The Bahadur representation for sample quantiles under weak dependence, *Statistics and Probability Letters*, 76(12), pp. 1238-1244.

Weiss, A. A. (1991) Estimating nonlinear dynamic models using least absolute error estimation, *Econometric Theory*, 7(1), pp. 46-68.

Xiao, Z. and Koenker, R. (2008) Conditional quantile estimation for GARCH models, Mimeo.

Yoshihara, K. (1995) The Bahadur representation of sample quantile for sequences of strongly mixing random variables, *Statistics and Probability Letters*, 24(4), pp. 299-304.

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

Appendix

Proof of Proposition 1

Before proceeding with the proof, we lay down some definitions. To simplify notation, define $Q_{X,n,\tau} \equiv Q_X(\tau|Z_n) \equiv Z_n'\gamma(\tau)$ so that the dependence on Z_n is suppressed. Furthermore, the proof restricts $\tau_X = \tau_Y = \tau$ without loss of generality. Let $\hat{\hat{\beta}}_N$ represent the point estimator of β from the non-corrected second step regression. Also, recall that $\tilde{u}_n$ such that

$$\begin{aligned}\tilde{u}_n =& Y_n - h(Q_{X,n,\tau}, \beta(\tau)) \\ =& Y_n - h(\hat{Q}_{X,n,\tau}, \beta(\tau)) - (h(Q_{X,n,\tau}, \beta(\tau)) - h(\hat{Q}_{X,n,\tau}, \beta(\tau))),\end{aligned}$$

so that the τ^{th} conditional quantile of $\tilde{u}$ is normalized to zero. Using this normalization, let the first order condition be represented by a $p \times 1$ vector

$$\begin{aligned}& W_N(\hat{Q}_{X,\tau}, \hat{\hat{\beta}}_N(\tau)) \\ \equiv& \frac{1}{N}\sum_{n=1}^{N} h_\beta(\hat{Q}_{X,n,\tau}, \hat{\hat{\beta}}_N(\tau)) \\ & \times \psi_\tau\left(\tilde{u}_n - (h(\hat{Q}_{X,n,\tau}, \hat{\hat{\beta}}_N(\tau)) - h(\hat{Q}_{X,n,\tau}, \beta(\tau))) - (h(\hat{Q}_{X,n,\tau}, \beta(\tau)) - h(Q_{X,n,\tau}, \beta(\tau)))\right), \end{aligned} \tag{25}$$

where $\psi_\tau = \mathbb{I}(u < 0) - \tau$ and $\mathbb{I}(\cdot)$ is an indicator function. Furthermore, define a $p \times 1$ vector

$$\begin{aligned}& \bar{W}_N(\hat{Q}_{X,\tau}, \hat{\hat{\beta}}_N(\tau)) \\ \equiv& \frac{1}{N}\sum_{n=1}^{N} h_\beta(Q_{X,n,\tau}, \beta(\tau)) \\ & \times \psi_\tau\left(\tilde{u}_n - (h(\hat{Q}_{X,n,\tau}, \hat{\hat{\beta}}_N(\tau)) - h(\hat{Q}_{X,n,\tau}, \beta(\tau))) - (h(\hat{Q}_{X,n,\tau}, \beta(\tau)) - h(Q_{X,n,\tau}, \beta(\tau)))\right). \end{aligned} \tag{26}$$

In other words, a bar over $W_N(\hat{Q}_{X,n,\tau}, \hat{\hat{\beta}}_N(\tau))$ represents the case in replacing $h_\beta(\hat{Q}_{X,n,\tau}, \hat{\hat{\beta}}_N(\tau))$ in W_N with $h_\beta(Q_{X,n,\tau}, \beta(\tau))$. In addition, given the true parameter $\beta(\tau)$, define the expectation counterpart of (25) and (26) with the expectation operator $E[\cdot]$ taken over $\tilde{u}$,

$$W(\hat{Q}_{X,\tau}, \hat{\hat{\beta}}_N(\tau)) \equiv E[W_N(\hat{Q}_{X,\tau}, \hat{\hat{\beta}}_N(\tau))]$$

and

$$\bar{W}(\hat{Q}_{X,\tau}, \hat{\hat{\beta}}_N(\tau)) \equiv E[\bar{W}_N(\hat{Q}_{X,\tau}, \hat{\hat{\beta}}_N(\tau))].$$

By definition, $\hat{\hat{\beta}}_N(\tau) = \text{argmin}(\beta)[W_N(\hat{Q}_{X,\tau},\beta)]$ and $\beta(\tau) = \text{argmin}(\beta)[W(Q_{X,\tau},\beta)]$. Except for points where $Y_n = h(\hat{Q}_{X,n,\tau},\hat{\hat{\beta}}_N(\tau))$, $\|W_N(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))\|$ may be set to zero while $\|W(Q_{X,\tau},\beta(\tau))\|$ is zero by definition, where $\|\cdot\|$ is the Euclidean norm, that is, $\|\cdot\| = (tr(A'A))^{1/2}$.[10] In addition, note that $E[\mathbb{I}(\tilde{u} < 0)] = \tau$. Using the argument from Xiao and Koenker (2008), we will establish that $\hat{\hat{\beta}}_N(\tau)$ is root-N-consistent, which crucially depends on the fact $|\hat{Q}_{X,n,\tau} - Q_{X,n,\tau}| = |Z_n'(\hat{\gamma}_N(\tau) - \gamma(\tau))| = O_p(N^{-1/2})$. Consider

$$
\begin{aligned}
&\|\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))\| \\
\leq &\|\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau)) - \bar{W}(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))\| \quad (27)\\
+&\|\bar{W}_N(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau)) - \bar{W}_N(Q_{X,\tau},\beta(\tau)) - (\bar{W}(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau)) - \bar{W}(Q_{X,\tau},\beta(\tau)))\| \quad (28)\\
+&\|\bar{W}_N(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))\| \quad (29)\\
+&\|\bar{W}_N(Q_{X,\tau},\beta(\tau))\|. \quad (30)
\end{aligned}
$$

Lemma 4 shows that (30) has a rate of $O_p(N^{-1/2})$. Next, let $\bar{W}_X$ represent a $p \times N$ matrix of derivatives of $\bar{W}$ with respect to the first argument and $\bar{W}_{XX}$ represent a $p \times N \times N$ tensor of second derivatives. For (27), express

$$
\begin{aligned}
&\|\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau)) - \bar{W}(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))\| \\
\leq &\|\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau)) - \bar{W}(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau)) - \bar{W}_X(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))(\hat{Q}_{X,\tau} - Q_{X,\tau})\| \\
+&\|\bar{W}_X(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))(\hat{Q}_{X,\tau} - Q_{X,\tau}) - \bar{W}_X(Q_{X,\tau},\beta(\tau))(\hat{Q}_{X,\tau} - Q_{X,\tau})\| \\
+&\|\bar{W}_X(Q_{X,\tau},\beta(\tau))(\hat{Q}_{X,\tau} - Q_{X,\tau})\|.
\end{aligned}
$$

Note, for instance, that since $\bar{W}_X(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau)) = \bar{W}_X(Z\hat{\gamma}_N(\tau),\hat{\hat{\beta}}_N(\tau))$, the above expression can also be obtained by Taylor expansion around $\gamma(\tau)$ instead, resulting in a $p \times q$ matrix of derivatives $\bar{W}_X(Z\gamma_N(\tau),\hat{\hat{\beta}}_N)Z$. Using Lemma 1, which shows that $\|\bar{W}_X(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))\| = O(1)$ and $\|\bar{W}_{XX}(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))\| = O(1)$, we have the following

$$\|\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau)) - \bar{W}(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau)) - \bar{W}_X(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))(\hat{Q}_{X,\tau} - Q_{X,\tau})\| = O_p(\|\hat{\gamma}_N(\tau) - \gamma(\tau)\|^2),$$

$$\|\bar{W}_X(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))(\hat{Q}_{X,\tau} - Q_{X,\tau}) - \bar{W}_X(Q_{X,\tau},\beta(\tau))(\hat{Q}_{X,\tau} - Q_{X,\tau})\| = O_p(\|\hat{\hat{\beta}}_N(\tau) - \beta(\tau)\|)O_p(\|\hat{\gamma}_N(\tau) - \gamma(\tau)\|),$$

and

$$\|\bar{W}_X(Q_{X,\tau},\beta(\tau))(\hat{Q}_{X,\tau} - Q_{X,\tau})\| = O_p(\|\hat{\gamma}_N(\tau) - \gamma(\tau)\|).$$

[10] Otherwise, we may impose a rate of $O_p(N^{-1/2})$ in He and Shao (2000).

Consolidating these results, we have

$$\|\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau)) - \bar{W}(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))\| = o_p(N^{-1/2}) + o_p(1)O_p(\|\hat{\hat{\beta}}_N(\tau)-\beta(\tau)\|) + O_p(N^{-1/2}).$$

Now, (28) is $o_p(1)$ since it satisfies stochastic equicontinuity established in Lemma 3, so that

$$\sup_{\tilde{\beta}(\tau)\in\Theta(\tau)} \|\bar{W}_N(Q_{X,\tau},\tilde{\beta}(\tau)) - \bar{W}_N(Q_{X,\tau},\beta(\tau)) - \left(\bar{W}(Q_{X,\tau},\tilde{\beta}(\tau)) - \bar{W}(Q_{X,\tau},\beta(\tau))\right)\| = o_p(N^{-1/2}).$$

Next, rewrite (29) as

$$\begin{aligned}
&\|\bar{W}_N(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))\| \\
\leq&\|\bar{W}_N(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau)) - \bar{W}_N(Q_{X,\tau},\beta(\tau)) - (\bar{W}(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau)) - \bar{W}(Q_{X,\tau},\beta(\tau)))\| \\
+&\|\bar{W}_N(Q_{X,\tau},\beta(\tau))\| \\
+&\|\bar{W}(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))\|. \qquad (31)
\end{aligned}$$

For (31), we have $\|\bar{W}(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))\| = (O_p(1)+o_p(1))\|\hat{\hat{\beta}}(\tau)-\beta(\tau)\|$. Hence, by stochastic equicontinuity and the fact that $\|\bar{W}_N(Q_{X,\tau},\beta(\tau))\| = O_p(N^{-1/2})$ established in Lemma 4,

$$\|\bar{W}_N(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))\| \leq o_p(N^{-1/2}) + O_p(N^{-1/2}) + (O_p(1)+o_p(1))\|\hat{\hat{\beta}}(\tau)-\beta(\tau)\|.$$

Hence, collecting the terms, we finally arrive at

$$\|\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))\| \leq (O_p(1)+o_p(1))\|\hat{\hat{\beta}}(\tau)-\beta(\tau)\| + O_p(N^{-1/2}). \qquad (32)$$

By Mean Value Theorem, we expand $\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))$ around $Q_{X,\tau}$ and $\beta(\tau)$

$$\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau)) = \bar{W}_\beta(\tilde{Q}_{X,\tau},\tilde{\beta}(\tau))(\hat{\hat{\beta}}_N(\tau)-\beta(\tau)) + \bar{W}_X(\tilde{Q}_{X,\tau},\tilde{\beta}(\tau))(\hat{Q}_{X,\tau}-Q_{X,\tau}),$$

with $\beta(\tau)\wedge\hat{\hat{\beta}}_N(\tau) < \tilde{\beta}(\tau) < \hat{\hat{\beta}}_N(\tau)\vee\beta(\tau)$ and $\hat{Q}_{X,\tau}\wedge Q_{X,\tau} < \tilde{Q}_{X,\tau} < \hat{Q}_{X,\tau}\vee Q_{X,\tau}$. With some rearrangement, this in turn implies that

$$\begin{aligned}
&|\|\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))\| - \|\bar{W}_\beta(\tilde{Q}_{X,\tau},\tilde{\beta}(\tau))(\hat{\hat{\beta}}_N(\tau)-\beta(\tau))\|| \\
\leq&\|\bar{W}_X(\tilde{Q}_{X,\tau},\tilde{\beta}(\tau))(\hat{Q}_{X,\tau}-Q_{X,\tau})\| \\
\leq&O_p(\|\hat{\gamma}_N(\tau)-\gamma(\tau)\|) \qquad (33)
\end{aligned}$$

Using the fact that $O_p(\|\hat{\gamma}_N(\tau)-\gamma(\tau)\|) = O_p(N^{-1/2})$ and $\bar{W}_\beta(\tilde{Q}_{X,\tau},\tilde{\beta}(\tau)) = O(1)$, and since $\|\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))\|$ is $O_p(\|\hat{\hat{\beta}}_N(\tau)-\beta(\tau)\|)$ or $O_p(N^{-1/2})$, we may conclude that $O_p(\|\hat{\hat{\beta}}_N(\tau)-\beta(\tau)\|) = O_p(N^{-1/2})$. The next objective is to obtain the linear representation for $\hat{\hat{\beta}}_N(\tau)$. Consider

$$
\begin{aligned}
&\|\bar{W}_N(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))-\bar{W}_N(Q_{X,\tau},\beta(\tau))-(\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))-\bar{W}(Q_{X,\tau},\beta(\tau)))\| \\
\leq&\|\bar{W}_N(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))-\bar{W}_N(Q_{X,\tau},\beta(\tau))-(\bar{W}(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))-\bar{W}(Q_{X,\tau},\beta(\tau)))\| \qquad (34)\\
&+\|\bar{W}_N(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))-\bar{W}_N(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau))-(\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))-\bar{W}(Q_{X,\tau},\hat{\hat{\beta}}_N(\tau)))\|. \qquad (35)
\end{aligned}
$$

By stochastic equicontinuity, (34) is $o_p(N^{-1/2})$ while (35) is $o_p(N^{-1/2})$ if stochastic equicontinuity can also be established for this term. Hence,

$$
\bar{W}_N(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))=\bar{W}_N(Q_{X,\tau},\beta(\tau))+\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))-\bar{W}(Q_{X,\tau},\beta(\tau))+o_p(N^{-1/2}). \qquad (36)
$$

Since $\|W_N(\hat{Q}_{X,\tau},\hat{\beta}_N(\tau))-\bar{W}_N(\hat{Q}_{X,\tau},\hat{\beta}_N(\tau))\|=O_p(N^{-3/4})$ follows from Lemma 5, we have

$$
\bar{W}_N(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))=W_N(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))+o_p(N^{-1/2}).
$$

Using the above, this implies that (36) may be rewritten as

$$
0=\bar{W}_N(Q_{X,\tau},\beta(\tau))+\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))-\bar{W}(Q_{X,\tau},\beta(\tau))+o_p(N^{-1/2}),
$$

where we have used the definition that $W_N(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))=0$ except for a finite number of points. Expanding $\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))-\bar{W}(Q_{X,\tau},\beta(\tau))$ around $\gamma(\tau)$ and $\beta(\tau)$, and recalling that $\hat{Q}_{X,\tau}=Z\hat{\gamma}_N(\tau)$ and $Q_{X,\tau}=Z\gamma(\tau)$, we have

$$
\begin{aligned}
&\bar{W}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))-\bar{W}(Q_{X,\tau},\beta(\tau))\\
=&\frac{1}{N}\sum_{n=1}^{N}f_{\tilde{u},n}(0)h_\beta(Q_{X,n,\tau},\beta(\tau))h_\beta(Q_{X,n,\tau},\beta(\tau))'(\hat{\hat{\beta}}_N(\tau)-\beta(\tau))\\
&+\frac{1}{N}\sum_{n=1}^{N}f_{\tilde{u},n}(0)h_\beta(Q_{X,n,\tau},\beta(\tau))h_X(Q_{X,n,\tau},\beta(\tau))Z_n'(\hat{\gamma}_N(\tau)-\gamma(\tau))\\
&+O_p(\|\hat{\hat{\beta}}_N(\tau)-\beta(\tau)\|^2)+O_p(\|\hat{\gamma}_N(\tau)-\gamma(\tau)\|^2)+O_p(\|\hat{\gamma}_N-\gamma(\tau)\|\|\hat{\hat{\beta}}_N(\tau)-\beta(\tau)\|).
\end{aligned}
$$

Let $Q_N=N^{-1}\sum_{n=1}^{N}f_{\tilde{u},n}(0)h_\beta(Q_{X,n,\tau},\beta(\tau))h_\beta(Q_{X,n,\tau},\beta(\tau))'$, the linear representation is

$$
\begin{aligned}
&\sqrt{N}(\hat{\hat{\beta}}_N(\tau)-\beta(\tau))\\
=&-Q_N^{-1}\frac{1}{\sqrt{N}}\sum_{n=1}^{N}h_\beta(Q_{X,n,\tau},\beta(\tau))\psi_\tau(\tilde{u}_n)\\
&-{Q_N}^{-1}\frac{1}{\sqrt{N}}\sum_{n=1}^{N}f_{\tilde{u},n}(0)h_\beta(Q_{X,n,\tau},\beta(\tau))h_X(Q_{X,n,\tau},\beta(\tau))Z_n'(\hat{\gamma}_N(\tau)-\gamma(\tau))+o_p(1).
\end{aligned}
$$

Based on the comments following A4 in Section 3, we may also express the linear representation equivalently as

$$\begin{aligned}
&\sqrt{N}(\hat{\hat{\beta}}_N(\tau) - \beta(\tau)) \\
=& - Q_N^{-1}\frac{1}{\sqrt{N}}\sum_{n=1}^{N} h_\beta(Q_{X,n,\tau},\beta(\tau))\psi_\tau(\tilde{u}_n) \\
&- Q_N{}^{-1}\frac{1}{\sqrt{N}}\sum_{n=1}^{N} f_{Y,n}(Q_Y(\tau|Q_{X,n,\tau}))h_\beta(Q_{X,n,\tau},\beta(\tau))h_X(Q_{X,n,\tau},\beta(\tau))Z_n'(\hat{\gamma}_N(\tau)-\gamma(\tau)) + o_p(1),
\end{aligned}$$

where $Q_N = N^{-1}\sum_{n=1}^{N} f_{Y,n}(Q_Y(\tau|Q_{X,n,\tau}))h_\beta(Q_{X,n,\tau},\beta(\tau))h_\beta(Q_{X,n,\tau},\beta(\tau))'$. □

Proof of Proposition 2

Let $\hat{\beta}_N$ denote the point estimator of β from the correction second stage regression. Define

$$\begin{aligned}
\tilde{u}_n &= Y_n - h(\hat{Q}_{X,n,\tau},\beta(\tau)) + h_X(\hat{Q}_{X,n,\tau},\beta(\tau))(\hat{Q}_{X,n,\tau} - Q_{X,\tau}) - O_p(\|\hat{\gamma}_N(\tau)-\gamma(\tau)\|^2) \\
&= Y_n - \breve{h}(\hat{Q}_{X,n,\tau},\beta(\tau)) - O_p(\|\hat{\gamma}_N(\tau)-\gamma(\tau)\|^2)
\end{aligned}$$

and

$$\begin{aligned}
&\breve{W}_N(\hat{Q}_{X,\tau},\hat{\beta}_N(\tau)) \\
:=&\frac{1}{N}\sum_{n=1}^{N}\breve{h}_\beta(\hat{Q}_{X,n,\tau},\hat{\beta}_N(\tau))\psi_\tau\left(\tilde{u}_n + O_p(\|\hat{\gamma}_N(\tau)-\gamma(\tau)\|^2) - \left(\breve{h}(\hat{Q}_{X,n,\tau},\hat{\beta}_N(\tau)) - \breve{h}(\hat{Q}_{X,n,\tau},\beta(\tau))\right)\right),
\end{aligned}$$

where $\tilde{u}_m$ is the innovation with its τ^{th} conditional quantile normalized to zero. Since $\breve{h}(Q_{X,n,\tau},\beta(\tau)) = h(Q_{X,n,\tau},\beta(\tau))$, notice that $\breve{W}_N(Q_{X,\tau},\beta(\tau)) = W_N(Q_{X,\tau},\beta(\tau))$. In addition, we define

$$\begin{aligned}
&\bar{\breve{W}}_N(\hat{Q}_{X,\tau},\hat{\beta}_N(\tau)) \\
:=&\frac{1}{N}\sum_{n=1}^{N}\breve{h}_\beta(Q_{X,n,\tau},\beta(\tau))\psi_\tau\left(\tilde{u}_n + O_p(\|\hat{\gamma}_N(\tau)-\gamma(\tau)\|^2) - \left(\breve{h}(\hat{Q}_{X,n,\tau},\hat{\beta}_N(\tau)) - \breve{h}(\hat{Q}_{X,n,\tau},\beta(\tau))\right)\right).
\end{aligned}$$

Hence, $E[\bar{\breve{W}}_N(Q_{X,\tau},\beta(\tau))] = E[\bar{W}_N(Q_{X,\tau},\beta(\tau))] = 0$. By the arguments in Proposition 1, we can established that $\hat{\beta}_N(\tau)$ is root-N consistent. Following this, we derive the linear representation based on stochastic equicontinuity estabilished in Lemma 3, which results in

$$\|\bar{\breve{W}}_N(\hat{Q}_{X,\tau},\hat{\beta}_N(\tau)) - \bar{\breve{W}}_N(Q_{X,\tau},\beta(\tau)) - (\bar{\breve{W}}(\hat{Q}_{X,\tau},\hat{\beta}_N(\tau)) - \bar{\breve{W}}(Q_{X,\tau},\beta(\tau)))\| = o_p(N^{-1/2}).$$

In addition, Lemma 5 implies that

$$\breve{\breve{W}}_N(\hat{Q}_{X,\tau},\hat{\beta}_N(\tau)) = \breve{W}_N(\hat{Q}_{X,\tau},\hat{\beta}_N(\tau)) + o_p(N^{-1/2}),$$

where $\breve{W}_N(\hat{Q}_{X,\tau},\hat{\beta}_N(\tau)) = 0$ except for a finite number of points. Consider,

$$\begin{aligned}
&\breve{\breve{W}}(\hat{Q}_{X,\tau},\hat{\beta}_N(\tau)) - \breve{\breve{W}}(Q_{X,\tau},\beta(\tau)) \\
=&\frac{1}{N}\sum_{n=1}^{N}\breve{h}_\beta(Q_{X,n,\tau},\beta(\tau))E\big[\psi_\tau\left(\tilde{u}_n + O_p(\|\hat{\gamma}_N(\tau)-\gamma(\tau)\|^2) - (\breve{h}(\hat{Q}_{X,n,\tau},\hat{\beta}_N(\tau)) - \breve{h}(\hat{Q}_{X,n,\tau},\beta(\tau)))\right) \\
&-\psi_\tau\left(\tilde{u}_n\right)\big]. \qquad (37)
\end{aligned}$$

Taking Taylor expansion of the expectation term in (37), we have

$$\begin{aligned}
&E\left[\psi_\tau\left(\tilde{u}_n + O_p(\|\hat{\gamma}_N(\tau)-\gamma(\tau)\|^2) - (\breve{h}(\hat{Q}_{X,n,\tau},\hat{\beta}_N(\tau)) - \breve{h}(\hat{Q}_{X,n,\tau},\beta(\tau)))\right) - \psi_\tau\left(\tilde{u}_n\right)\right] \\
=&F_{\tilde{u},n}(O_p(\|\hat{\gamma}_N(\tau)-\gamma(\tau)\|^2) + (\breve{h}(\hat{Q}_{X,n,\tau},\hat{\beta}_N(\tau)) - \breve{h}(\hat{Q}_{X,n,\tau},\beta(\tau)))) - F_{\tilde{u},n}(0) \\
=&f_{\tilde{u},n}(F^{-1}_{\tilde{u},n}(\tau))\left[O_p(\|\hat{\gamma}_N(\tau)-\gamma(\tau)\|^2) + (\breve{h}(\hat{Q}_{X,n,\tau},\hat{\beta}_N(\tau)) - \breve{h}(\hat{Q}_{X,n,\tau},\beta(\tau)))\right] + o_p(N^{-1/2}) \\
=&f_{\tilde{u},n}(F^{-1}_{\tilde{u},n}(\tau))\left[O_p(\|\hat{\gamma}_N(\tau)-\gamma(\tau)\|^2) + \breve{h}_\beta(\hat{Q}_{X,n,\tau},\beta(\tau))'(\hat{\beta}_N(\tau)-\beta(\tau)) + O_p(\|\hat{\beta}_N(\tau)-\beta(\tau)\|^2)\right] \\
&+ o_p(N^{-1/2}) \\
=&f_{\tilde{u},n}(F^{-1}_{\tilde{u},n}(\tau))\left[O_p(N^{-1}) + \breve{h}_\beta(\hat{Q}_{X,n,\tau},\beta(\tau))'(\hat{\beta}_N(\tau)-\beta(\tau))\right] + o_p(N^{-1/2}),
\end{aligned}$$

with

$$\begin{aligned}
&\breve{h}_\beta(\hat{Q}_{X,n,\tau},\beta(\tau))'(\hat{\beta}_N(\tau)-\beta(\tau)) \\
=&h_\beta(\hat{Q}_{X,n,\tau},\beta(\tau))'(\hat{\beta}_N(\tau)-\beta(\tau)) - h_{X\beta}(\hat{Q}_{X,n,\tau},\beta(\tau))'(\hat{\beta}_N(\tau)-\beta(\tau))(\hat{Q}_{X,n,\tau}-Q_{X,n,\tau}) \\
=&h_\beta(\hat{Q}_{X,n,\tau},\beta(\tau))'(\hat{\beta}_N(\tau)-\beta(\tau)) + O_p(\|\hat{\beta}_N(\tau)-\beta(\tau)\|\|\hat{\gamma}_N(\tau)-\gamma(\tau)\|) \\
=&h_\beta(Q_{X,n,\tau},\beta(\tau))'(\hat{\beta}_N(\tau)-\beta(\tau)) + O_p(\|\hat{\beta}_N(\tau)-\beta(\tau)\|\|\hat{\gamma}_N(\tau)-\gamma(\tau)\|),
\end{aligned}$$

where second line follows from the definition of $\breve{h}_\beta(\hat{Q}_{X,n,\tau},\beta(\tau))$ and the last line follows from expanding $h_\beta(\hat{Q}_{X,n,\tau},\beta(\tau))$ around $Q_{X,n,\tau}$. Since $O_p(\|\hat{\beta}_N(\tau)-\beta(\tau)\|\|\hat{\gamma}_N(\tau)-\gamma(\tau)\|) = o_p(N^{-1/2})$, (37) may be expressed as

$$\begin{aligned}
&\breve{\breve{W}}(\hat{Q}_{X,\tau},\hat{\beta}_N(\tau)) - \breve{\breve{W}}(Q_{X,\tau},\beta(\tau)) \\
&= \frac{1}{N}\sum_{n=1}^{N} f_{\tilde{u},n}(0)h_\beta(Q_{X,n,\tau},\beta(\tau))h_\beta(Q_{X,n,\tau},\beta(\tau))'(\hat{\beta}_N(\tau)-\beta(\tau)) + o_p(N^{-1/2}) \\
&= Q_N(\hat{\beta}_N(\tau)-\beta(\tau)).
\end{aligned}$$

Using the fact that $\breve{\bar{W}}_N(Q_{X,\tau}, \beta(\tau)) = W_N(Q_{X,\tau}, \beta(\tau))$ and $\breve{h}(Q_{X,n,\tau}, \beta(\tau)) = h(Q_{X,n,\tau}, \beta(\tau))$, the asymptotic representation for $\hat{\beta}_N(\tau)$ is

$$
\begin{aligned}
&\sqrt{N}(\hat{\beta}_N(\tau) - \beta(\tau)) \\
=& - Q_N^{-1} \frac{1}{\sqrt{N}} \sum_{n=1}^{N} h_\beta(Q_{X,n,\tau}, \beta(\tau)) \psi_\tau(\tilde{u}_n) + o_p(1). \quad \square
\end{aligned}
$$

Some Lemmas

The following lemmas are derived for the function h. Under appropriate assumptions for $\breve{h}$ similar to that for h, the lemmas also hold with $\breve{h}$ replacing h.

Lemma 1. *$\|\bar{W}_X(Z\bar{\gamma}, \bar{\beta})\| = O(1)$, $\|\bar{W}_{XX}(Z\bar{\gamma}, \bar{\beta})\| = O(1)$, $\|\bar{W}_\beta(Z\bar{\gamma}, \bar{\beta})\| = O(1)$, $\|\bar{W}_{\beta\beta}(Z\bar{\gamma}, \bar{\beta})\| = O(1)$ and $\|\bar{W}_{\beta X}(Z\bar{\gamma}, \bar{\beta})\| = O(1)$ uniformly in $\bar{\beta} \in \Theta$ and $\bar{\gamma} \in \Gamma$.*

Proof: For an arbitrary $\bar{\beta} \in \Theta$ and $\bar{\gamma} \in \Gamma$, where the latter implies $\bar{Q}_{X,n,\tau} = Z_n'\bar{\gamma}$, consider

$$
\begin{aligned}
&\bar{W}(\bar{Q}_{X,\tau}, \bar{\beta}) \\
=&\frac{1}{N} \sum_{n=1}^{N} E[h_\beta(Q_{X,n,\tau}, \beta(\tau)) \\
&\times \psi_\tau\left(\tilde{u}_n - \left(h(\bar{Q}_{X,n,\tau}, \bar{\beta}) - h(\bar{Q}_{X,n,\tau}, \beta(\tau))\right) - \left(h(\bar{Q}_{X,n,\tau}, \beta(\tau)) - h(Q_{X,n,\tau}, \beta(\tau))\right)\right)] \\
=&\frac{1}{N} \sum_{n=1}^{N} E[h_\beta(Q_{X,n,\tau}, \beta(\tau)) \psi_\tau\left(\tilde{u}_n - (h(\bar{Q}_{X,n,\tau}, \bar{\beta}) - h(Q_{X,n,\tau}, \beta(\tau)))\right)] \\
=&\frac{1}{N} \sum_{n=1}^{N} E[h_\beta(Q_{X,n,\tau}, \beta(\tau)) E[\psi_\tau\left(\tilde{u}_n - (h(\bar{Q}_{X,n,\tau}, \bar{\beta}) - h(Q_{X,n,\tau}, \beta(\tau)))\right) | Z_n]] \\
=&\frac{1}{N} \sum_{n=1}^{N} E[h_\beta(Q_{X,n,\tau}, \beta(\tau)) (F_{\tilde{u},n}((h(\bar{Q}_{X,n,\tau}, \bar{\beta}) - h(Q_{X,n,\tau}, \beta(\tau))) - F_{\tilde{u},n}(0))].
\end{aligned}
$$

Let the j element of $h_\beta(Q_{X,n,\tau}, \beta(\tau))$ be $h_\beta^{(j)}(Q_{X,n,\tau}, \beta(\tau))$. Following A4, which assumes that $f_{\tilde{u},n}$ and $f'_{\tilde{u},n}$ are bounded above by C, the following holds:

$$
\begin{aligned}
\|\bar{W}_X^{(j)}\|^2 =&\frac{1}{N^2} \sum_{n=1}^{N} (E[h_\beta^{(j)}(Q_{X,n,\tau}, \beta(\tau)) f_{\tilde{u},n}((h(\bar{Q}_{X,n,\tau}, \bar{\beta}) - h(Q_{X,n,\tau}, \beta(\tau))) h_X(\bar{Q}_{X,n,\tau}, \bar{\beta})])^2 \\
\leq& C^2 \frac{1}{N} \sum_{n=1}^{N} (E[h_\beta^{(j)}(Q_{X,n,\tau}, \beta(\tau)) h_X(\bar{Q}_{X,n,\tau}, \bar{\beta})])^2,
\end{aligned}
$$

where the summation over the observations follows from the fact that $\|\bar{W}_X^{(j)}\|^2 = tr(\bar{W}_X^{(j)\prime} \bar{W}_X^{(j)})$ and $\bar{W}_X^{(j)}$

is a N-vector. In addition, we have

$$\|\bar{W}_{XX}^{(j)}\|^2 = \frac{1}{N^2}\sum_{n=1}^{N}(E[h_\beta^{(j)}(Q_{X,n,\tau},\beta(\tau))(f_{\tilde{u},n}((h(\bar{Q}_{X,n,\tau},\bar{\beta}) - h(Q_{X,n,\tau},\beta(\tau)))h_{XX}(\bar{Q}_{X,n,\tau},\bar{\beta})$$
$$+ f'_{\tilde{u},n}((h(\bar{Q}_{X,n,\tau},\bar{\beta}) - h(Q_{X,n,\tau},\beta(\tau)))h_X(\bar{Q}_{X,n,\tau},\bar{\beta})^2])^2$$
$$\leq C^2\frac{1}{N}\sum_{n=1}^{N}(E[h_\beta^{(j)}(Q_{X,n,\tau},\beta(\tau))(h_{XX}(\bar{Q}_{X,n,\tau},\bar{\beta}) + h_X(\bar{Q}_{X,n,\tau},\bar{\beta})^2)])^2,$$

$$\|\bar{W}_{\beta}^{(j)}\|^2 = \sum_{k=1}^{p}(\frac{1}{N}\sum_{n=1}^{N}E[h_\beta^{(j)}(Q_{X,n,\tau},\beta(\tau))(f_{\tilde{u},n}((h(\bar{Q}_{X,n,\tau},\bar{\beta}) - h(Q_{X,n,\tau},\beta(\tau)))h_\beta^{(k)}(\bar{Q}_{X,n,\tau},\bar{\beta})])^2$$
$$\leq C^2\sum_{k=1}^{p}(\frac{1}{N}\sum_{n=1}^{N}E[h_\beta^{(j)}(Q_{X,n,\tau},\beta(\tau))h_\beta^{(k)}(\bar{Q}_{X,n,\tau},\bar{\beta})])^2,$$

$$\|\bar{W}_{\beta\beta}^{(j)}\|^2 = \sum_{k=1}^{p}(\frac{1}{N}\sum_{n=1}^{N}E[h_\beta^{(j)}(Q_{X,n,\tau},\beta(\tau))(f_{\tilde{u},n}((h(\bar{Q}_{X,n,\tau},\bar{\beta}) - h(Q_{X,n,\tau},\beta(\tau)))h_{\beta\beta}^{(kk)}(\bar{Q}_{X,n,\tau},\bar{\beta})$$
$$+ f_{\tilde{u},n}((h(\bar{Q}_{X,n,\tau},\bar{\beta}) - h(Q_{X,n,\tau},\beta(\tau)))h_\beta^{(k)}(\bar{Q}_{X,n,\tau},\bar{\beta})^2)])^2$$
$$\leq C^2\sum_{k=1}^{p}(\frac{1}{N}\sum_{n=1}^{N}E[h_\beta^{(j)}(Q_{X,n,\tau},\beta(\tau))(h_{\beta\beta}^{(kk)}(\bar{Q}_{X,n,\tau},\bar{\beta}) + h_\beta^{(k)}(\bar{Q}_{X,n,\tau},\bar{\beta})^2)])^2,$$

and

$$\|\bar{W}_{\beta X}^{(j)}\|^2 = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{p}(E[h_\beta^{(j)}(Q_{X,n,\tau},\beta(\tau))[f_{\tilde{u},n}((h(\bar{Q}_{X,n,\tau},\bar{\beta}) - h(Q_{X,n,\tau},\beta(\tau)))h_{\beta X}^{(k)}(\bar{Q}_{X,n,\tau},\bar{\beta})$$
$$+ f'_{\tilde{u},n}((h(\bar{Q}_{X,n,\tau},\bar{\beta}) - h(Q_{X,n,\tau},\beta(\tau)))h_\beta^{(k)}(\bar{Q}_{X,n,\tau},\bar{\beta})h_X(\bar{Q}_{X,n,\tau},\bar{\beta})'])^2$$
$$\leq C\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{p}(E[h_\beta^{(j)}(Q_{X,n,\tau},\beta(\tau))(h_{\beta X}^{(k)}(\bar{Q}_{X,n,\tau},\bar{\beta}) + h_\beta^{(k)}(\bar{Q}_{X,n,\tau},\bar{\beta})h_X(\bar{Q}_{X,n,\tau},\bar{\beta})')])^2.$$

Since A3 assumes that $E|h_\beta^{(k)}(\bar{Q}_{X,n,\tau},\bar{\beta})|^4$, $E|h_X(\bar{Q}_{X,n,\tau},\bar{\beta})|^4$, $E|h_{XX}(\bar{Q}_{X,n,\tau},\bar{\beta})|^4$ and $E|h_{\beta\beta}^{(kk)}(\bar{Q}_{X,n,\tau},\bar{\beta})|^2$ and $E|h_{\beta X}^{(k)}(\bar{Q}_{X,n,\tau},\bar{\beta})|^4$ are bounded above by $\Delta < \infty$, it follows from Cauchy-Schwarz inequality, this implies that $\|\bar{W}_X^{(j)}\| = O(1)$, $\|\bar{W}_{XX}^{(j)}\| = O(1)$, $\|\bar{W}_\beta^{(j)}\| = O(1)$, $\|\bar{W}_{\beta\beta}^{(j)}\| = O(1)$ and $\|\bar{W}_{\beta X}^{(j)}\| = O(1)$. Furthermore, since this is true for each j, Lemma 1 follows. □

Before proving Lemma 3, we briefly review the notion of weak dependence. Let $\mathcal{F}_m^n = \sigma(X_i : m \leq i \leq n, i \in \mathbb{N})$ be the σ-field generated by the random variables $X_m, \dots, X_n$, $1 \leq m \leq n \leq \infty$. A variable X_i is said to be α-mixing if the mixing coefficient

$$\alpha(n) = \sup_{m\in\mathbb{N}} \sup_{\mathcal{F}_1^m, \mathcal{F}_{m+n}^\infty} |P(A \cup B) - P(A)P(B)|$$

goes to zero as n increases without bound. For the proof, we restate a lemma due to Bosq (1998), which is stated as Lemma 2 in Sun (2006):

Lemma 2. *Let X_n be a zero mean real-valued process of strongly mixing random variables where $\alpha(N) = O(N^{-\zeta})$ for some some $\zeta > 0$. Suppose that there exists $c > 0$ such that*

$$E|X_n|^\varphi \leq c^{\varphi-2}\varphi!E|X_n|^2 < \infty, \qquad t = 1, \dots, N, \quad \varphi \geq 3,$$

then for each $N \geq 2$, each integer $q \in [1, N/2]$, each $t > 0$ and each $\varphi \geq 3$,

$$P(|\sum_{n=1}^N X_n| > Nt) \leq a_1 \exp\left(-\frac{at^2}{25m_2^2 + 5ct}\right) + a_2(\varphi)\alpha\left(\lfloor\frac{N}{q+1}\rfloor\right),$$

where

$$a_1 = 2\frac{N}{q} + 2\left(1 + \frac{at^2}{25m_2^2 + 5ct}\right) \qquad \text{with} \quad m_2^2 = \max_{1\leq t\leq N} E[X_n^2]$$

and

$$a_2(\varphi) = 11N\left(1 + \frac{5m_\varphi^{\varphi/(2\varphi+1)}}{t}\right) \qquad \text{with} \quad m_\varphi = \max_{1\leq t\leq N}(E|X_n|^\varphi)^{1/\varphi}. \quad \square$$

Recall that the notation $E[\bar{W}_N(\hat{Q}_{X,\tau}, \hat{\beta}_N(\tau)) = \bar{W}(\hat{Q}_{X,\tau}, \hat{\beta}_N(\tau))]$ is used when proving Propositions 1 and 2.

Lemma 3. *(Stochastic Equicontinuity) Under A1-A4,*

$$\begin{aligned}&\sup_{\tilde{\beta}\in\Theta}\|\bar{W}_N(Q_{X,\tau}, \tilde{\beta}) - \bar{W}_N(Q_{X,\tau}, \beta(\tau)) - \left(\bar{W}(Q_{X,\tau}, \tilde{\beta}) - \bar{W}(Q_{X,\tau}, \beta(\tau))\right)\| \\ &= O_p(N^{-3/4+\lambda}\log N), \quad a.s. \end{aligned} \tag{38}$$

where $\lambda \in ((11/4 + p)/(1 + \zeta(1-\nu)), 1/2)$ with $\zeta > (4.5 + 2p)/(1-\nu)$, $\nu > 0$, p is the dimension of the parameter space Θ and $Q_N = N^{-1}\sum_n h_\beta(Q_{X,n,\tau}, \beta(\tau))h_\beta(Q_{X,n,\tau}, \beta(\tau))'$.

Proof: The proof is similar to Sun (2006) who derived the asymptotic representation for the sample quantile under weak dependence. Here, a pointwise relationship of (38) is first established where the bound is an exponential tail, which a chaining argument is later applied. For p-dimensional parameter space Θ, let $j = 1, \dots, p$ index the j element in the p-vector $\tilde{\beta}$. For each j, consider $\tilde{\beta}^{(j)}(\tau) - \beta^{(j)}(\tau) = r\epsilon_N$, where $r = 0, \pm1, \dots, \pm\delta_N$, $\delta_N = \lceil N^{1/4}\rceil$ and $\epsilon_N = k_p^{-1}N^{-3/4}\log N$ where k_p is a positive constant. Since

$\tilde{\beta}^{(j)}(\tau) - \beta^{(j)}(\tau)$ is divided into $(2\delta_N + 1)$ partitions for each j, there is a collection of $(2\delta_N + 1)^p$ partitions of p-dimensional cubes each indexed by $\mathcal{E}_i$. The cardinality has a rate of $O(N^{p/4})$ while for each $\mathcal{E}_i$, $\tilde{\beta}_i^{(j)}(\tau) - \beta^{(j)}(\tau) = O(N^{-1/2} \log N)$. Now, define

$$
\begin{aligned}
&\Omega(Q_{X,\tau}, \tilde{\beta}) \\
\equiv &\frac{1}{N} \sum_{n=1}^{N} h_\beta(Q_{X,n,\tau}, \beta(\tau))[\psi(\tilde{u}_n - (h(Q_{X,n,\tau}, \tilde{\beta}) - h(Q_{X,n,\tau}, \beta(\tau))) - \psi_\tau(\tilde{u}_n) \\
&- E[\psi(\tilde{u}_n - (h(Q_{X,n,\tau}, \tilde{\beta}) - h(Q_{X,n,\tau}, \beta(\tau))) - \psi_\tau(\tilde{u}_n)]] \\
= &\frac{1}{N} \sum_{n=1}^{N} h_\beta(Q_{X,n,\tau}, \beta(\tau))(V_n - p),
\end{aligned}
$$

where we have defined $V_n \equiv \psi_\tau(\tilde{u}_n - (h(Q_{X,n,\tau}, \tilde{\beta}) - h(Q_{X,n,\tau}, \beta(\tau))) - \psi_\tau(\tilde{u}_n)$. The expectation of V_n is p, which may in turn be expressed as $p = P(F_{\tilde{u},n}^{-1}(\tau) \leq \tilde{u}_n \leq h(Q_{X,n,\tau}, \tilde{\beta}) - h(Q_{X,n,\tau}, \beta(\tau))) = F_{\tilde{u},n}(h(Q_{X,n,\tau}, \tilde{\beta}) - h(Q_{X,n,\tau}, \beta(\tau))) - F_{\tilde{u},n}(F_{\tilde{u},n}^{-1}(\tau))$. Without loss of generality, let $h(Q_{X,n,\tau}, \tilde{\beta}) - h(Q_{X,n,\tau}, \beta(\tau)) > 0$. For $\tilde{\beta} \neq \beta(\tau)$ and using the assumption that $\tilde{\beta} - \beta(\tau) < k_p^{-1} N^{-1/2} \log N$ as well as A3, p may also be expressed as $0 < p < C k_p^{-1} N^{-1/2} \log N$. Since p is the probability where $V_n = 1$, the following relationship holds

$$
E|V_n - p|^\varphi = |1 - p|^\varphi p + |p|^\varphi |1 - p| = (1 - p)p(p^{\varphi - 1} + (1 - p)^{\varphi - 1}) < p < \infty. \tag{39}
$$

Notice that $E|V_n - p|^\varphi \leq E|V_n - p|^2 \leq p$ for $\varphi \geq 3$. For the j element of $h_\beta(Q_{X,n,\tau}, \beta(\tau))$, consider $\varphi = 3$ so that

$$
\begin{aligned}
&E|h_\beta^{(j)}(Q_{X,n,\tau}, \beta(\tau))(V_n - p)|^3| \\
\leq &E[|h_\beta^{(j)}(Q_{X,n,\tau}, \beta(\tau))|^3 |V_n - p|^3] \\
= &E[|h_\beta^{(j)}(Q_{X,n,\tau}, \beta(\tau))|^3 E[|V_n - p|^3 | X_n]] \\
\leq &E[|h_\beta^{(j)}(Q_{X,n,\tau}, \beta(\tau))|^3 p] \\
\leq &C \max_k |\tilde{\beta}^{(k)} - \beta^{(k)}(\tau)| \sum_{k=1}^{p} E[|h_\beta^{(j)}(Q_{X,n,\tau}, \beta(\tau))|^3 |h_\beta^{(k)}(Q_{X,n,\tau}, \beta(\tau))|] \\
\leq &pC \max_k |\tilde{\beta}^{(k)} - \beta^{(k)}(\tau)| \max_k E|h_\beta^{(k)}(Q_{X,n,\tau}, \beta(\tau))|^4 \\
\leq &pC\Delta \max_k |\tilde{\beta}^{(k)} - \beta^{(k)}(\tau)| \\
= &pC\Delta k_p^{-1} N^{-1/2} \log N \\
< &\infty,
\end{aligned}
$$

where the third line follows from the law of iterated expectation, the forth line follows from (39), fifth line follows from A4 and Taylor expansion of p around $\beta(\tau)$, and the seventh line follows from A3. Hence, we may apply Lemma 2 for $\varphi = 3$. For this purpose, note that $m_2^2 = E|h_\beta^{(j)}(Q_{X,n,\tau}, \beta(\tau))(V_n - p)|^2 \leq k_p^{-1} N^{-1/2} \log N$ for each j, where k_p^{-1} subsumes the other constants in the expression. Also, note that $\Omega(Q_{X,n,\tau}, \tilde{\beta}(\tau))$ is a p vector with j element $\Omega^{(j)}(Q_{X,\tau}, \tilde{\beta}(\tau))$. Considering $\tilde{\beta} \in E_i$ and substituting $c = 1$ and $t = N^{-3/4+\lambda} \log N$, where $\lambda > 0$ is a constant, yields

$$P(|\Omega^{(j)}(Q_{X,\tau}, \tilde{\beta}(\tau))| > N^{-3/4+\lambda} \log N) = \quad \text{TERM 1} \quad + \quad \text{TERM 2}.$$

For some q, we have

$$\text{TERM 1} = a_1 \exp\left(-\frac{qt^2}{25m_2^2 + 5ct}\right) \leq a_1 O(\exp(-qN^{-1+\lambda} k_p \log N)) = O(\frac{N}{q}) O(\exp(-qN^{-1+\lambda} \log N^{k_p})),$$

where we have used the fact that $m_2^2 < p$ and $ct = o(m_2^2)$. In addition,

$$\begin{aligned}
&\text{TERM 2} \\
=& a_2(\varphi)\alpha\left(\lfloor\frac{N}{q+1}\rfloor\right) \\
=& 11N(1 + \frac{5(E|V_n - p|^\varphi)^{1/(2\varphi+1)}}{t}))\alpha\left(\lfloor\frac{N}{q+1}\rfloor\right) \\
\leq& 11(N + N\frac{5(E|V_n - p|^2)^{1/(2\varphi+1)}}{t}))\alpha\left(\lfloor\frac{N}{q+1}\rfloor\right) \\
\leq& 11(N + N\frac{5p^{1/(2\varphi+1)}}{t}))\alpha\left(\lfloor\frac{N}{q+1}\rfloor\right) \\
\leq& 11(N + N\frac{5(k_p^{-1}N^{-1/2+\lambda}\log N)^{1/(2\varphi+1)}}{N^{-3/4+\lambda}\log N})\alpha\left(\lfloor\frac{N}{q+1}\rfloor\right) \\
\leq& 11(N + N\frac{5(k_p^{-1}N^{-1/2+\lambda}\log N)^{1/(2\varphi+1)}}{N^{-3/4+\lambda}\log N})\alpha(\frac{N}{q}) \\
\leq& O_p(N^{7/4-\lambda}\log N^{-1}(N^{-1/2+\lambda}\log N)^{1/(2\varphi+1)})\alpha(\frac{N}{q}).
\end{aligned}$$

Upon choosing $q = N^{1-\lambda} \log N^\lambda$, TERM 1 becomes

$$\text{TERM 1} \leq O(\frac{N^\lambda}{\log N^\lambda}) O(\exp(-(\log N)^{k_p+\lambda})) \leq O(\frac{N^\lambda}{\log N^\lambda}) O(N^{-(k_p+\lambda)}) \leq O(\frac{1}{N^{k_p}}\frac{1}{\log N^\lambda}) = o(N^{-k_p}).$$

where in the first line, we have used the fact that $(\log N)^{k_p+\lambda} > (k_p + \lambda)\log N$ for $N > 1$. For TERM 2, consider some $\nu > 0$ such that

$$
\begin{aligned}
&\text{TERM 2} \\
\leq & O_p(N^{7/4-\lambda}\log N^{-1}(N^{-1/2+\lambda}\log N)^{1/(2\varphi+1)})O_p(N^{-\lambda\zeta}\log N^{\lambda\zeta}) \\
\leq & O_p(N^{7/4-\lambda(1+\zeta(1-\nu))})(N^{-1/2-\nu\lambda\zeta(2\varphi+1)}\log N^{1/2+(\lambda\zeta-1)(2\varphi+1)})^{1/(2\varphi+1)} \\
\leq & O_p(N^{7/4-\lambda(1+\zeta(1-\nu))})o_p(1).
\end{aligned}
$$

The $o_p(1)$ term comes from imposing the condition $-1/2-\nu\lambda\zeta(2\varphi+1)+1/2+(\lambda\zeta-1)(2\varphi+1) < 0$ by choosing an appropriate value of $\nu > 0$. For almost sure convergence, ζ is chosen to satisfy $7/4-\lambda(1+\zeta(1-\nu))+p/4 < -1$. The inclusion of $p/4$ is needed for the chaining argument, as the number of cubes grows at a rate of $O(\delta_N^p) = O(N^{p/4})$. Our choice of ζ implies the restriction of $\zeta > (4.5+2p)/(1-\nu)$. For this ζ and $\nu > 0$, $\lambda \in ((11/4+p)/(1+\zeta(1-\nu)), 1/2)$. To complete the argument, choose $k_p > 1+p/4$ to obtain a fast enough rate of convergence for TERM 1. □

Lemma 4. *Under A1 and A3,* $\|\bar{W}_N(Q_{X,\tau}, \beta(\tau))\| = O_p(N^{-1/2})$.

Proof: Chebyshev inequality implies

$$
\begin{aligned}
& P(\|\frac{1}{N}\sum_{n=1}^{N} h_\beta(Q_{X,n,\tau}, \beta(\tau))\psi_\tau(\tilde{u}_n)\| > t) \\
\leq & E\|\frac{1}{N}\sum_{n=1}^{N} h_\beta(Q_{X,n,\tau}, \beta(\tau))\psi_\tau(\tilde{u}_n)\|^2/t^2 \\
= & \frac{1}{N^2}E[\sum_{j=1}^{p}(\sum_{n=1}^{N} h_\beta^{(j)}(Q_{X,n,\tau}, \beta(\tau))\psi_\tau(\tilde{u}_n))^2]/t^2 \\
= & \frac{1}{N^2}\sum_{j=1}^{p}E[(\sum_{n=1}^{N} h_\beta^{(j)}(Q_{X,n,\tau}, \beta(\tau))\psi_\tau(\tilde{u}_n))^2]/t^2.
\end{aligned}
$$

Recognizing that $E[(\sum_{n=1}^{N} h_\beta^{(j)}(Q_{X,n,\tau}, \beta(\tau))\psi_\tau(\tilde{u}_n))^2] = Var(E[(\sum_{n=1}^{N} h_\beta^{(j)}(Q_{X,n,\tau}, \beta(\tau))\psi_\tau(\tilde{u}_n))^2])$, we have

$$
\begin{aligned}
&Var(\sum_{n=1}^{N} h_{\beta}^{(j)}(Q_{X,n,\tau},\beta(\tau))\psi_{\tau}(\tilde{u}_n))\\
=&\sum_{n=1}^{N} Var(h_{\beta}^{(j)}(Q_{X,n,\tau},\beta(\tau))\psi_{\tau}(\tilde{u}_n)) + \sum_{n\neq m} E[h_{\beta}^{(j)}(Q_{X,m,\tau},\beta(\tau))\psi_{\tau}(\tilde{u}_m), h_{\beta}^{(j)}(Q_{X,n,\tau},\beta(\tau))\psi_{\tau}(\tilde{u}_n)]\\
=&\sum_{n=1}^{N} \tau(1-\tau)E|h_{\beta}^{(j)}(Q_{X,n,\tau},\beta(\tau))|^2 + \sum_{n\neq m} E[h_{\beta}^{(j)}(Q_{X,m,\tau},\beta(\tau))h_{\beta}^{(j)}(Q_{X,n,\tau},\beta(\tau))\psi_{\tau}(\tilde{u}_m)\psi_{\tau}(\tilde{u}_n)]\\
\leq&\sum_{n=1}^{N} \tau(1-\tau)E|h_{\beta}^{(j)}(Q_{X,n,\tau},\beta(\tau))|^2 + \sum_{n\neq m} (E|(h_{\beta}^{(j)}(Q_{X,m,\tau},\beta(\tau))h_{\beta}^{(j)}(Q_{X,n,\tau},\beta(\tau)))|^2)^{1/2}(E[\psi_{\tau}(\tilde{u}_m)\psi_{\tau}(\tilde{u}_n)])^{1/2}\\
\leq&\sum_{n=1}^{N} \tau(1-\tau)E|h_{\beta}^{(j)}(Q_{X,n,\tau},\beta(\tau))|^2\\
&+2\sum_{n\neq m}((P(\tilde{u}_n,\tilde{u}_m\leq 0)-\tau^2)\alpha(|m-n|))^{1/2}(E|h_{\beta}^{(j)}(Q_{X,m,\tau},\beta(\tau))h_{\beta}^{(j)}(Q_{X,n,\tau},\beta(\tau))|^2)^{1/2}\\
\leq&\max_n E|h_{\beta}^{(j)}(Q_{X,n,\tau},\beta(\tau))|^2 \sum_{n=1}^{N}\tau(1-\tau)\\
&+2\max_n (E|h_{\beta}^{(j)}(Q_{X,n,\tau},\beta(\tau))|^4)^{1/2}\sum_{n\neq m}((P(\tilde{u}_n,\tilde{u}_m\leq 0)-\tau^2)\alpha(|m-n|))^{1/2}\\
\leq&\Delta\left(\sum_{n=1}^{N}(1-\tau^2)+2\sum_{n\neq m}((1-\tau^2)\alpha(|m-n|))^{1/2}\right)\\
=&\Delta\left(N(1-\tau^2)+4(1-\tau^2)^{1/2}\sum_{n=2}^{N}\alpha(n-1)^{1/2}\right),
\end{aligned}
$$

where we have applied the covariance for the summands of α-mixing sequences by Doukhan (1994) in the second inequality. Let $\tilde{\zeta}=\zeta/2$, where $-\zeta$ is the size of mixing. Note that $n^{-\tilde{\zeta}} \leq \int_{n-1}^{n} j^{-\tilde{\zeta}}dj$ follows from the fact that $n^{-\tilde{\zeta}}$ is the lower sum of the Riemann integral above. Therefore, for some constant C such that $\alpha(n)^{1/2}=Cn^{-\tilde{\zeta}}$

$$
\lim_{N\to\infty}\sum_{n=2}^{N}\alpha(n-1)^{1/2} \leq \lim_{N\to\infty}\sum_{n=2}^{N} C\int_{n-1}^{n} j^{-\tilde{\zeta}}dj = C\int_{1}^{\infty} j^{-\tilde{\zeta}}dj = C\frac{1}{\tilde{\zeta}-1}=O(1).
$$

This implies that $Var(\sum_{n=1}^{N} h_{\beta}^{(j)}(Q_{X,n,\tau},\beta(\tau))\psi_{\tau}(\tilde{u}_n))=O(N)$. Since, $P(\|\frac{1}{N}\sum_{n=1}^{N} h_{\beta}(Q_{X,n,\tau},\beta(\tau))\psi_{\tau}(\tilde{u}_n)\| > t)=O(N^{-1})$, the conclusion follows by letting $t=O(N^{-1/2})$. $\square$

Lemma 5. *Suppose $\hat{\hat{\beta}}_N$ is a sequence of p vector such that $\hat{\hat{\beta}}_N^{(k)}-\beta^{(k)}(\tau)=C_{\beta}^{k}N^{-1/2}\epsilon_N^{(k)}$ and $\hat{\gamma}_N$ is a q vector such that $\hat{\gamma}_N^{(l)}-\gamma^{(l)}(\tau)=C_{\gamma}^{l}N^{-1/2}\nu_N^{(l)}$, where C_{β}^{k} and C_{γ}^{l} are constants and $E|\epsilon_N^{(k)}|^4<\infty$, $E|\nu_N^{(l)}|^4<\infty$ and $E|\epsilon_N^{(k)}\nu_N^{(l)}|^4<\infty$ for all k, l, and N. Then, under A3 and A7, $\|W_N(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))-\bar{W}_N(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))\|=O_p(N^{-3/4})$.*

Proof: Following the argument in Lemma 2, consider the j element

$$
\begin{aligned}
&E|W_N^{(j)}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))-\bar{W}_N^{(j)}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))| \\
=&E|\frac{1}{N}\sum_{n=1}^{N}\left(h_\beta^{(j)}(\hat{Q}_{X,n,\tau},\hat{\hat{\beta}}(\tau))-h_\beta^{(j)}(Q_{X,\tau},\beta(\tau))\right) \\
&\times\psi_\tau\left(\tilde{u}_n-(h(\hat{Q}_{X,n,\tau},\hat{\hat{\beta}}_N(\tau)))-h(\hat{Q}_{X,n,\tau},\beta(\tau)))-(h(\hat{Q}_{X,n,\tau},\beta(\tau)))-h(Q_{X,n,\tau},\beta(\tau)))\right)| \\
\leq&\max_n(E|h_\beta^{(j)}(Z_n'\hat{\gamma}_N(\tau),\hat{\beta}(\tau))-h_\beta^{(j)}(Z_n'\gamma(\tau),\beta(\tau))|^2)^{1/2} \\
\times&\frac{1}{N}\sum_{n=1}^{N}(E[\psi_\tau\left(\tilde{u}_n-(h(\hat{Q}_{X,n,\tau},\hat{\hat{\beta}}_N(\tau)))-h(\hat{Q}_{X,n,\tau},\beta(\tau)))-(h(\hat{Q}_{X,n,\tau},\beta(\tau)))-h(Q_{X,n,\tau},\beta(\tau)))\right)]^2)^{1/2},
\end{aligned}
$$

where the last line follows from Cauchy-Schwarz inequality. Now, letting $E[\psi_\tau]=p$, note that $E[\psi_\tau]^2=Var(\psi_\tau)+(E[\psi_\tau])^2=p(1-p)+p^2\leq 2p$. Following this argument, we have

$$
\begin{aligned}
&E|W_N^{(j)}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))-\bar{W}_N^{(j)}(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))| \\
\leq&(\max_n E|h_\beta^{(j)}(Z_n'\hat{\gamma}_N(\tau),\hat{\beta}(\tau))-h_\beta^{(j)}(Z_n'\gamma(\tau),\beta(\tau))|^2)^{1/2} \\
\times&\frac{1}{N}\sum_{n=1}^{N}(2E[\psi_\tau\left(\tilde{u}_n-(h(\hat{Q}_{X,n,\tau},\hat{\hat{\beta}}_N(\tau)))-h(\hat{Q}_{X,n,\tau},\beta(\tau)))-(h(\hat{Q}_{X,n,\tau},\beta(\tau)))-h(Q_{X,n,\tau},\beta(\tau)))\right)])^{1/2}.
\end{aligned}
$$

Consider a Taylor expansion,

$$
\begin{aligned}
&E|h_\beta^{(j)}(Z_n'\hat{\gamma}_N(\tau),\hat{\beta}(\tau))-h_\beta^{(j)}(Z_n'\gamma(\tau),\beta(\tau))|^2 \\
=&E|\sum_{l=1}^{q}h_{\beta X}^{(j)}(Q_{X,n,\tau},\beta(\tau))Z_n^{(l)}(\hat{\gamma}_N^{(l)}(\tau)-\gamma^{(l)}(\tau))+\sum_{k=1}^{p}h_{\beta\beta}^{(jk)}(Q_{X,n,\tau},\beta(\tau))(\hat{\hat{\beta}}_N^{(k)}(\tau)-\beta^{(k)}(\tau)) \\
&+2\sum_{k=1}^{p}\sum_{l=1}^{q}h_{\beta\beta X}^{(jk)}(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))(\hat{\hat{\beta}}_N^{(k)}(\tau)-\beta^{(k)}(\tau))Z_n^{(l)}(\hat{\gamma}_N^{(l)}(\tau)-\gamma^{(l)}(\tau))|^2 \\
\leq&((E|\sum_{l=1}^{q}h_{\beta X}^{(j)}(Q_{X,n,\tau},\beta(\tau))Z_n^{(l)}(\hat{\gamma}_N^{(l)}(\tau)-\gamma^{(l)}(\tau))|^2)^{1/2}+(E|\sum_{k=1}^{p}h_{\beta\beta}^{(jk)}(Q_{X,n,\tau},\beta(\tau))(\hat{\hat{\beta}}_N^{(k)}(\tau)-\beta^{(k)}(\tau))|^2)^{1/2} \\
&+2(E|\sum_{k=1}^{p}\sum_{l=1}^{q}h_{\beta\beta X}^{(jk)}(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))(\hat{\hat{\beta}}_N^{(k)}(\tau)-\beta^{(k)}(\tau))Z_n^{(l)}(\hat{\gamma}_N^{(l)}(\tau)-\gamma^{(l)}(\tau))|^2)^{1/2})^2,
\end{aligned}
$$

where $\hat{\hat{\beta}}_N(\tau)\wedge\beta(\tau)\leq\tilde{\beta}(\tau)\leq\hat{\hat{\beta}}_N(\tau)\vee\beta(\tau)$ and $\hat{Q}_{X,n,\tau}\wedge\hat{Q}_{X,n,\tau}\leq\tilde{Q}_{X,n,\tau}\leq\hat{Q}_{X,n,\tau}\vee\hat{Q}_{X,n,\tau}$. Analyzing one term at the time, we first examine

$$
\begin{aligned}
&E|\sum_{l=1}^{q} h_{\beta X}^{(j)}(Q_{X,n,\tau},\beta(\tau))Z_n^{(l)}(\hat{\gamma}_N^{(l)}(\tau)-\gamma^{(l)}(\tau))|^2\\
\leq&[\sum_{l=1}^{q}(E|h_{\beta X}^{(j)}(Q_{X,n,\tau},\beta(\tau))Z_n^{(l)}(\hat{\gamma}_N^{(l)}(\tau)-\gamma^{(l)}(\tau))|^2)^{1/2}]^2\\
\leq&[\sum_{l=1}^{q}((E|h_{\beta X}^{(j)}(Q_{X,n,\tau},\beta(\tau))Z_n^{(l)}|^4)^{1/2}(E|\hat{\gamma}_N^{(l)}(\tau)-\gamma^{(l)}(\tau)|^4)^{1/2})^{1/2}]^2\\
\leq&[\sum_{l=1}^{q}(E|h_{\beta X}^{(j)}(Q_{X,n,\tau},\beta(\tau))Z_n^{(l)}|^4)^{1/4}(E|\hat{\gamma}_N^{(l)}(\tau)-\gamma^{(l)}(\tau)|^4)^{1/4}]^2\\
\leq&q^2\Delta^{1/2}N^{-1}\max_l(E|C_\gamma^l\nu_N^l|^4)^{1/2}\\
=&O(N^{-1}).
\end{aligned}
$$

The second line follows from Minkowski's inequality, the third line follows from Cauchy-Schwarz inequality, the fifth line follows from A3. In addition,

$$
\begin{aligned}
&E|\sum_{k=1}^{p} h_{\beta\beta}^{(jk)}(Q_{X,n,\tau},\beta(\tau))(\hat{\hat{\beta}}_N^{(k)}(\tau)-\beta^{(k)}(\tau))|^2\\
\leq&[\sum_{k=1}^{p}(E|h_{\beta\beta}^{(jk)}(Q_{X,n,\tau},\beta(\tau))|^4)^{1/4}(E|\hat{\hat{\beta}}_N^{(k)}(\tau)-\beta^{(k)}(\tau)|^4)^{1/4}]^2\\
\leq&p^2\Delta^{1/2}N^{-1}\max_k(E|C_\beta^k\epsilon_N^k|^4)^{1/2}\\
=&O(N^{-1}).
\end{aligned}
$$

Finally, notice that

$$
\begin{aligned}
&E|\sum_{k=1}^{p}\sum_{l=1}^{q} h_{\beta\beta X}^{(jk)}(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))(\hat{\hat{\beta}}_N^{(k)}(\tau)-\beta^{(k)}(\tau))Z_n^{(l)}(\hat{\gamma}_N^{(l)}(\tau)-\gamma^{(l)}(\tau))|^2\\
\leq&[\sum_{k=1}^{p}\sum_{l=1}^{q}(E|h_{\beta\beta X}^{(jk)}(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))Z_n^{(l)}|^4)^{1/4}(E|(\hat{\hat{\beta}}_N^{(k)}(\tau)-\beta^{(k)}(\tau))(\hat{\gamma}_N^{(l)}(\tau)-\gamma^{(l)}(\tau))|^4)^{1/4}]^2\\
\leq&p^2q^2\Delta^{1/2}\max_{k,l}(E|(\hat{\hat{\beta}}_N^{(k)}(\tau)-\beta^{(k)}(\tau))(\hat{\gamma}_N^{(l)}(\tau)-\gamma^{(l)}(\tau))|^4)^{1/2}\\
\leq&p^2q^2\Delta^{1/2}N^{-1}\max_{k,l}(E|C_\beta^k\epsilon_N^kC_\gamma^l\nu_N^l|^4)^{1/2}\\
\leq&O(N^{-1}).
\end{aligned}
$$

Together, the above implies that $\max_n(E|h_\beta^{(j)}(Z_n'\hat{\gamma}_N(\tau),\hat{\beta}(\tau))-h_\beta^{(j)}(Z_n'\gamma(\tau),\beta(\tau))|^2)^{1/2}=O(N^{-1/2})$. Next, note that

$$
\begin{aligned}
&E[\psi_\tau\left(\tilde{u}_n - (h(\hat{Q}_{X,n,\tau},\hat{\hat{\beta}}_N(\tau))) - h(\hat{Q}_{X,n,\tau},\beta(\tau))) - (h(\hat{Q}_{X,n,\tau},\beta(\tau))) - h(Q_{X,n,\tau},\beta(\tau)))\right)] \\
=&F_{\tilde{u},n}(h(\hat{Q}_{X,n,\tau},\hat{\hat{\beta}}_N(\tau))) - h(\hat{Q}_{X,n,\tau},\beta(\tau))) - (h(\hat{Q}_{X,n,\tau},\beta(\tau))) - h(Q_{X,n,\tau},\beta(\tau))) - F_{\tilde{u},n}(0) \\
\leq&C[h_\beta(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))'(\hat{\hat{\beta}}_N(\tau)-\beta(\tau)) + h_X(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))Z_n'(\hat{\gamma}_N(\tau)-\gamma(\tau))].
\end{aligned}
$$

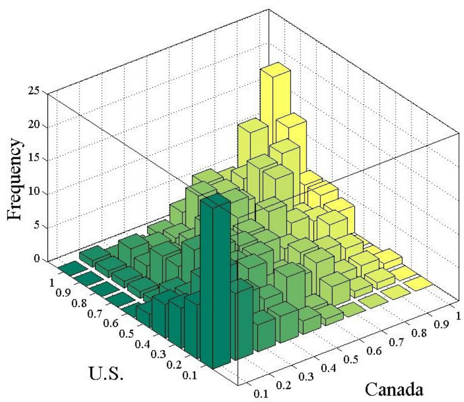
Using the above, we have

$$
\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}(2E[\psi_\tau\left(\tilde{u}_n - (h(\hat{Q}_{X,n,\tau},\hat{\hat{\beta}}_N(\tau))) - h(\hat{Q}_{X,n,\tau},\beta(\tau))) - (h(\hat{Q}_{X,n,\tau},\beta(\tau))) - h(Q_{X,n,\tau},\beta(\tau)))\right)])^{1/2} \\
\leq&\sqrt{2}C\frac{1}{N}\sum_{n=1}^{N}[h_\beta(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))'(\hat{\hat{\beta}}_N(\tau)-\beta(\tau)) + h_X(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))Z_n'(\hat{\gamma}_N(\tau)-\gamma(\tau))]^{1/2} \\
\leq&\sqrt{2}C\frac{1}{N}\sum_{n=1}^{N}[|h_\beta(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))'(\hat{\hat{\beta}}_N(\tau)-\beta(\tau))|^{1/2} + |h_X(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))Z_n'(\hat{\gamma}_N(\tau)-\gamma(\tau))|^{1/2}] \\
\leq&\sqrt{2}C\frac{1}{N}\sum_{n=1}^{N}[\|h_\beta(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))\|^{1/2}\|\hat{\hat{\beta}}_N(\tau)-\beta(\tau)\|^{1/2} + \|h_X(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))\|^{1/2}\|Z_n'(\hat{\gamma}_N(\tau)-\gamma(\tau))\|^{1/2}] \\
\leq&\sqrt{2}C[\|\hat{\hat{\beta}}_N(\tau)-\beta(\tau)\|^{1/2}\frac{1}{N}\sum_{n=1}^{N}\|h_\beta(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))\|^{1/2} + \|Z_n'(\hat{\gamma}_N(\tau)-\gamma(\tau))\|^{1/2}\frac{1}{N}\sum_{n=1}^{N}\|h_X(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))\|^{1/2}] \\
=&O_p(N^{-1/4})\frac{1}{N}\sum_{n=1}^{N}\|h_\beta(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))\|^{1/2} + O_p(N^{-1/4})\frac{1}{N}\sum_{n=1}^{N}\|h_X(\tilde{Q}_{X,n,\tau},\tilde{\beta}(\tau))\|^{1/2} \\
=&O_p(N^{-1/4}),
\end{aligned}
$$

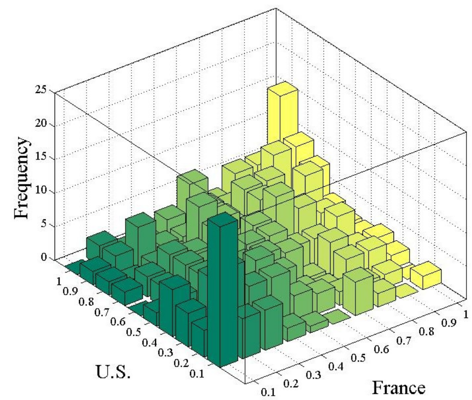
where the second last line follows from the fact that both $\hat{\gamma}_N(\tau)$ and $\hat{\hat{\beta}}_N(\tau)$ have the rate of $O_p(N^{-1/2})$ and the last line follows from A7. Hence, $E\|W_N(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau)) - \bar{W}_N(\hat{Q}_{X,\tau},\hat{\hat{\beta}}_N(\tau))\| = O_p(N^{-3/4})$. The conclusion follows from Markov inequality. □

Figure 1: **Bivariate Histograms of U.S. Returns with Canada, France, Germany and U.K. Returns.**

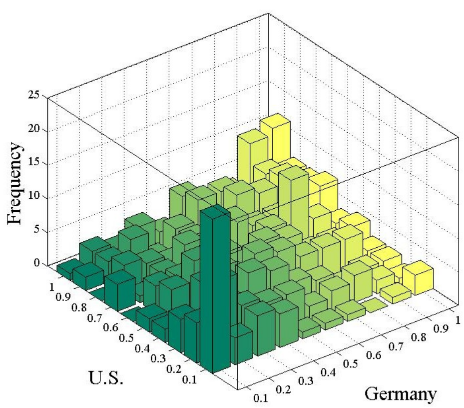
This figure plots the number of observations of X (U.S. returns) and Y (Canada, France, Germany or U.K. returns) that falls into bins defined by the quantiles of the returns. Specifically, the bins are defined by the intervals $Q_X(\tau_i) < X \leq Q_X(\tau_i)$ and $Q_Y(\tau_j) < Y \leq Q_Y(\tau_j)$, for $\tau_i, \tau_j = 0, 0.1, ..., 1$. The axis labels indicates the upper bound of the interval. For $\tau = 0$, we set $Q_X(0) = \min(X) - 1$ and $Q_Y(0) = \min(Y) - 1$.



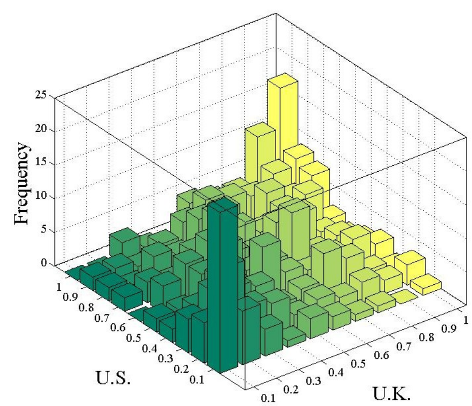
A. Canada



B. France



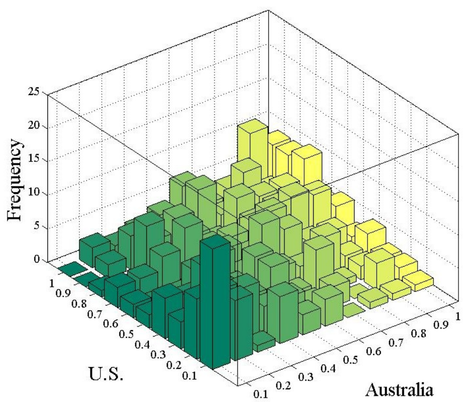
C. Germany



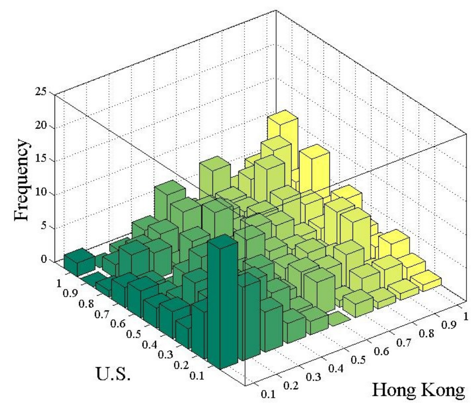
D. U.K.

Figure 2: **Bivariate Histograms of U.S. Returns with Australia, Hong Kong, Japan and Singapore.**

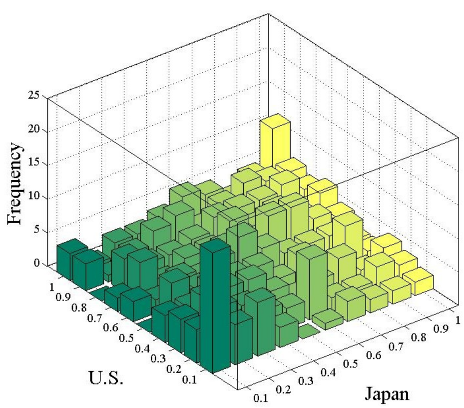
This figure plots the number of observations of X (U.S. returns) and Y (Australia, Hong Kong, Japan or Singapore returns) that falls into bins defined by the quantiles of the returns. Specifically, the bins are defined by the intervals $Q_X(\tau_i) < X \leq Q_X(\tau_i)$ and $Q_Y(\tau_j) < Y \leq Q_Y(\tau_j)$, for $\tau_i, \tau_j = 0, 0.1, ..., 1$. The axis labels indicates the upper bound of the interval. For $\tau = 0$, we set $Q_X(0) = \min(X) - 1$ and $Q_Y(0) = \min(Y) - 1$.



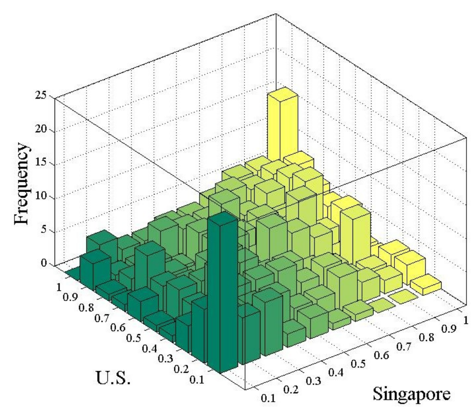
A. Australia



B. Hong Kong



C. Japan



D. Singapore

Figure 3: **Exceedence Correlation for Canada, France, Germany and U.K.**

This figure plots the correlations between U.S. returns and the returns to eight other countries using the method of exceedance of Longin and Solnik (2001). The correlations are to be interpreted as follows. At $\tau \leq 0.5$, this is the correlation between returns that are *less than or equal* to their respective $100\tau^{\text{th}}$ percentiles. At $\tau \geq 0.5$, this is the correlation between returns that are *greater than or equal* to their respective $100\tau^{\text{th}}$ percentiles.

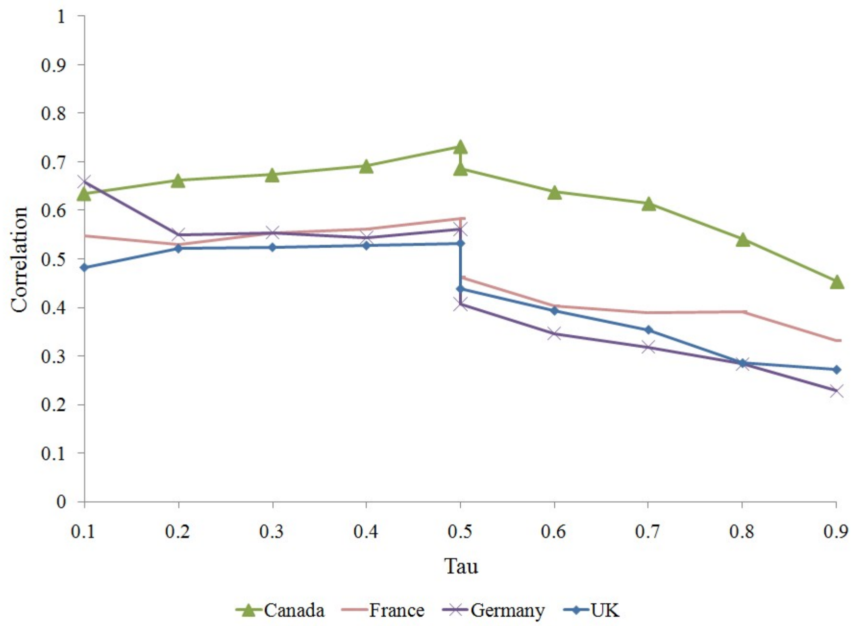


Figure 4: **Exceedence Correlation for Australia, Hong Kong, Japan and Singapore.**

This figure plots the correlations between U.S. returns and the returns to eight other countries using the method of exceedance of Longin and Solnik (2001). The correlations are to be interpreted as follows. At $\tau \leq 0.5$, this is the correlation between returns that are *less than or equal* to their respective $100\tau^{\text{th}}$ percentiles. At $\tau \geq 0.5$, this is the correlation between returns that are *greater than or equal* to their respective $100\tau^{\text{th}}$ percentiles.

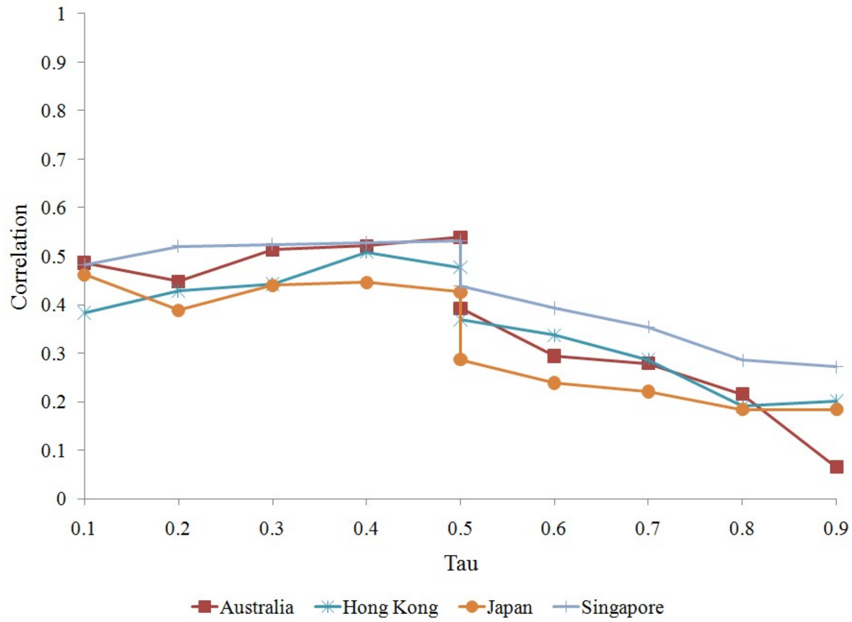
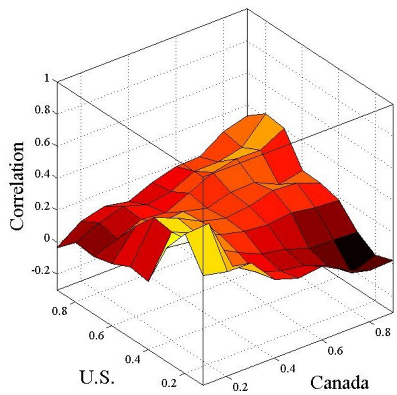
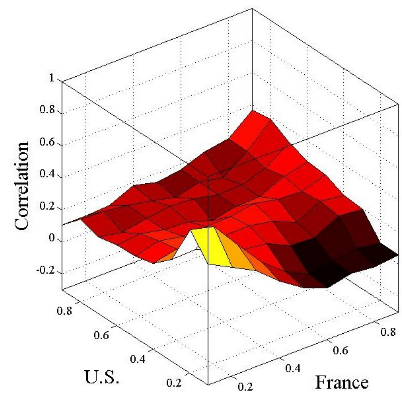


Figure 5: **Correlation of U.S. Returns Quantiles with Canada, France, Germany and U.K. Returns Quantiles: Gaussian Copula.**

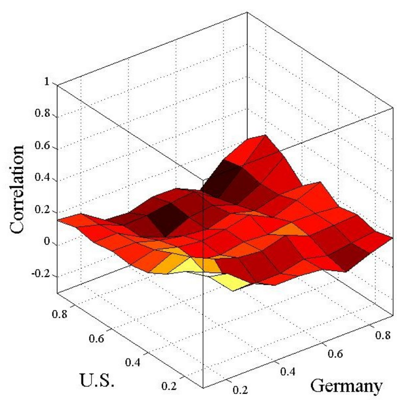
This figure plots the correlations between U.S. returns quantiles and returns quantiles of Canada, France, U.K. and Germany. The marginal distributions for the returns series are standard normal and the copula function is Gaussian. The equation estimated is based on (23).



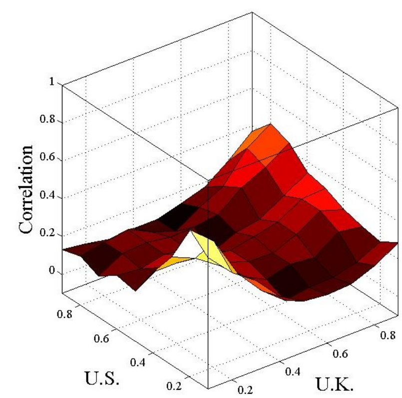
A. Canada



B. France



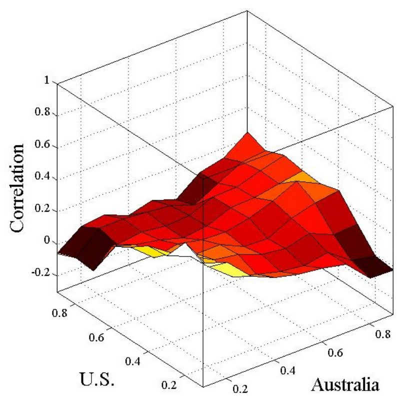
C. Germany



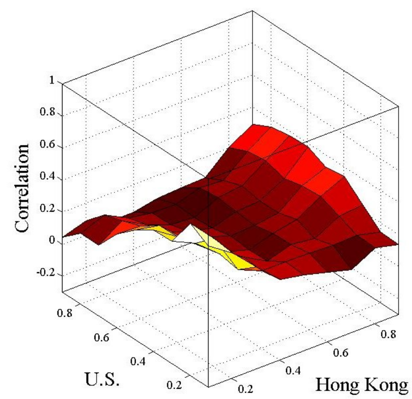
D. U.K.

Figure 6: **Correlation of U.S. Returns Quantiles with Australia, Hong Kong, Japan and Singapore Returns Quantiles: Gaussian Copula.**

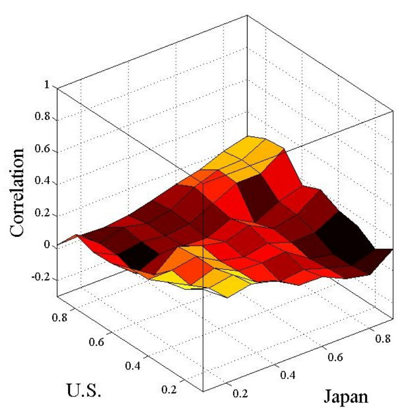
This figure plots the correlations between U.S. returns quantiles and returns quantiles of Australia, Hong Kong, Japan, Singapore. The marginal distributions for the returns series are standard normal and the copula function is Gaussian. The equation estimated is based on (23).



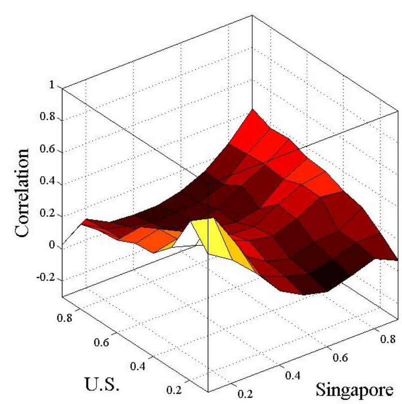
A. Australia



B. Hong Kong



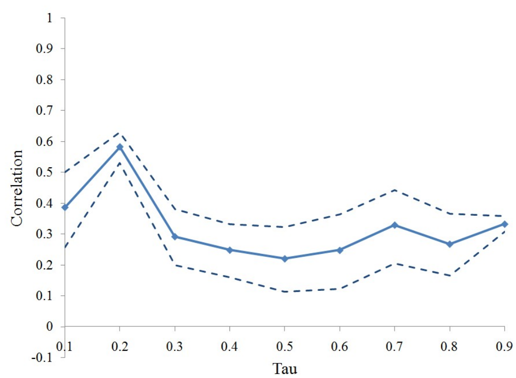
C. Japan



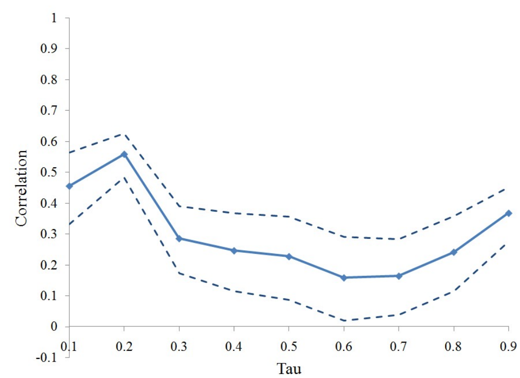
D. Singapore

Figure 7: **Main Diagonal Correlation of U.S. Returns Quantiles with Canada, France, Germany and U.K. Returns Quantiles: Gaussian Copula**

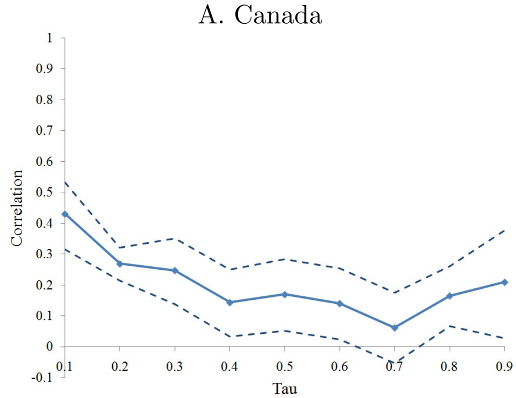
This figure plots the correlations of the main diagonals in Panels A and B in Figure 4. For Panel A and B in this figure, the marginal distributions for the returns series are standard normal and the copula function is Gaussian. For Panel C and D, the marginal distributions and the copula function are Student-t with ten degrees of freedom. The X-axis labels the quantiles of both U.S. and foreign stock returns. The dashed lines represent the 95 percent confidence bands constructed using the asymptotic standard errors calculated from (9).



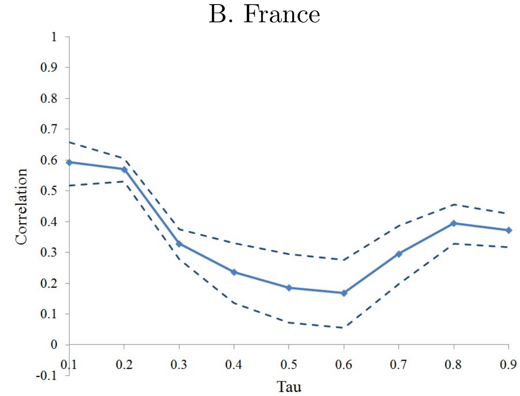
A. Canada



B. France



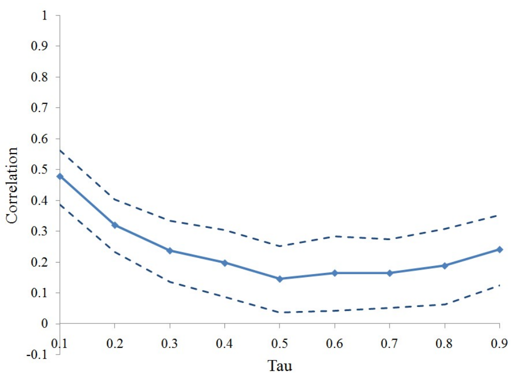
C. Germany



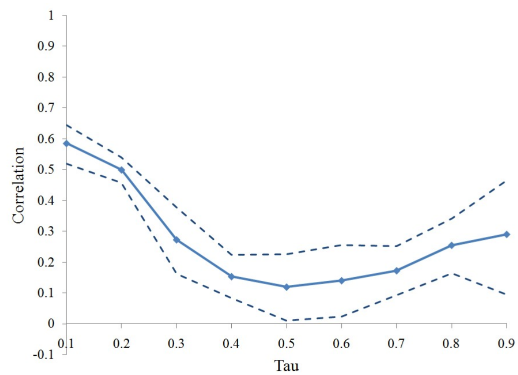
D. U.K.

Figure 8: **Main Diagonal Correlation of U.S. Returns Quantiles with Australia, Hong Kong, Japan and Singapore Returns Quantiles: Gaussian Copula**

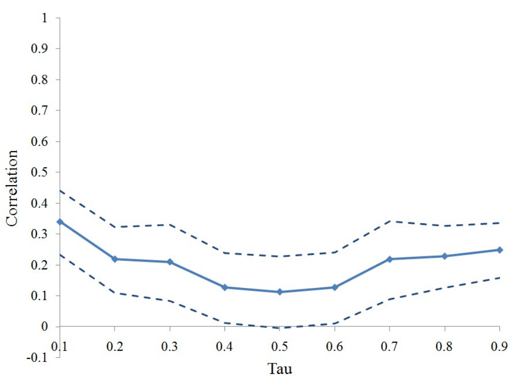
This figure plots the correlations of the main diagonals in Panels A and B in Figure 4. For Panel A and B in this figure, the marginal distributions for the returns series are standard normal and the copula function is Gaussian. For Panel C and D, the marginal distributions and the copula function are Student-t with ten degrees of freedom. The X-axis labels the quantiles of both U.S. and foreign stock returns. The dashed lines represent the 95 percent confidence bands constructed using the asymptotic standard errors calculated from (9).



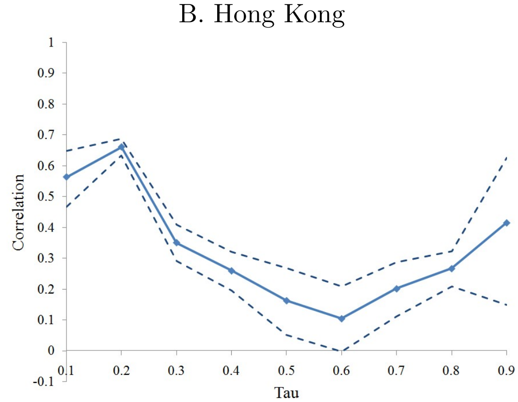
A. Australia



B. Hong Kong



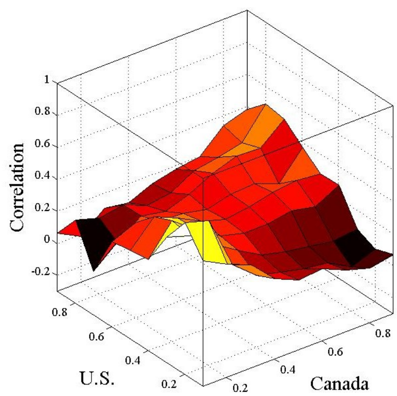
C. Japan



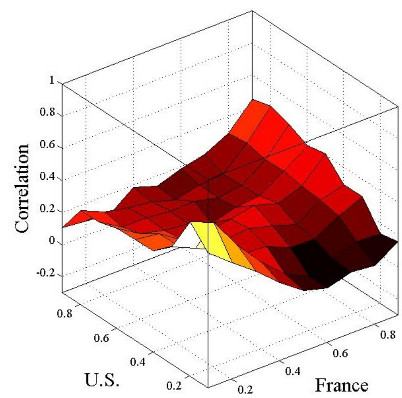
D. Singapore

Figure 9: **Correlation of U.S. Returns Quantiles with Canada, France, Germany and U.K. Returns Quantiles: Student-t_{10} Copula**

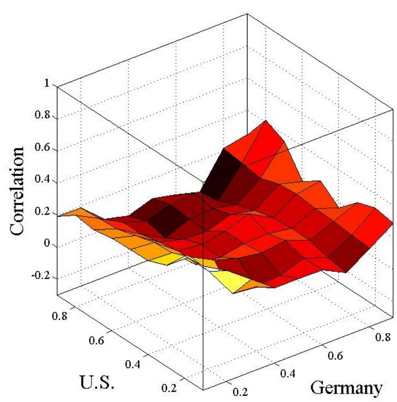
This figure plots the correlations between U.S. returns quantiles and returns quantiles of Canada, France, U.K. and Germany. The marginal distributions for the returns series and the copula function are all Student-t_{10}. The equation estimated is based on (24).



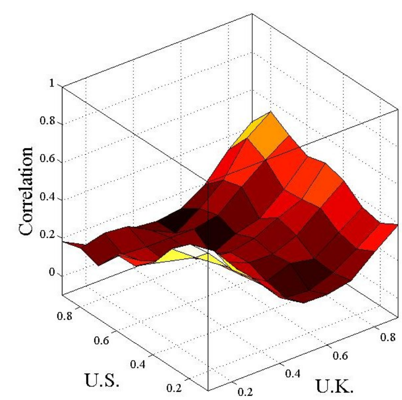
A. Canada



B. France



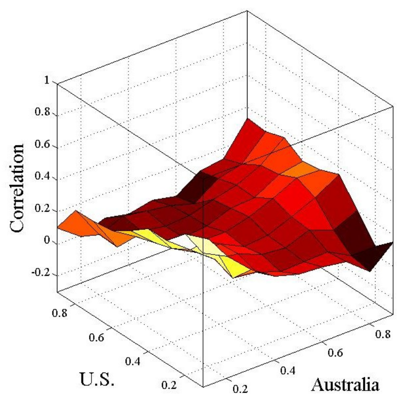
C. Germany



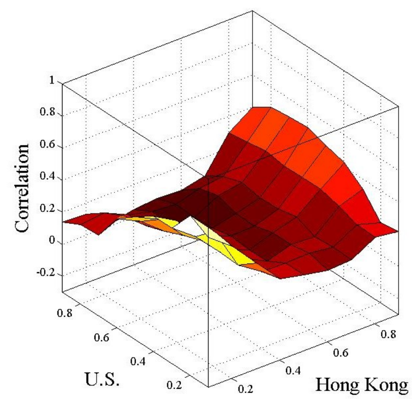
D. U.K.

Figure 10: **Correlation of U.S. Returns Quantiles with Australia, Hong Kong, Japan and Singapore Returns Quantiles: Student-t_{10} Copula.**

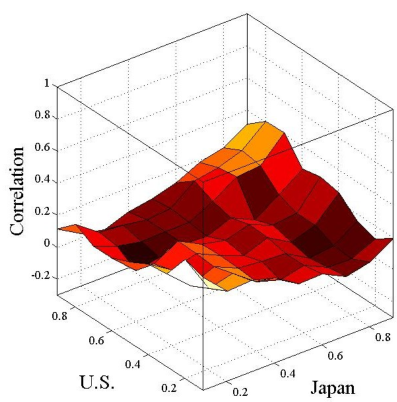
This figure plots the correlations between U.S. returns quantiles and returns quantiles of Australia, Hong Kong, Japan, Singapore. The marginal distributions for the returns series and the copula function are all Student-t_{10}. The equation estimated is based on (24).



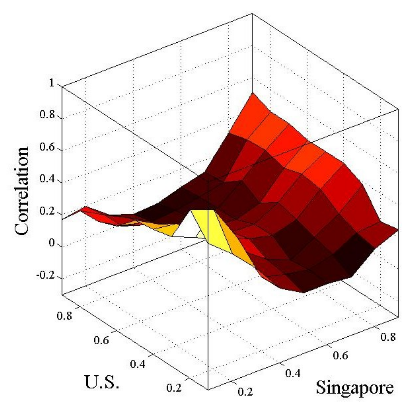
A. Australia



B. Hong Kong



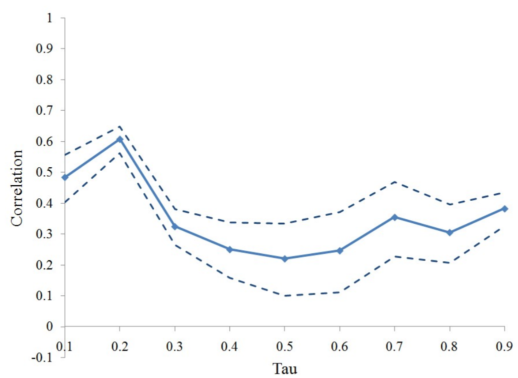
C. Japan



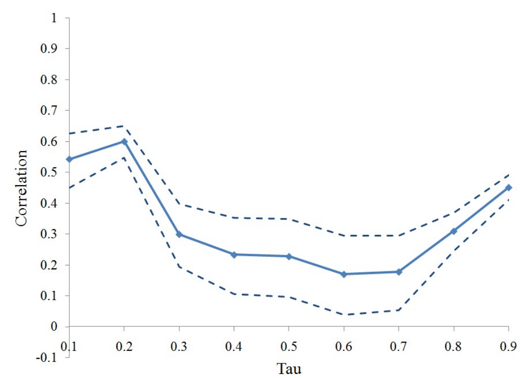
D. Singapore

Figure 11: **Main Diagonal Correlation of U.S. Returns Quantiles with Canada, France, Germany and U.K. Returns Quantiles: Student-t Copula**

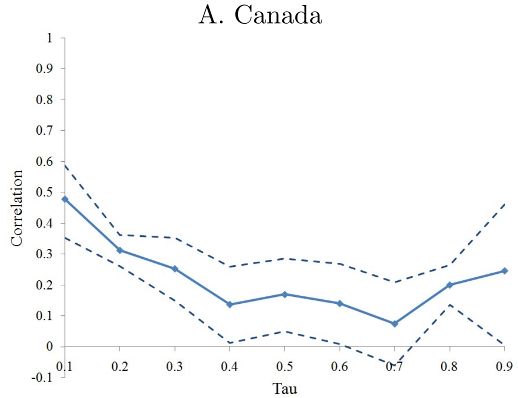
This figure plots the correlations of the main diagonals in Panels A and B in Figure 4. For Panel A and B in this figure, the marginal distributions for the returns series are standard normal and the copula function is Gaussian. For Panel C and D, the marginal distributions and the copula function are Student-t with ten degrees of freedom. The X-axis labels the quantiles of both U.S. and foreign stock returns. The dashed lines represent the 95 percent confidence bands constructed using the asymptotic standard errors calculated from (9).



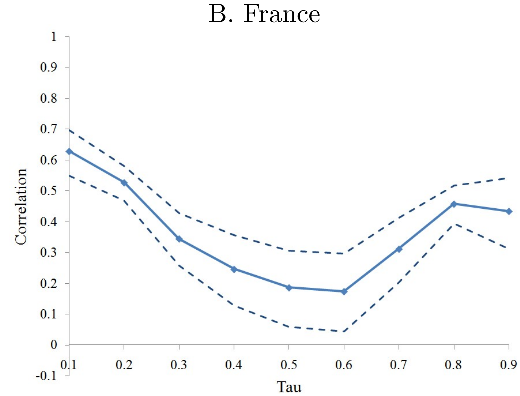
A. Canada



B. France



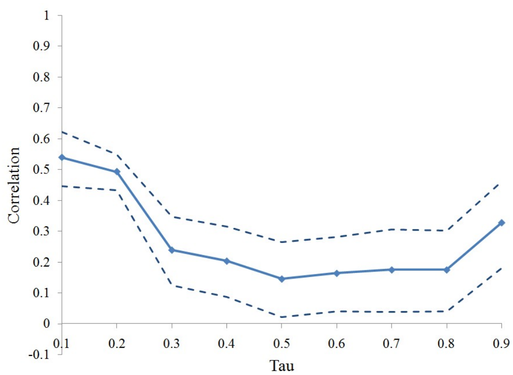
C. Germany



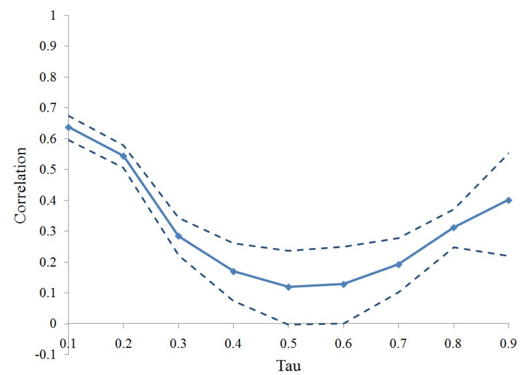
D. U.K.

Figure 12: **Main Diagonal Correlation of U.S. Returns Quantiles with Australia, Hong Kong, Japan and Singapore Returns Quantiles: Student-t Copula**

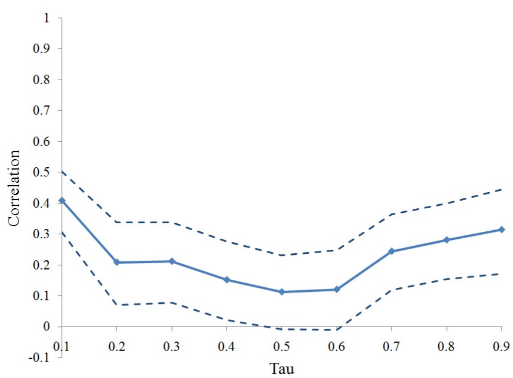
This figure plots the correlations of the main diagonals in Panels A and B in Figure 4. For Panel A and B in this figure, the marginal distributions for the returns series are standard normal and the copula function is Gaussian. For Panel C and D, the marginal distributions and the copula function are Student-t with ten degrees of freedom. The X-axis labels the quantiles of both U.S. and foreign stock returns. The dashed lines represent the 95 percent confidence bands constructed using the asymptotic standard errors calculated from (9).



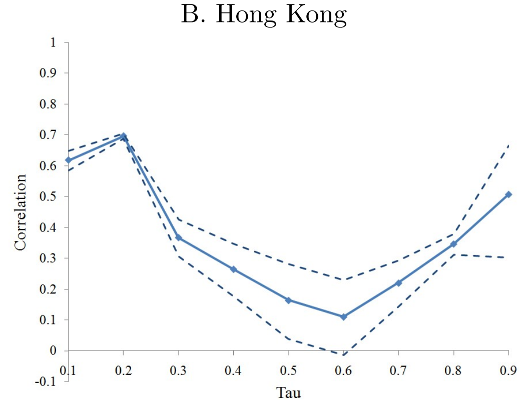
A. Australia



B. Hong Kong



C. Japan



D. Singapore

Table 1

Depth for the Eight Country Pairs.

This table estimates $Depth = |\rho_{0.1} - \rho_{0.5}| + |\rho_{0.9} - \rho_{0.5}|$, where ρ_τ is the correlation between the $100\tau^{\text{th}}$ percentile returns of the U.S. and the foreign country. The correlation is estimated from the Gaussian copula model based on (23).

	Canada	France	Germany	U.K.	Australia	Hong Kong	Japan	Singapore
Depth	0.2780	0.3691	0.3003	0.5938	0.4286	0.6367	0.3643	0.6525

CHAPTER 2

Modeling Quantile Dependence: A New Look at the Effects of Monetary Policy on Output Growth

Nicholas C.S. Sim*

ABSTRACT

Are the effects of monetary policy on output growth asymmetric? Does the same monetary policy stance influence output growth differently when output growth is high or low? These are questions that may be addressed together in a unified framework through our novel econometric methodology that models the quantile of output growth on the quantile of monetary policy shock, where restrictive (expansive) policies are represented by the left (right) tail of the policy shock distribution. We examine the asymptotic properties of the model, which combines series estimation and quantile regression methods. We find that the right tail of output growth is generally more sensitive to M1 money supply shocks while both tails of output growth are more sensitive to M2 money supply shocks than is the center of the distribution. When non-neutral, restrictive rather than expansive monetary policies have more influential effects on output growth.

JEL Classification: E5, C49, C50.
Key Words: Monetary Policy, Output Growth, Quantile Regression, Quantile Dependence, Series Estimation.

*Corresponding address: Department of Economics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467 USA. Tel: +1-617-552-6347, Fax: +1-617-552-2308, Email: nicholas.sim@bc.edu.

1 Introduction

The relationship between output and monetary policy has been a topic of considerable debate in the past decades. From the practical perspective, understanding the reaction of output to changes in the monetary instrument provides justification and guidance for the conduct of monetary policy. Since output and monetary policy could be nonlinearly related, the objective of the paper is to propose a new econometric methodology using quantile regression to uncover the nonlinearities that may exist in this relationship.

One of the earliest investigation along this line is to examine the asymmetric response of output growth to money supply shocks. Led by Cover (1992), it is now well-known that output growth reduced more strongly following a negative money supply shock than it increased following a positive money supply shock of the same magnitude. The econometric methodology employed by Cover and similar variations by other subsequent researchers[1] involves separating the estimated money supply shocks into positive and negative ones, then regressing output growth on these positive and negative shocks. Money supply shocks are deemed to have asymmetric effects on output growth if the coefficients on the positive and negative shocks are statistically distinguishable.

While convenient, Cover's approach implicitly assumes that the money supply shock has a zero population mean, which is necessary for identifying episodes of monetary contraction and expansion. Should this assumption of zero mean be violated, then some estimated money supply shocks may incorrectly identify the actual policy stance since identifying contractions or expansions solely rests on the signs of these shocks.

Therefore, this paper investigates the relationship between output growth and monetary policy in the tradition of Cover by proposing a new quantile regression methodology. This methodology relaxes the assumption that the policy shock has a zero population mean while still enabling us to uncover any potential asymmetric influence exerted by the policy stance on output. The insight of the quantile-based methodology comes from observing that the

[1]See, *inter alia*, DeLong and Summers (1988), Morgan (1993), Rhee and Rich (1995), Karras (1996), Senda (2001), and Parker and Rothman (2004).

quantile of the monetary shock contains information about the stance of monetary policy. In this respect, even if the true mean of the shock cannot be established, the new method still permits one to rank the policy environments on a spectrum ranging from the least expansive (or equivalently, the most restrictive) to the most expansive (or equivalently, the least restrictive) based on quantiles of the identified policy shocks.

To further elucidate the idea of ranking the policy environments, one may assert that *relative* to the median, a 10^{th} percentile policy shock reflects a more restrictive monetary policy stance while a 90^{th} percentile shock reflects a more expansive one. Using such a strategy, the econometric objective of the paper would then involve constructing a model of output growth as a function of the *quantile* of monetary policy shocks. Another potential dimension of nonlinearity that is unexplored in Cover's analysis is to allow output growth to react differently to the monetary policy stance contingent on whether output growth is high or low. This will enable policymakers to ascertain whether certain monetary policy objectives, as indexed by the quantiles of the monetary policy shocks, will be more effective in some economic states than others, as indexed by the quantiles of output growth. Therefore, a unified econometric framework that can simultaneously accommodate the two dimensions of nonlinearities will be one that models the *quantile* of output growth as a function of the *quantile* of monetary policy shocks.

It is crucial to clarify that the notion of expansive and restrictive policy reflects a ranking concept and does not imply that the policy is expansionary or contractionary,[2] so that for instance, an *expansive* policy may not necessarily be the same as an *expansionary* policy. While an expansionary environment is geared towards boosting output growth, an expansive environment is one where monetary policy is more favorable for output growth relative to another policy stance. This implies that even if the 10th and 20th percentile shocks are both contractionary policies, a fact which cannot be determined empirically,

[2]This interpretation of monetary policy stance bears a resemblance to the one employed by Conover et al. (1999) who examined the stock market's response to monetary policy. In their paper, the period when an interest rate cut that was preceded by a rate increase is considered an expansive policy environment. But the following period characterized by an unchanged interest rate is also considered as expansive even though there are no further rate cuts.

the 20th percentile shock is expansive relative to the 10th percentile shock, as the 10th percentile shock is purported to reduce output growth more aggressively than does the 20th percentile shock. Therefore, expansive and restrictive policies are only sensible if they are interpreted in relation to some reference policy position.

As mentioned, the econometric framework developed in this paper is based on the quantile regression paradigm. Typically, quantile regression focuses on modeling the conditional quantile of the dependent variable, as oppose to ordinary least squares regression that models its conditional mean. In this paper, the key departure from the standard quantile regression model is to allow the regressor to be itself a quantile. In order to construct the so-called *quantile-quantile* or QQ model, one must specify a system of equations having a recursive triangular structure. For a bivariate QQ model, two equations are required. The primary equation models the quantile of the dependent variable, i.e. the quantile of output growth, conditioned on a quantile regressor, i.e. the quantile of monetary policy shock, while the secondary equation is used to model the quantile of the regressor.

A similar recursive system of this nature was also examined in the seminal paper by Ma and Koenker (2006) for the parametric QQ framework. As opposed to Ma and Koenker, the main contribution of this paper is to allow the intercept and the slope parameters of the primary regression equation to be unknown functions of the model's innovation terms, which conditional on the information set map directly into the desired quantile of the dependent variable. We consider a linear triangular system of equations to be consistent with Cover's methodology, giving rise to a nonparametric model in the sense that the coefficients are nonparametric functions of the innovation terms. The estimation then uses power series expansion of the nonparametric intercept and slope parameters, employing truncation arguments similar to Newey (1997) while allowing the truncation parameter to grow with the sample size.[3] We show in Monte Carlo experiments that regressions based

[3] This is similar to having a finite-dimension parameter space whose dimension increases with the sample size, as first examined by Huber (1973) for M-estimation, then specialized to M-estimation with non-differentiable objective functions by He and Shao (2000). In addition, Zernov et al. (2009) examined the asymptotic properties of infinite dimensional quantile regressions. Their paper is similar as they also employed a truncation argument in their analysis.

on cubic or quartic expansions are able to estimate the dependence of quantiles well with reasonable bias properties and root-mean-squared errors.

The rest of the paper is organized as follows. Section 2 motivates why the new methodology may be more suitable for investigating nonlinearities in the money-output relationship. Section 3 reviews the standard quantile regression framework, discusses the quantile dependence model and motivates the power series estimation approach for modeling quantile dependence. Section 4 discusses the asymptotic properties of the estimator while Section 5 provides Monte Carlo evidence on the performance of the series regression. The empirical section is presented in Section 6 and Section 7 concludes.

2 Literature Review and Motivation

Over the past decades, the question of whether monetary policy is neutral had generated much interest as the conduct of monetary policy is only meaningful when it has the ability to influence real variables such as output and unemployment. By now, it is widely recognized that the money-output relationship is an asymmetric one, where the effects of monetary policy is akin to *pushing on the string* as it is a more effective tool for contracting than expanding the economy.

The pushing on the string hypothesis had gained much traction in empirical work following the econometric treatment by Cover (1992). Cover's methodology, which is popular due to its simplicity, first identifies monetary policy shocks as residuals from an autoregression involving a chosen monetary instrument. Following this preliminary regression step, output growth is regressed on the positive and negative monetary shocks as separate variables. Using various money supply processes to model monetary policy, Cover found that the contemporaneous effects of negative money supply shocks were generally statistically significant and the most effective but the positive money supply shocks, both contemporaneous and lags, were generally not. Furthermore, the coefficients on the contemporaneous negative shocks could be as large as 0.75, implying that a one percent contraction in money supply growth could cause an contemporaneous contraction of 0.75 percent in output growth. Moreover, the coefficients on the positive shocks were often close to zero.

The two-step procedure of Cover was also employed by DeLong and Summers (1988) and Lee (2000) to investigate the asymmetric effects of monetary policy during the pre-World War I and the interwar periods respectively. Both papers concurred with Cover's findings. Beyond the U.S., Karras (1996) focused on 18 European countries and uncovered similar asymmetric responses as well. The asymmetry is also robust to adopting a different monetary instrument other than money supply, for instance, to using the Federal funds rate by Morgan (1993) for the U.S. and the short-term interbank rate by Florio (2005) for Italy.[4]

In order to justify Cover's and indeed much of the literature's interpretation that monetary expansions and contractions can be identified by the signs of the estimated shocks, it is essential that the true mean of the monetary innovation is zero. However, ordinary least squares regression cannot separately identify the population mean of the innovation from the constant of the regression model. Should the population mean be nonzero, for instance, if the monetary policy shock has a negative population mean, then some estimated shocks with positive signs could in fact be negative innovations. Consequently, certain episodes of contractions could be misrepresented as expansions in this example.

While we cannot determine if the true population mean of the monetary shock is zero, there is some evidence that the distribution of the measures of monetary policy stance is skewed. We first examine the distribution of the monetary policy stance indicated by the Boschen and Mills (1995) index, which is updated by Weise (2007).[5] The index is a categorical variable taking on five possible values, [-2, -1, 0, 1, 2], with -2 representing the most contractionary stance where "monetary policy strongly emphasized reducing inflation" to 2 representing the most expansionary stance where "monetary policy strongly emphasized promoting real growth" (Boschen and Mills, 1995 p. 43). Zero, which indexes the months

[4]In addition to research along the lines of Cover, the relationship between output and money supply shocks may also exhibit other forms of nonlinearities. For instance, output responds to money supply shocks differently when the variance of the shocks is small versus when it is large (Ravn and Sola, 2004), or when the relationship exhibits regime switching behavior (Lo and Piger, 2003).

[5]The index begins from January 1968 and ends on December 2000, giving a total of 396 observations.

when monetary stance is deemed neutral, is the mode with 133 observations. For contractionary stance, there are 124 and 43 months indexed by -1 and -2 respectively, in contrast to the 74 and 22 months of expansion indexed by 1 and 2. Thus, the Boschen and Mills index suggests that the policy stance during this sample period is more often contractionary than expansionary. Unless serendipity has it that the policy shock has a zero population mean despite having a skewed distribution, misrepresentation of the true policy stance by the signs of estimated shocks may be nontrivial especially when using low frequency macroeconomic time series.[6]

Further evidence of this skewness may be found for the distribution of the shocks estimated as residuals of autoregressions and structural vector autoregressions (SVAR). Based on an autoregression, the skewness is 0.0897 for M1 money supply growth residuals and 0.4091 for M2 money supply growth residuals.[7] The positive skewness is also confirmed by the SVAR that uses the ordering of industrial production growth, consumer price inflation, gold price inflation, either M1 or M2 money supply growth, change in nonborrowed reserves and change in total reserves, where gold price inflation proxies for commodities price inflation as in Sims (1992). Using M1 money supply growth as the monetary instrument, the skewness of estimated shock is 0.4179 while using M2 money supply growth as the instrument, the skewness becomes 0.7311. The positive skewness found in the exercise suggests that the mean of the shock is greater than the median. If one believes that the median but not necessarily the mean is zero, then the population mean of the monetary policy stance is positive, which violates the implicit assumption of Cover's methodology.

The final piece of evidence follows from Conover et al. (1999), where we consider a proxy for monetary policy stance based on the change in the discount rate. A monetary environment is said to be restrictive if the discount rate breaks a "weakly" decreasing trend. For instance, starting from January 1970 to October 1970, the discount rate was 6

[6]Romer and Romer (2004) computed a monetary policy measure based on evidence from the FOMC minutes. However, to address the issue that the monetary action may be accommodating the future economic environment, they regress their raw monetary policy measure on a set of forecasted macroeconomic variables. This, however, presupposes that the true mean of the shock is zero.

[7]The sample starts from January 1970 to January 2009 and the lags are selected using a log-likelihood ratio test.

percent, but fell to 5.85 in November 1970. So, November 1970 is characterized as expansive. December 1970 is also expansive since the discount rate fell further to 5.52 percent. The expansive environment is terminated in July 1971 as the discount rate was increased from 4.75 to 4.88 percent.

Using this characterization of the monetary policy environment, we found that 58.00 percent of the monthly sample starting January 1970 and ending December 2002 was expansive. In addition, we repeated the exercise using the Federal funds target rate from January 1971 and to January 2009 and found that 56.55 percent of the sample was expansive. Taken together, this is evidence that the distribution of the monetary policy shock is likely to be skewed, which in turn raises questions on whether the population mean of the monetary policy shock is zero.

Therefore, this paper proposes interpreting the quantile of monetary policy shock as indexing the policy stance. While one cannot completely avoid misidentifying contractions or expansions since the population mean of the policy shock is unidentified, the estimated shock is nevertheless useful for indicating whether a policy stance is restrictive or expansive *relative* to a reference policy position. This motivates modeling output growth as a function of the quantiles of monetary policy shocks, which will be used to indicate the stance of the monetary policy.

3 The Model

This section will quickly review the standard quantile regression framework using linear models with additive errors for the purpose of illustration. It will then be followed by a discussion of the quantile-quantile model which is the main econometric contribution of this paper.

3.1 A Standard Quantile Regression Framework

Linear econometric models generally exhibit location or scale shift or both. The location shift model arises when the conditional quantiles are differentiated only by the intercept while the slope coefficients remain the same. The simplest special case of a location shift model with a single X_t regressor takes the form of

$$Y_t = \alpha_0 + \alpha_1 X_t + u_t, \tag{1}$$

where u_t is the innovation term. To appreciate what quantile regression does to (1), we may rearrange this equation as

$$Y_t = (\alpha_0 + u_t) + \alpha_1 X_t,$$

so that the model can be interpreted as having a *random-intercept* term $\alpha_0(u_t)$

$$Y_t = \alpha_0(u_t) + \alpha_1 X_t,$$

where $\alpha_0(u_t)$ expresses the intercept α_0 as a function of u_t. Hence, conditioned on X_t, the τ^{th} quantile of Y_t is obtained when u_t is the τ^{th} quantile also. Therefore, the conditional quantile of Y_t becomes

$$\begin{aligned} Q_Y(\tau|X_t) &= (\alpha_0 + F_u^{-1}(\tau)) + \alpha_1 X_t \\ &= \alpha_0(\tau) + \alpha_1 X_t, \end{aligned}$$

which demonstrates how the innovation term is associated with the conditional quantile of Y_t by shifting the intercept term. Another way to look at (1) is to define $u_t(\tau) = u_t - F_u^{-1}(\tau)$ so that the τ^{th} quantile of $u_t(\tau)$ is repositioned at zero. Substituting $u_t = u_t(\tau) + F_u^{-1}(\tau)$

into (1), we may express Y_t as

$$\begin{aligned} Y_t &= (\alpha_0 + F_u^{-1}(\tau)) + \alpha_1 X_t + u_t(\tau) \\ &= Q_Y(\tau|X_t) + u_t(\tau). \end{aligned}$$

This representation is convenient for elucidating what estimation in quantile regression entails. Here, estimating $Q_Y(\tau|X_t)$ involves searching for both the intercept and slope coefficients that set the τ^{th} quantile of $\hat{u}_t(\tau)$, the sample analog of the $u_t(\tau)$, to zero.[8] An extension of the location shift model is the location and scale shift model where both intercept and slope parameters may vary with u_t. The simplest special case of location and scale shift model with a single X_t regressor has the structure of

$$Y_t = \alpha_0 + (\alpha_1 + \delta u_t) X_t + u_t, \tag{2}$$

so that u_t acts as the shifter of both intercept and slope parameters. Equation (2) arises naturally as a model with conditional heteroskedasticity having an error term $u_t + \delta u_t X_t$. Given the monotonicity of Y_t with respect to u_t conditioning on X_t, this suggests that $Q_Y(\tau|X_t)$ can be obtained as $Q_Y(\tau|X_t) = \alpha_0 + (\alpha_1 + \delta F_u^{-1}(\tau))X_t + F_u^{-1}(\tau)$. As before, we substitute $u_t(\tau) = u_t + F_u^{-1}(\tau)$ into (2) so that (2) may be rewritten as

$$\begin{aligned} Y_t &= (\alpha_0 + F_u^{-1}(\tau)) + (\alpha_1 + \delta F_u^{-1}(\tau))X_t + u_t(\tau)(1 + \delta X_t) \\ &= Q_Y(\tau|X_t) + u_t(\tau|X_t). \end{aligned} \tag{3}$$

The computational work in estimating $Q_Y(\tau|X_t)$ then involves searching for the parameters that set the τ^{th} conditional quantile of $\hat{u}_t(\tau|X_t)$, the sample analog of $u_t(\tau|X_t)$, is zero.

[8] In practice, this computational problem translates into minimizing the quantile regression objective function proposed by Koenker and Bassett (1978), which in turn can be expressed as a linear programming problem. Given that p is the number of parameters, the linear programming solution will generate p zeros of $\hat{u}_t(\tau)$ so that the solution interpolates between these p observations. If nonlinear programming based on the interior point algorithm of Koenker and Park (1996) is used, then zero will also emerge as the τ^{th} quantile of $\hat{u}_t(\tau)$.

For the actual implementation, the conditional quantile is estimated by minimizing the first moment of the "check" function $\rho_\tau(u) = (\tau - \mathbb{I}(u < 0))u$, where $\mathbb{I}(\,.\,)$ is an indicator function. In the population context, the population parameters $\alpha(\tau)$ are those that minimize

$$\alpha(\tau) = \underset{\alpha}{\operatorname{argmin}}\, E[\rho_\tau(Y_t - \alpha' X_t)],$$

so that these parameters also set the population score function $E[X_t \psi_\tau(Y_t - \alpha' X_t)]$ to zero, where $\psi_\tau(u) = \tau - \mathbb{I}(u < 0)$. This follows from the fact that $E[\mathbb{I}(u_t(\tau) < 0)|X_t] = \tau$ given that zero is the τ^{th} quantile of $u_t(\tau)$, so that the population score function evaluated at $\alpha(\tau)$ is zero.

The actual estimation involves replacing the population quantile objective function with the sample analog

$$\hat{\alpha}(\tau) = \underset{\alpha}{\operatorname{argmin}}\, T^{-1} \sum_{t=1}^{T} \rho_\tau(Y_t - \alpha' X_t),$$

which is differentiable except at $Y_t = \alpha' X_t$, yielding the sample score function as

$$W(\hat{\alpha}(\tau)) = T^{-1} \sum_{t=1}^{T} X_t \psi_\tau(Y_t - \hat{\alpha}(\tau)' X_t), \tag{4}$$

which is zero except on set of measure zero.

3.2 A Quantile-Quantile (QQ) Framework

The previous subsection demonstrates how conditional quantiles are generated when the innovation term shifts the intercept and slope parameters. In this respect, a quantile regression framework is also a random coefficients framework, except the coefficients are influenced by a single innovation term.[9] Unlike the standard quantile regression framework, a quantile dependence framework allows the regressor to be a conditional quantile itself. Using the random-coefficient interpretation of quantile regression, the basic framework expressing the

[9]Technically speaking, we can say that the coefficients are *comonotone*, meaning that they are each monotonic in a common innovation term.

relationship between two quantiles, the quantiles of $Y_{1,t}$ and $Y_{2,t}$, can first be written as

$$Y_{1,t} = \alpha_0(w_t, u_t) + \alpha_1(w_t, u_t)' X_{1,t} + \alpha_2(w_t, u_t) Y_{2,t} \tag{5}$$

and

$$Y_{2,t} = \beta_0(w_t) + \beta_1(w_t)' X_{2,t}. \tag{6}$$

Here, w_t is the innovation of $Y_{2,t}$ so that conditioned on $X_{2,t}$, $Q_{Y_2}(\tau_2|X_{2,t})$ is obtained when w_t is $F_w^{-1}(\tau_2|X_{2,t})$. Similarly, assuming that w_t and u_t are independent, where u_t is the innovation of $Y_{1,t}$. Then, conditioning on $X_{1,t}$ and $Y_{2,t}$, $Q_{Y_1}(\tau_1|X_{1,t}, Y_{2,t})$ is obtained when u_t is $F_u^{-1}(\tau_1|X_{1,t}, Y_{2,t})$. This system of conditional quantile functions can be expressed as

$$Q_{Y_1}(\tau_1|X_{1,t}, Y_{2,t}) = \alpha_0(w_t, F_u^{-1}(\tau_1)) + \alpha_1(w_t, F_u^{-1}(\tau_1))' X_{1,t} + \alpha_2(w_t, F_u^{-1}(\tau_1)) Y_{2,t} \tag{7}$$

and

$$Q_{Y_2}(\tau_2|X_{2,t}) = \beta_0(F_w^{-1}(\tau_2)) + \beta_1(F_w^{-1}(\tau_2))' X_{2,t}, \tag{8}$$

where for the identification concern, $Y_{2,t}$ is identified by an exclusionary restriction whereby $X_{2,t}$ contains at least one variable excluded from $X_{1,t}$. The next step is to obtained the dependence between the quantiles which (7) has yet to express. This is obtained by setting w_t in (7) to its τ_2^{th} quantile so that $Y_{2,t}$ becomes its τ_2^{th} quantile, yielding

$$Q_{Y_1}(\tau_1|X_{1,t}, Q_{Y_2}(\tau_2|X_{2,t})) = \alpha_0(\tau_2, \tau_1) + \alpha_1(\tau_2, \tau_1)' X_{1,t} + \alpha_2(\tau_2, \tau_1) Q_{Y_2}(\tau_2|X_{2,t}), \tag{9}$$

where we denote $\alpha_i(\tau_2, \tau_1) \equiv \alpha_i(F_w^{-1}(\tau_2), F_u^{-1}(\tau_1))$. In order to obtain the QQ model as (9) expresses, a recursive system of structural equations such as (5) and (6) must be specified, so that $Y_{2,t}$ may influence $Y_{1,t}$ but not vice-versa. This setup is similar to the one examined by Ma and Koenker (2006) with two important differences. First, Ma and Koenker considered a nonlinear system while we specialize it to a linear model. Second, Ma and Koenker considered a fully parametric setup for both regressors as well as innovation terms. For the linear model, following Ma and Koenker would entail specifying how u_t and w_t enter the α parameters, which is avoided in our approach.

Two interesting facts emerge from the QQ model. First, (9) suggests that the influence by the quantile regressor may also come indirectly from α_0 and α_1 as these parameters may be functions of w_t as well. Second, if one wishes to obtain the coefficient on the quantile regressor, i.e. $\alpha_2(\tau_2, \tau_1)$, it does not matter if the regressor is actually $Q_{Y_2}(\tau_2|X_{2,t})$. This will be explained in the next section when we introduce a power series approach to estimate $\alpha_2(\tau_2, \tau_1)$ while allowing it to be the coefficient on $Y_{2,t}$ instead.

3.3 A Power Series Estimation Approach

For this approach, the α coefficients must be analytic in w_t so that there exists a power series expansion of α in w_t of all order. Without loss of generality, let the dimension of $X_{1,t}$ be one. To motivate the power series method, first rewrite (5) by adding and subtracting some terms

$$\begin{aligned}
Y_{1,t} =& \alpha_0(F_w^{-1}(\tau_2), u_t) + \alpha_1(F_w^{-1}(\tau_2), u_t)X_{1,t} + \alpha_2(F_w^{-1}(\tau_2), u_t)Y_{2,t} + [\alpha_0(w_t, u_t) - \alpha_0(F_w^{-1}(\tau_2), u_t)] \\
&+ [\alpha_1(w_t, u_t) - \alpha_1(F_w^{-1}(\tau_2), u_t)]X_{1,t} + [\alpha_2(w_t, u_t) - \alpha_2(F_w^{-1}(\tau_2), u_t)]Y_{2,t} \\
=& \alpha_0(F_w^{-1}(\tau_2), u_t) + \alpha_1(F_w^{-1}(\tau_2), u_t)X_{1,t} + \alpha_2(F_w^{-1}(\tau_2), u_t)Y_{2,t} + \Psi_t(w_t, u_t),
\end{aligned}$$

where $\Psi_t(w_t, u_t)$ is a nuisance quantity aggregating the bracketed terms. Insofar $\Psi_t(w_t, u_t)$ can be controlled in the regression, we may estimate the conditional quantile function of Y_1 as

$$\hat{Q}_{Y_1}(\tau_1|X_{1,t}, Y_{2,t}) = \hat{\alpha}_0(\tau_2, \tau_1) + \hat{\alpha}_1(\tau_2, \tau_1)X_{1,t} + \hat{\alpha}_2(\tau_2, \tau_1)Y_{2,t} + \widehat{\Psi},$$

where $\widehat{\Psi}$ controls for Ψ. Since $\Psi_t(w_t, u_t)$ contains the difference $\alpha_i(w_t, u_t) - \alpha_i(F_w^{-1}(\tau_2), u_t)$, one way to control $\Psi_t(w_t, u_t)$ is to employ a power series expansion of $\alpha_i(w_t, u_t)$ in the first argument around $F_w^{-1}(\tau_2)$. Then using the fact that $w_t(\tau_2) = w_t - F_w^{-1}(\tau_2)$, the expansion yields

$$\Psi_t(w_t(\tau_2), u_t) = \lim_{M\to\infty} \sum_{k=1}^{M} \left[\frac{\alpha_{0,k}(u_t)}{k!} w_t(\tau_2)^k + \frac{\alpha_{1,k}(u_t)}{k!} w_t(\tau_2)^k X_{1,t} + \frac{\alpha_{2,k}(u_t)}{k!} w_t(\tau_2)^k Y_{2,t} \right]$$

where $\alpha_{i,k}(u_t)$ is the k-derivative of α_i around $F_w^{-1}(\tau_2)$ so that the only variable remaining in the derivatives of α_i is u_t. For feasible estimation, we truncate the infinite series, utilizing the regression function

$$\begin{aligned} & H_{Y_1}(w_t(\tau_2), u_t; \alpha, \varphi) \\ =& \alpha_0(\tau_2, u_t) + \alpha_1(\tau_2, u_t) X_1 + \alpha_2(\tau_2, u_t) Y_{2,t} \\ & + \varphi_{\tau_2,0,K_0}(u_t)' P_{0,K_0}(w_t(\tau_2)) + \varphi_{\tau_2,1,K_1}(u_t)' P_{1,K_1}(w_t(\tau_2)) + \varphi_{\tau_2,2,K_2}(u_t)' P_{2,K_2}(w_t(\tau_2)), \end{aligned}$$

where $P_{i,K_i}(w(\tau_2))$ is the K_i polynomial in $w_t(\tau_2)$ while $\varphi_{\tau_2,1,K_1}$ is a parameter vector associated with the derivatives of α_i up to the order K_i where the expansion of α_i is taken around $F_w^{-1}(\tau_2)$. For instance, the parameter vector is $\varphi_{\tau_2,1,K_1} = (\varphi_{\tau_2,1,1}, \varphi_{\tau_2,1,2}, \ldots, \varphi_{\tau_2,1,K_1})$ while the polynomial contains

$$P_{1,K_1}(w_t(\tau_2)) = (w_t(\tau_2), w_t(\tau_2)^2/2!, \ldots, w_t(\tau_2)^{K_1}/K_1!) X_{1,t},$$

so that

$$\varphi'_{\tau_2,1,K_1} P_{1,K_1}(w_t(\tau_2)) = \varphi_{\tau_2,1,1} w_t(\tau_2) X_{1,t} + \varphi_{\tau_2,1,2} \frac{w_t(\tau_2)^2}{2!} X_{1,t} + \ldots + \varphi_{\tau_2,1,K_1} \frac{w_t(\tau_2)^{K_1}}{K_1!} X_{1,t}.$$

Since we have expanded w_t around $F_w^{-1}(\tau_2)$, the only innovation term remaining in φ_{τ_2,i,K_i} is u_t. In other words, the expansion separates u_t from w_t in α so that after controlling for $w_t(\tau_2)$ in the nuisance term, all the parameters will be functions of u_t alone. Hence α, with its w-argument now anchored at $F_w^{-1}(\tau_2)$, can be estimated using the standard quantile regression framework treating u_t as the only source of innovation.

With the truncation, $H_{Y_1}(w_t(\tau_2), F_u^{-1}(\tau_1); \alpha, \varphi)$ may be used to approximate the conditional quantile of Y_1. Hence, the difference between the true and the approximate conditional quantile of Y_1 is $\Gamma_{0,t} + \Gamma_{1,t} + \Gamma_{2,t}$, where $\Gamma_{i,t}$ defines a remainder term associated with the series expansion of α_i. To consistently estimate the conditional quantile function, it is imperative for $\Gamma_{i,t}$ to disappear asymptotically as the number of approximating terms K_i in the polynomial grow with the sample size. This issue is related to estimating a model with an increasing parameter dimension, first considered by Huber (1973) and recently generalized by He and Shao (2000) to M-estimation where discontinuities in the score function are permitted. Similar to Huber, Newey (1997) examined the conditions for consistency and asymptotic normality for series estimation in the ordinary least squares framework, and unlike Huber, he accounted for the remainder term that is introduced by using the method of truncation.

The two-step estimation procedure is summarized as follows:

1. *Obtain $\hat{w}(\tau_2)$ as the residual from the τ_2 quantile regression of $Y_{2,t}$ on $X_{2,t}$ by estimating*
$$\hat{\beta}(\tau_2) = \underset{\beta}{\operatorname{argmin}}\, T^{-1} \sum_{t=1}^{T} \rho_{\tau_2}(Y_{2,t} - \beta' X_{2,t}).$$

2. *Using $\hat{w}(\tau_2)$, estimate*
$$(\hat{\alpha}(\tau_2, \tau_1), \hat{\varphi}_{\tau_2,K}(\tau_1)) = \underset{\alpha}{\operatorname{argmin}}\, T^{-1} \sum_{t=1}^{T} \rho_{\tau_1}(Y_{1,t} - H_{Y_1}(\hat{w}_t(\tau_2), u_t; \alpha, \varphi)).$$

Inference will be based on the asymptotic distribution which is derived in the next section.

4 Asymptotic Theory

We first define some notation. Define $\alpha_i(\tau_2, u) \equiv \alpha_{i,\tau_2}(u)$ for $i = 0, 1, 2$ so that $\boldsymbol{\alpha}_{\tau_2}(u) = [\alpha_{0,\tau_2}(u)\ \alpha_{1,\tau_2}(u)\ \alpha(u)_{2,\tau_2}]'$. Let the original information vector at time t, not including the polynomials from the series expansion, be $\mathbf{X}_{1,t} = [1\ X_{1,t}'\ Y_{2,t}]$. The design matrix is thus a $T \times p$ matrix $\mathbf{X}_1$. Without loss of generality, we consider a one-dimensional X_1 so that $p = 3$.

As a result of the series expansion, the additional regressors will form a $T \times \bar{\lambda}$ matrix of polynomials $\mathbb{P}_{\bar{\lambda}}(w(\tau_2)) = [P_{0,K_0}(w(\tau_2))\ P_{1,K_1}(w(\tau_2))\ P_{2,K_2}(w(\tau_2))]$, where the number of terms in the polynomials is $\bar{\lambda}(T) = K_0(T) + K_1(T) + K_2(T)$. The notation for the polynomials makes it explicit that the polynomials are functions of $w(\tau_2)$. The design matrix will then include the original regressors $\mathbf{X}_1$ and polynomials $\mathbb{P}_{\bar{\lambda}}(w(\tau_2))$ to form $\mathbb{X}_1(w(\tau_2)) = [\mathbf{X}_1\ \mathbb{P}_{\bar{\lambda}}(w(\tau_2))]$, which has $\lambda = p + \bar{\lambda}$ dimensions. For feasible estimation, $w(\tau_2)$ must be replaced with its fitted counterpart $\hat{w}(\tau_2)$ estimated from a preliminary step. Therefore, the actual regression employs the polynomials $\hat{\mathbb{P}} \equiv \mathbb{P}_{\bar{\lambda}}(\hat{w}(\tau_2))$ and thus the design matrix $\hat{\mathbb{X}}_1 \equiv \mathbb{X}_1(\hat{w}(\tau_2))$. With appropriate regularity conditions, we have $\hat{w}(\tau_2) = w(\tau_2) + o_p(1)$ so that $\hat{\mathbb{X}}_1 = \mathbb{X}_1 + o_p(1)$, which is true as long as the estimated parameters in the first-step regression are consistent, i.e. $\hat{\gamma}(\tau_2) = \gamma(\tau_2) + o_p(1)$.

Let the coefficients on the polynomials be $\boldsymbol{\varphi}_{\tau_2} = [\varphi'_{0,\tau_2}\ \varphi'_{1,\tau_2}\ \varphi'_{2,\tau_2}]'$, bearing in mind that they are functions of the $Y_{1,t}$ innovation term u_t. Hence, the combined parameter vector is a λ-dimension vector $\boldsymbol{\theta}_{\tau_2} = [\boldsymbol{\alpha}'_{\tau_2}\ \boldsymbol{\varphi}'_{\tau_2}]'$. Since truncation of the infinite series is employed, doing so introduces a remainder term associated with each of the α parameters that are expanded. Suppressing the arguments, the remainder term is a multiplication of a $T \times 3$ vector $\Gamma = [\Gamma_0\ \Gamma_1\ \Gamma_2]$ and a 3×1 vector of ones denoted by i_3, where Γ_i is a $T \times 1$ vector of the remainder term associated with estimating α_i. In period t notation, Γ_t is a 3×1 vector.

Define $u_t(\tau_1) = Y_{1,t} - Q_{Y_1}(\tau_1|\mathbb{X}_{1,t})$ so that $Q_{u(\tau_1)}(\tau_1|\mathbb{X}_{1,t}) = 0$. The model, as we recall, is a system of equations comprising of

$$Y_{1,t} = \boldsymbol{\theta}_{\tau_2}(\tau_1)'\mathbb{X}_{1,t} + \Gamma'_t i_3 + u_t(\tau_1)$$

$$Y_{2,t} = \gamma(\tau_2)' X_{2,t} + w_t(\tau_2)$$

where $\Gamma'_t i_3 = Q_{Y_1}(\tau_1|\mathbb{X}_{1,t}) - \boldsymbol{\theta}_{\tau_2}(\tau_1)'\mathbb{X}_{1,t}$ reflects the fact that $\boldsymbol{\theta}_{\tau_2}(\tau_1)'\mathbb{X}_{1,t}$ only approximates the quantile of $Y_{1,t}$. Since $\mathbb{X}_{1,t}$ is unknown, feasible estimation requires replacing $\mathbb{X}_{1,t}$ with $\hat{\mathbb{X}}_{1,t}$ after obtaining $\hat{w}_t(\tau_2)$ from the second equation. This introduces a generated regressor problem that will have implications for inference. That using generated regressors, i.e. $\hat{\mathbb{X}}_{1,t}$, may give rise to issues for inference comes from the fact that we are actually estimating

$$Y_{1,t} = \boldsymbol{\theta}_{\tau_2}(\tau_1)'\hat{\mathbb{X}}_{1,t} + \Gamma_t' i_3 + \underbrace{\boldsymbol{\theta}_{\tau_2}(\tau_1)'(\mathbb{X}_{1,t} - \hat{\mathbb{X}}_{1,t})}_{\Phi_t} + u_t(\tau_1)$$

where Φ_t is a term introduced by using $\hat{\mathbb{X}}_{1,t}$. If $\mathbb{X}_{1,t}$ is consistently estimated by $\hat{\mathbb{X}}_{1,t}$, the consistency of $\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1)$ will usually not be compromised. Nevertheless, using $\hat{\mathbb{X}}_{1,t}$ introduces an additional source of impreciseness that will lead to increasing the standard error of $\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1)$. This claim will be verified later in the section.

We now examine the large sample properties of $\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1)$ and derive its asymptotic distribution. The large sample theory utilizes the following assumptions:

A1. Let $\{Y_{1,t}, t \geq 1\}$ and $\{Y_{2,t}, t \geq 1\}$ be sequences of independent random variables defined on the probability space $(\Omega_1, \mathcal{F}_{1,t}, P_1)$ and $(\Omega_2, \mathcal{F}_{2,t}, P_2)$ having a nondecreasing sub σ-fields $\mathcal{F}_{i,0} \subset \mathcal{F}_{i,1} \subset \ldots \subset \mathcal{F}_i$ for $i = 1, 2$, where $\mathcal{F}_{i,0}$ is the trivial σ-field, $\mathcal{F}_{1,t-1} = \sigma(\{\mathbb{X}_{1,j}\}_{j=0}^{t}, \{Y_{1,k}\}_{k=0}^{t-1})$ and $\mathcal{F}_{2,t-1} = \sigma(\{X_{2,j}\}_{j=0}^{t}, \{Y_{2,k}\}_{k=0}^{t-1})$.

A2. The λ-dimensional parameter space $\boldsymbol{\Theta}$ is compact.

A3. There exists a constant $s > (1-a)/2a$ and a sequence of numbers $K\lambda^{-s}$ such that $\max_i \max_t |\Gamma_{i,t}| < K\lambda^{-s}$, where K is some constant and $\lambda = O(T^a)$ for $a \in (0, 1/2)$.

A4. The cumulative distribution function of $u(\tau_1)$, denoted by F, is continuously differentiable with density f that is bounded above by a constant $C_f^{\max}$ and bounded below by a constant $C_f^{\min}$ at $u(\tau_1) = 0$.

A5. Define $\hat{D}_{1,T} = \hat{\mathbb{X}}_1'\hat{\mathbb{X}}_1/T$ and $D_{1,T} = \mathbb{X}_1'\mathbb{X}_1/T$, where the latter converges to a positive definite matrix D. In addition, for each $\tilde{\boldsymbol{\theta}}_{\tau_2} \in \boldsymbol{\Theta}$, define $\hat{Q}_{1,T}(\tilde{\boldsymbol{\theta}}_{\tau_2}) = \hat{\mathbb{X}}_1'\hat{\mathbb{F}}(\tilde{\boldsymbol{\theta}}_{\tau_2})\hat{\mathbb{X}}_1/T$ and $Q_{1,T}(\tilde{\boldsymbol{\theta}}_{\tau_2}) = \mathbb{X}_1'\mathbb{F}(\tilde{\boldsymbol{\theta}}_{\tau_2})\mathbb{X}_1/T$, where the latter converges to a positive definite matrix

$Q_1(\tilde{\boldsymbol{\theta}}_{\tau_2})$. The minimum eigenvalues of $Q_{1,T}(\tilde{\boldsymbol{\theta}}_{\tau_2})$ and $Q_1(\tilde{\boldsymbol{\theta}}_{\tau_2})$, i.e. $K_{\min}(Q_{1,T}(\tilde{\boldsymbol{\theta}}_{\tau_2}))$ and $K_{\min}(Q_1(\tilde{\boldsymbol{\theta}}_{\tau_2}))$, are bounded away from zero for all T and uniformly in $\tilde{\boldsymbol{\theta}}_{\tau_2} \in \boldsymbol{\Theta}$. $\mathbb{F}(\tilde{\boldsymbol{\theta}})$ is a diagonal matrix with t element $f(\eta \Upsilon_t(\tilde{\boldsymbol{\theta}}_{\tau_2}))$ and $\hat{\mathbb{F}}(\tilde{\boldsymbol{\theta}}_{\tau_2})$ is a diagonal matrix with t element $f(\eta \hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}}_{\tau_2}))$, where $0 < \eta < 1$ and $\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}}_{\tau_2}) = (\tilde{\boldsymbol{\theta}}_{\tau_2} - \boldsymbol{\theta}_{\tau_2}(\tau_1))'\hat{\mathbb{X}}_{1,t} + \boldsymbol{\theta}_{\tau_2}(\tau_1)'(\hat{\mathbb{X}}_{1,t} - \mathbb{X}_{1,t}) - \Gamma_t' i_3$ and $\Upsilon_t(\tilde{\boldsymbol{\theta}}_{\tau_2}) = (\tilde{\boldsymbol{\theta}}_{\tau_2} - \boldsymbol{\theta}_{\tau_2}(\tau_1))'\mathbb{X}_{1,t} - \Gamma_t' i_3$.

A6. Let the j element of $\mathbb{X}_{1,t}$ be $\mathbb{X}_{1,t}^{(j)}$. Then, there exists a constant Δ such that $E|\mathbb{X}_{1,t}^{(j)}|^3 \leq \Delta < \infty$ for all t and $j = 1, \ldots, p$.

A7. $\hat{\gamma}(\tau_2)$ is a consistent estimator of $\gamma(\tau_2)$.

By assuming that $\mathbb{X}_{1,t}$ is $\mathcal{F}_{t,1}$-measurable, A1 implicitly captures the fact that conditioning on $\mathcal{F}_{t,1}$ implies conditioning on w_t also. The independence assumption in A1, while is a strong one, is reasonable for our empirical objective as the innovation terms are interpreted as unexpected shocks to output growth and monetary stance. Assumption A3 is required to bound the remainder term, which is also required in Newey (1997). In particular, the parameter a in A3 controls for the rate in which the dimension may increase. It also controls the speed in which the remainder term must converge to zero. In the extreme case where a tends to zero, the remainder term converges to zero extremely quickly, so that the convergence of $\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1)$ will tend to the rate of root-T.

Assumption A4 requires the density function to be bounded above and at $u(\tau_1) = 0$, the density must be bounded above zero. In expressing A4, $u(\tau_1)$ is assumed to be homoskedastic, although the case for conditional heteroskedasticity can be easily extended. Assumption A5 and A6 impose the existence of certain moments. Assumption A7 implies that $\hat{w}_t(\tau_2)$ is a consistent estimator of $w_t(\tau_2)$ as $\hat{w}_t(\tau_2) - w_t(\tau_2) = -(\hat{\gamma}(\tau_2) - \gamma(\tau_2))'X_{2,t}$ and $\hat{\gamma}(\tau_2)$ converges to $\gamma(\tau_2)$ in probability by A7.

We first proceed by establishing consistency through Proposition 1, then the rate of convergence through Proposition 2. The rate of convergence, not surprisingly, is slower than root-T given the increasing dimension of the design matrix. From Proposition 2, we may derive the linear representation for $\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1)$, which may be used to obtain the asymptotic distribution. The technical details of the proofs are relegated to the appendix.

Proposition 1. *(Consistency) Under A1-A7,* $\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1) - \boldsymbol{\theta}_{\tau_2}(\tau_1) = o_p(1)$.

That $\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1)$ converges at a rate slower than root-T has been established previously for ordinary least squares regression. This can also be established for quantile regression as Proposition 2 claims.

Proposition 2. *(Convergence Rate) Under A1-A7,* $\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1) - \boldsymbol{\theta}_{\tau_2}(\tau_1) = O_p(\sqrt{\lambda/T})$.

The rate of convergence of $\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1) - \boldsymbol{\theta}_{\tau_2}(\tau_1)$ may be inferred from Proposition 2 as $O_p(T^{-(1-a)/2})$. The parameter a clearly demonstrates the tension between the remainder term and the speed of convergence. If the remainder term converges slowly, as it is the case if a is close to $1/2$, then convergence to a limiting distribution will also be slow. If a is close to zero, then this convergence rate will be near root-T. Note that we obtain the same range for a as compared to Zernov et al. (2009), where they also examined the asymptotic properties of quantile regression with infinite dimension using similar truncation methods. In their paper, shrinking the remainder to zero requires the dimension of the regressors to grow at a polynomial rate controlled by $a \in (0, 1/2)$, which is permitted by A3 in our paper.

In quantile regression, the linear (Bahadur) representation is commonly used to verify the conditions for Central Limit Theorem and to derive the formula for the asymptotic covariance matrix. This representation has been derived as part of the proof of Proposition 2 as

$$\begin{aligned}
&\sqrt{T}(\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1) - \boldsymbol{\theta}_{\tau_2}(\tau_1)) \\
=&Q_1^{-1}T^{-1/2}\sum_{t=1}^{T}\mathbb{X}_{1,t}\psi_{\tau_1}\left(Y_t - \boldsymbol{\theta}_{\tau_2}(\tau_1)'\mathbb{X}_{1,t}\right) \\
&- Q_1^{-1}T^{-1}\sum_{t=1}^{T}E[f_t(\eta\Upsilon_t(\boldsymbol{\theta}_{\tau_2}(\tau_1))\mathbb{X}_{1,t}\boldsymbol{\theta}_{\tau_2}(\tau_1)']\mathbb{X}_{1,t}^{d}X_{2,t}'\sqrt{T}(\hat{\gamma}(\tau_2) - \gamma(\tau_2)) \\
&+ o_p(1).
\end{aligned} \tag{10}$$

Let γ be a p_2 vector. In addition, let $Q_{3,T} = T^{-1}\sum_{t=1}^{T} E[f_t(\eta\Upsilon(\boldsymbol{\theta}_{\tau_2}(\tau_1)))\mathbb{X}_{1,t}\boldsymbol{\theta}_{\tau_2}(\tau_1)']\mathbb{X}_{1,t}^{d}X_{2,t}'$ be a $\lambda \times p_2$ matrix that converges to Q_3 with full column rank. If $\sqrt{T}(\hat{\gamma}(\tau_2) - \gamma(\tau_2))$ is asymptotically normal under appropriate moment conditions, the asymptotic distribution of $\sqrt{T}(\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1) - \boldsymbol{\theta}(\tau_1))$ depends on the asymptotic distribution of $T^{-1/2}\sum_{t=1}^{T}\mathbb{X}_{1,t}\psi_{\tau_1}(Y_t - \boldsymbol{\theta}_{\tau_2}(\tau_1)'\mathbb{X}_{1,t})$ as well. Let V_K be the asymptotic covariance matrix of $\sqrt{T}(\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1) - \boldsymbol{\theta}_{\tau_2}(\tau_1))$, which may be derived from (10) as

$$\Omega_{\hat{\gamma}(\tau_2)} = \tau_1(1-\tau_1)Q_1^{-1}D_1Q_1^{-1} + Q_1^{-1}Q_3\Omega_{\hat{\gamma}(\tau_2)}Q_3'Q_1^{-1}, \tag{11}$$

where $\Omega_{\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1)}$ is a $p_2 \times p_2$ asymptotic covariance matrix of $\sqrt{T}(\hat{\gamma}(\tau_2) - \gamma(\tau_2))$, which also can be expressed as $\Omega_{\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1)} = \tau_2(1-\tau_2)Q_2^{-1}D_2Q_2^{-1}$, where Q_2 is the $p_2 \times p_2$ limit matrix of $Q_{2,T} = T^{-1}\sum_{t=1}^{T} g(G_{w,t}^{-1}(\tau_2))X_{2,t}X_{2,t}'$ and D_2 is the $p_2 \times p_2$ limit matrix of $D_{2,T} = T^{-1}\sum_{t=1}^{T} X_{2,t}X_{2,t}'$.

Proposition 3. *(Asymptotic Normality) Under A1-A7,* $\sqrt{T}\Omega_{\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1)}^{-1/2}(\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1) - \boldsymbol{\theta}_{\tau_2}(\tau_1)) \Rightarrow N(0, I)$.

The asymptotic covariance matrix expressed in (11) is a general one that includes the possibility conditional heteroskedasticity. For the actual estimation, it is much more computationally convenient to treat w_t and u_t as both conditionally homoskedastic instead. In this case, we achieve further simplication of the covariance matrix formula, since $Q_1(\boldsymbol{\theta}_{\tau_2}(\tau_1)) = f(F^{-1}(\tau_1))D_1$, where $f(F^{-1}(\tau_1))$ in turn is $f(0)$ since $F^{-1}(\tau_1) = 0$. Under A5 and A7, D_1 may be consistently estimated using $\hat{D}_{1,T}$ while $\hat{f}(0)$ may be estimated as the inverse of the quantile density function, i.e. $s(\tau_1) = 1/f(F^{-1}(\tau_1))$, using the nonparametric method of Siddiqui (1961) and the bandwidth proposed by Hall and Sheather (1988). Details of the procedure are available in Koenker (2005). Since $f(\eta\Upsilon_t(\boldsymbol{\theta}_{\tau_2}(\tau_1))$ converges to $f(0)$ in probability, $\hat{f}(0)$ will be used to estimate Q_3, which is estimated by $\hat{Q}_{3,T} = T^{-1}\sum_{t=1}^{T}\hat{f}(0)\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1)'\hat{\mathbb{X}}_{1,t}^{d}\hat{\mathbb{X}}_{1,t}X_{2,t}'$. To estimate $\Omega_{\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1)}$, we estimate D_2 using $D_{2,T}$ and Q_2 using $\hat{g}(0)D_{2,T}$, where $\hat{g}(0)$ is the inverse of the nonparametric quantile density estimator and absence of the circumflex over $D_{2,T}$ expresses the fact that no generated regressors are used in the first step regression.

For robust estimation of the covariance matrix robustly under conditionally heteroskedasticity errors, we may first estimate the covariance matrix of $\sqrt{T}(\hat{\gamma}(\tau_2) - \gamma(\tau_2))$ using $\hat{\Omega}_{\hat{\gamma}(\tau_2)} = \tau_2(1-\tau_2)\hat{Q}_{2,T}^{-1} D_{2,T} \hat{Q}_{2,T}^{-1}$, where $\hat{Q}_{2,T} = T^{-1}\sum_{t=1}^{T} \hat{g}_{t,\tau_2} X_{2,t} X_{2,t}'$ and $\hat{g}_{t,\tau_2}$ is the Hendricks-Koenker quantile density estimator (see Koenker 2005, p. 80), that is

$$\hat{g}_{t,\tau_2} = \max\left\{0, \frac{2b_k}{X_{2,t}'\hat{\gamma}(\tau_2+b_k) - X_{2,t}'\hat{\gamma}(\tau_2-b_k) - e}\right\}$$

where e is a small number to prevent division by zero and b_k is the bandwidth where the Bofinger (1975) and Hall and Sheather (1988) bandwidths are the possible candidates. Then to estimate the covariance matrix of $\sqrt{T}(\hat{\boldsymbol{\theta}}_{\tau_2}(\tau_1) - \boldsymbol{\theta}_{\tau_2}(\tau_1))$ robustly, we use the robust estimator $\hat{\Omega}_{\hat{\gamma}(\tau_2)}$ together with $\hat{D}_{1,T}$, $\hat{Q}_{1,T} = T^{-1}\sum_{t=1}^{T} \hat{f}_{t,\tau_1} \hat{\mathbb{X}}_{1,t} \hat{\mathbb{X}}_{1,t}'$ and $\hat{Q}_{3,T} = T^{-1}\sum_{t=1}^{T} \hat{f}_{t,\tau_1} \hat{\mathbb{X}}_{1,t} \hat{\mathbb{X}}_{1,t}^{d}{}' \boldsymbol{\theta}_{\tau_2}(\tau_1) X_{2,t}'$, where $\hat{f}_{t,\tau_1}$ is the Hendricks-Koenker density estimator.

5 Monte Carlo Simulation

In this section, we compare the performance of the model under various assumptions about the order of the polynomial. Following the convention in Section 3.2, we let α_0 be the intercept, α_1 be the coefficient on $X_{1,t}$ and α_2 be the coefficient on $Y_{2,t}$. We consider three cases:

Case 1: w_t innovations in α_2 only and series expansions for α_2 only.

Consider the data generating process

$$Y_{1,t} = \alpha_0 + \alpha_1 X_{1,t} + (\alpha_2 + \delta(\lambda e^{w_t} + u_t))Y_{2,t} \tag{12}$$

$$Y_{2,t} = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + w_t \tag{13}$$

where $(\alpha_0, \alpha_1, \alpha_2, \delta, \lambda) = (3, 4, 4, 5, 3)$, $(\beta_1, \beta_2, \beta_3) = (1, 2, 3)$, $X_{1,t} \sim t_3$, $X_{2,t} \sim N(15, 2)$,

$w_t \sim N(0, 0.5)$ and $u_t \sim N(0, 1)$. This generating process is similar to the benchmark model of Ma and Koenker (2006), except that we specify e^{w_t} in (12) while they simply used w_t instead. In doing so, we incorporate a nonlinear feature in how w_t enters the slope coefficient on $Y_{2,t}$. We want to estimate α_2 based on the equation

$$Y_{1,t} = \alpha_0 + \alpha_1 X_{1,t} + \alpha_2(w_t, u_t) Y_{2,t}$$

understanding that $\alpha_2(w_t, u_t)$, from the researcher's perspective, is an unknown function of w_t and u_t. The true value of $\alpha_2(\tau_2, \tau_1)$, which is of interest, is $4+5(3\exp(F_w^{-1}(\tau_2))+F_u^{-1}(\tau_1))$ and our objective is to estimate this value as best as we can. To do so, we consider series expansions up to the quartic polynomial. Following the discussion in Section 3, we first employ the regression function of

$$\hat{H}_{Y_{1,t}} = \alpha_0 + \alpha_1 X_{1,t} + \alpha_2 Y_{2,t} + \sum_{k=1}^{I} \varphi_k \frac{\hat{w}_t(\tau_2)^k}{I!} Y_{2,t} \tag{14}$$

where k indexes the power of $\hat{w}_t$ and $I = 1$ to 4. Here, the estimate of interest is $\hat{\alpha}_2(\tau_2, \tau_1)$ in (18) and the summation term controls for the nuisance term as mentioned before. By employing (14), we are hypothesizing that the researcher knows that only w_t enters into the slope coefficient on $Y_{2,t}$, thus justifying the expansion for α_2 alone. The Monte Carlo experiment is carried out by simulating data from (12) and (13) and estimating $\hat{\alpha}_2(\tau_2, \tau_1)$ for each simulation. We consider a grid of $\tau = [0.1, 0.2, \ldots, 0.9]$, with a total of nine categories in τ, resulting in 81 regressions corresponding to each τ_1 and τ_2 located on the grid. We employ 200 simulations with 1000 observations and report the average of the estimated $\hat{\alpha}_2(\tau_2, \tau_1)$.

In Figure 1, we present the surface of $\alpha_2(\tau_2, \tau_1)$ (Panel A) together with the estimated surfaces $\hat{\alpha}_2(\tau_2, \tau_1)$ based on the linear expansion model (Panel B) to the quartic expansion model (Panel E). In these plots, the larger values are more lightly shaded. From the figure, we can see that the shapes of the estimated surfaces are very similar to the shape of the true surface. For the linear expansion model, the estimated surface deviates slightly from the true surface in the extreme quantiles. For instance, in a 10th percentile regression of $Y_{1,t}$, the estimated slope coefficient on the 10th percentile of $Y_{2,t}$ appears to be smaller than the true value, i.e. $\hat{\alpha}_2(0.1, 0.1) < \alpha_2(0.1, 0.1)$. However, the estimated surface becomes very close to the true surface even in the extremes when a quadratic, cubic or quartic model is used.

In Table 1, we report the true parameter values, the estimated values, the remainder and the root-mean-squared errors based on regressions confined to $\tau_1 = \tau_2 = \tau$. Generally, the root-mean-squared error is similar across the four regression models. The bias, except at $\tau = 0.9$, generally declines as we move from a linear to a quadratic model and further declines when a cubic or quartic model is used. However, the bias from the quartic model is not always the least among all the four models, although it is the case in six of the nine categories of τ.

Case 2: w_t innovation in α_2 only and series expansions for all α_0, α_1 and α_2.

Here, the data is generated from (12) and (13). However, we assume that the researcher does not know if w_t enters into the other coefficients also, and thus takes this possibility into account by including the expansions for α_0 and α_1. The regression function in this case is

$$\hat{H}_{Y_{1,t}} = \alpha_0 + \alpha_1 X_{1,t} + \alpha_2 Y_{2,t} + \sum_{k=1}^{I} \varphi_{k,0} \frac{\hat{w}_t(\tau_2)^k}{I!} + \sum_{k=1}^{I} \varphi_{k,1} \frac{\hat{w}_t(\tau_2)^k}{I!} X_{1,t} + \sum_{k=1}^{I} \varphi_{k,2} \frac{\hat{w}_t(\tau_2)^k}{I!} Y_{2,t} \tag{15}$$

Equation (15) contains two additional summation terms than (14), reflecting the fact the expansions for α_0 and α_1 are included. The estimated surfaces are shown in Figure 2 while Table 2 reports the true parameter values, the estimated values, the bias and the root-mean-squared errors.

Figure 2 shows that the shape of the true parameter surface is satisfactorily estimated by all the models. In addition, as in Case 1, the bias becomes smaller when higher order polynomials are used. Nevertheless, the bias from this overfitted model is generally larger than that of Case 1, reflecting the relative impreciseness of the regression function used here. The root-mean-squared error is also noticeably larger for the $\tau = 0.1$ and $\tau = 0.9$ regressions when the linear model is used. Again, the cubic and quartic models are recommended as they yield the smallest bias in eight of the nine categories of τ. The only exception is $\tau = 0.1$, where the quadratic model yields the smallest bias.

Case 3: w_t innovation in α_0, α_1 and α_2 only and series expansions for α_2 only.

What happens if w_t enters all three coefficients but series expansion is only applied for α_2? For instance, suppose the true data generating function of $Y_{2,t}$ is

$$Y_{1,t} = (\alpha_0 + \tilde{\delta}(\tilde{\lambda}e^{w_t} + u_t)) + (\alpha_1 + \tilde{\delta}(\tilde{\lambda}e^{w_t} + u_t))X_{1,t} + (\alpha_2 + \delta(\lambda e^{w_t} + u_t))Y_{2,t} \qquad (16)$$

where $\tilde{\delta} = 10$ and $\tilde{\lambda} = 5$ and the other constants are the same as before. However, the regression function is based on (14). This corresponds to the case where the researcher is only interested in variations exhibited by α_2 and not by the other parameters, and thus adopts the regression function containing expansions for α_2 only. Certainly, this misspecifies the relation that w_t influences α_0 and α_1 since the regression function should include expansions for these parameters also.

However, as Figure 3 demonstrates, the shape of α_2 is satisfactorily estimated even in the presence of this misspecification. This result is useful as it suggests that the researcher can focus on expanding only the parameters of interest while leaving the others untouched. This advantage is also reinforced by Table 3. While the bias is generally larger when compared to Case 1, this may not always be so when compared to Case 2. For instance, both cubic and quartic regressions in Case 3 perform better than those in Case 2 at $\tau = 0.5$. In other words, underfitting may not necessarily perform worse than overfitting especially as far as dependence between the center of the distributions is concerned. If one is willing to tolerate a slightly larger bias occuring in the extremes, then to maintain parsimony, it is may not be necessary to expand every single parameter specified in the original model.

6 Empirical Results

As explained in the introduction, the signs of the estimated monetary policy shocks have been used to identify episodes of monetary expansion or contraction. The econometric methodology proposed by this paper comes from observing that the signs of these shocks, which are estimated from ordinary least squares, may not correctly reflect the true policy position unless the monetary innovation has a zero population mean.

In motivating the quantile dependence framework, our objective is to construct a model to investigate the money-output relationship that does not rely on the signs of the shocks. To do so, we exploit the idea that the quantiles of the monetary policy shock convey information about the policy stance by interpreting a lower quantile shock as restrictive relative to a higher quantile shock, and likewise a higher quantile shock as expansive relative to a lower quantile shock. Therefore, monetary policy stance will be described here as "restrictive" and "expansive" rather than "contractionary" and "expansionary".

The empirical model will be based on Cover (1992) who formulated a two-equation system.[10] First, the monetary policy shocks are identified as residuals in a monetary process equation as

$$m_t = \alpha_0 + \sum_{i=1}^{K_m} \alpha_{m,i} m_{t-i} + \sum_{i=1}^{K_x} \alpha'_{x,i} x_{t-i} + u_t \tag{17}$$

where m_t is a monetary instrument and x_t is a vector that contains other information variables. Cover employed M1 money supply growth as the policy instrument for his analysis on the post-war money-output relationship while DeLong and Summers (1988) employed M2 and M3 money supply growth when investigating the relationship during the pre-war and pre-Depression periods. There were also others who used non-money measures such as the Federal funds rate (Morgan, 1993) and short-term interbank offer rates (Florio, 2005). Having obtained the monetary policy shocks obtained from (17), output growth is then

[10] Florio (2004) surveyed the literature on the monetary policy-output nexus in the spirit of Cover.

regressed on the negative and positive shocks using

$$y_t = \beta_0 + \sum_{i=1}^{K_y} \beta_{y,i} y_{t-i} + \sum_{i=1}^{K_r} \beta_{r,i} dr_{t-i} + \sum_{i=0}^{K_u} (\beta_{u,i}^{+} u_{t-i}^{+} + \beta_{u,i}^{-} u_{t-i}^{-}) + w_t \tag{18}$$

where y_t is output growth and dr_t is the first differenced three-month Treasury yield. Whichever monetary instrument is used, the same conclusion generally emerges: output reduces more following a negative shock than it increases following a positive shock of the same size. This is reflected by the fact that the sum of $\beta_{u,i}^{+}$ is usually smaller than the sum of $\beta_{u,i}^{-}$, hence indicating that the monetary effect on output growth is asymmetric.

In Cover's benchmark model, output growth was regressed on the first lag of output growth, contemporaneous and lagged values of the first differenced Treasury yield, and contemporaneous positive and negative monetary shocks. Other extensions included lagged monetary shocks, but the contemporaneous negative shocks were typically the most important, statistically and size-wise, while the positive shocks were usually statistically insignificant. This motivates a more parsimonious setup to investigate how output growth reacts to a change in the quantile of a *contemporaneous* monetary shock, which can be expressed by rewriting (18) as

$$\begin{aligned} y_t &= (\beta_0 + \beta_u u_t + w_t) + \sum_{i=1}^{K_y} \beta_{y,i} y_{t-i} + \sum_{i=1}^{K_r} \beta_{r,i} dr_{t-i} \\ &= \beta_0(u_t, w_t) + \sum_{i=1}^{K_y} \beta_{y,i} y_{t-i} + \sum_{i=1}^{K_r} \beta_{r,i} dr_{t-i} \end{aligned} \tag{19}$$

where the second line writes β_0 as a random intercept term. Based on the quantile dependence framework, since the τ_1^{th} conditional quantile of y_t corresponds to the τ_1^{th} quantile of w_t, this implies that the τ_1^{th} conditional quantile of y_t depends on monetary policy asymmetrically if $\beta_0(\tau_2, \tau_1) \neq \beta_0(1 - \tau_2, \tau_1)$ holds.

Equation (19), which is a pure location shift model, may be generalized based on our earlier discussion on quantile dependence. Relaxing to a location and scale shift model, the quantile of m_t may be allowed to affect y_t through the presence of u_t in the slope parameters.

Therefore, assuming that the slopes are influenced by both u_t and w_t, a general output process can be written as

$$y_t = \beta_0(u_t, w_t) + \sum_{i=1}^{K_y} \beta_{y,i}(u_t, w_t) y_{t-i} + \sum_{i=1}^{K_r} \beta_{r,i}(u_t, w_t) dr_{t-i} \tag{20}$$

Now, although the asymmetric relationship between monetary policy and output growth is a stylized fact, the quantile dependence framework may nevertheless bring other aspects to light in the money-output relationship that cannot be estimated using conventional methods. First, while a negative monetary shock is known to influence output growth more strongly than positive shock, a traditional framework cannot determine if a larger negative shock exerts a larger *marginal* influence. For instance, using ordinary least squares regression, the only estimable quantity is the average marginal influence of the negative shock. This disregards the possibility that negative shocks of various sizes may influence output growth differently. Thus by using the quantile-based framework, one can get a sense of how much more sensitive output growth is when the policy stance becomes even more restrictive or expansive.

Second, the quantile dependence framework makes it possible to examine if the quantiles of output growth, say the 90th percentile and the median output growth, respond differently to the same monetary policy stance. As we will see, our estimation results reveal that the 90th percentile of output growth is more sensitive to variations in M1 money supply shocks than are the 10th and 50th percentiles. Therefore, in addition to addressing whether monetary policy affects a given level of output growth asymmetrically, our framework can also be used to uncover another possible dimension of nonlinearity that expresses how quantiles of output growth may respond differently to the same policy stance.

For the actual empirical implementation, we consider two policy instruments: M1 and M2 money supply.[11] For the monetary process equation, we regress the monetary instrument

[11] While the Federal funds rate is also a monetary policy instrument, however using this measure implies that Cover's model can no longer be extended to the QQ model since the upper quantiles of the Federal funds innovation will now represent restrictive policies while the lower quantiles will be expansive. This indexation scheme is opposite to the one when the money supply is used.

on twelve of its lags as well as the first lag of the first differenced Treasury yield.[12] For the output process equation, we consider a parsimonious specification in the form of

$$y_t = \beta_0(u_t, w_t) + \sum_{i=1}^{K_y} \beta_{y,i}(u_t, w_t) y_{t-i} + \beta_r(u_t, w_t) dr_{t-1} \tag{21}$$

Variations in the random intercept term are of particular interest given the location shift specification of Cover's model. In addition, by imposing a generalized structure in (21), we allow the monetary shock to also influence the slope parameters. We consider a model with eight and twelve lags of output growth, i.e. $K_y = 8$ or 12, and the first lag of the first differenced Treasury yield. To keep the exposition concise, the paper only reports the estimates for $\beta_0(\tau_2, \tau_1)$, $\beta_{y,1}(\tau_2, \tau_1)$ and $\beta_r(\tau_2, \tau_1)$ based on the cubic and quartic models. For parsimony, the paper only considers expanding the coefficients on the first four lags of output, i.e. $\beta_{y,i}$, for $i = 1, \ldots, 4$.

Monthly time series from Datastream is used while previous research typically employed quarterly time series. Output growth is defined as the growth rate of the industrial production index, replacing Gross Domestic Product when quarterly data is used. The starting date of the dataset is January 1970 and the ending date is January 2009. All growth variables are obtained by log-differencing and multiplying by 100. The standard errors are calculated under the assumption of homoskedasticity. First, we turn our attention to the main parameter of interest: the random intercept term.

The Random Intercept Term

Figures 4 and 5 plot the random intercept surface when the monetary instrument is M1 and M2 money supply growth respectively. Not surprisingly, the surface is downward sloping as the quantile of output growth declines, meaning that the random intercept term is reduced as we move towards the lower quantiles of output growth. When monetary policy influences output growth, the surface will also vary along the quantiles of the monetary policy shock.

[12] Cover also included in the information set the lagged government budget surplus, the ratio of unemployed over employed, and the lag of output growth. However, these variables are usually statistically insignificant.

In this case, the surface should be tilted towards the $(0, 0, z_{\min})$ vertex as we move towards the lower quantiles of monetary shock since these quantiles reflect an increasingly restrictive policy stance, thus further reducing the intercept.

Focusing on M1 money supply, when output growth is located in the upper quantiles, moving to a lower quantile monetary shock will tend to lower the intercept, which is equivalent to saying that a more restrictive monetary policy will tend to reduce output growth when output growth is large. However, changing to a more restrictive policy has little impact on the lower quantiles of output growth, i.e. the left tail of the output growth distribution. For M2 money supply, a more restrictive policy reduces the left-tail output growth more than M1 does, hence demonstrating that output growth is more sensitive to restrictive M2 than M1 money supply shocks when it is low.

To see the influence of monetary policy on output growth more clearly, we plot cross-sections of the surface dissected in the output dimension. These figures demonstrate how the intercept term responds to changes in the monetary shocks when output growth is at the 10th, 50th and 90th percentiles. The subplots based on regression specifications with eight and twelve lags of output growth are shown in Panel A and B respectively.

For M1 money supply, Figures 6 and 7 plot the cross-section estimates corresponding to the cubic and quartic models. The dotted horizontal line in each subplot reflects the value of the intercept at the median monetary shock, which will be used as the reference point so that quantiles below the median are more restrictive and those above are more expansive than the median policy stance.

First, focus on the cubic regression model whose results are shown in Figure 6. At the 10th percentile of output growth, both regression specifications with eight and twelve lags of output growth suggest that varying the quantiles of M1 money supply shock does little to shift the intercept away from the horizontal line, which is the value of the intercept corresponding to the median policy stance. When output growth is at the median, there is evidence that monetary policy is asymmetric. In Panel B (twelve lags of output growth), the median output growth stays around 0.23% per month when the magnitude of M1 money supply shock is at least the median. However, the median output growth declines to 0.13% per month when monetary shock declines to the 10th percentile. This asymmetry is also observed at the 90th percentile of output growth. In Panel B, changing the monetary stance from the median to the 90th percentile increases the 90th percentile of output growth by 0.13% per month, but decreases it by 0.2% per month when monetary shock is reduced from the median to the 10th percentile.

So far, the above findings can be summarized into two points: 1) the right tail of output growth tends to be the most sensitive to variations in the M1 money supply shock, and 2) whenever it is important, e.g. at the 90th percentile of output growth, the monetary shock tends to influence output growth more strongly for restrictive policies. These conclusions are also echoed by the quartic model as reported by Figure 7. Among the 10th, 50th and 90th percentiles of output growth, the latter is most sensitive to changes in the monetary policy stance where at the same time, monetary policy exhibits asymmetry. In Panel A (eight lags of output growth), changing the monetary policy stance from the median to the 90th percentile increases the 90th percentile of output growth by 0.10% per month but moving down to the 10th percentile shock it reduces by 0.17% per month. Similarly in Panel B, the impact on the 90th percentile of output growth is an increase of 0.16% per month moving from the median to the 90th percentile of the monetary shock and a decline of 0.21% per month moving from the median to the 10th percentile shock.

For the M2 money supply, Figures 8 and 9 plot the cross-section estimates corresponding to the cubic and quartic models respectively. Here, the 10th, 50th and 90th percentiles of output growth are more sensitive to M2 than M1 money supply shocks. Furthermore, the asymmetric influence of M2 money supply shocks are more pronounced. Focusing on the cubic regression model with results shown in Figure 8, Panel A shows that the 10th percentile of output growth increases only by 0.10% per month after changing the monetary stance from the median to the 90th percentile. However, the 10th percentile of output growth declines by 0.26% per month when the monetary stance shifts from the median to the 10th percentile. In Panel B, the same experiment reveals that 10th percentile of output growth expands only by 0.11% per month in the expansive direction but contracts by 0.19% per month in the restrictive direction. As already mentioned, the asymmetric effects of M2 money supply shocks are also evident at the median and the 90th percentile of output growth, where the latter is more sensitive than the median to changes in the policy stance.

Slope on the First Lag of Output Growth

Figures 10 and 11 plot the estimated slope surface on the first lag of output growth when the monetary instrument is M1 and M2 money supply growth respectively. Contrasting the intercept surfaces, both figures show that the slope surface tends to be flat across most quantiles of monetary shock and output growth, but become increasingly elevated at the lower quantiles of output growth. This implies that the left tail, rather than the right tail, of contemporaneous output growth tends to react more strongly to lagged output growth. Therefore, a stronger lagged output growth may effectively provide resistance against the slowdown in contemporaneous output growth during downturns, but not effectively boost growth during upturns.

This asymmetry can be seen more clearly when we examine the cross-sections of the estimated slope surface. For the sake of exposition, let us do so while fixing the monetary shock at the median. On these cross-sectional results, Figures 12 and 13 plot the cubic and quartic regression estimates based on M1 money supply. In Panel A of Figure 12 (regression with eight lags of output growth), the estimated slope parameter is 0.38 at the 10th percentile of output growth, which is greater than the estimates corresponding to the median and 90th percentile of output growth at 0.09 and 0.15 respectively. Similarly in Panel B (regression with twelve lags of output growth), the estimated slope is 0.30 at the 10th percentile of output growth, which again is greater than those to median and 90th percentile of output growth at 0.12 and 0.15 respectively.

For the quartic model, Panel A of Figure 13 shows that the estimated slope at the median monetary stance is 0.40, 0.09 and 0.13 for the 10th, 50th and 90th percentiles of output growth. Panel B tells a similar story, further supporting the fact that the left tail of contemporaneous output growth is more responsive to lagged output growth.

This asymmetry also shows up at times when the M2 money supply is used. Here, Figures 14 and 15 plot the cross-sections corresponding to the cubic and quartic models. For the cubic regression model, Panel B of Figure 14 shows that the slope is 0.17, 0.13 and 0.21 for the 10th, 50th and 90th percentiles of output growth, implying that the extreme quantiles, rather than the centrally located quantiles of output growth are more sensitive to lagged output growth. The asymmetry, seen when M1 money supply is used, re-emerges in the quartic model as Panel B of Figure 15 shows that the slope is 0.26, 0.14 and 0.16 at the 10th, 50th and 90th percentile of output growth.

Now, varying the monetary shock, we can see that the slope is only weakly affected by the changes in the M1 money supply shock.[13] This is also generally true when M2 money supply is used. Taken together, we may conclude that the relationship between the contemporaneous output growth quantiles and the first lag of output growth is generally robust to changes in monetary policy.

Slope on the First Lag of the First Differenced Treasury Yield

Figures 18 and 19 plot the estimated slope surface on the first lag of the first differenced Treasury yield when the monetary instrument is the M1 and M2 money supply growth respectively. Both figures show that the surface is typically elevated at the lower quantiles of output growth and monetary shock. However, when M2 money supply is used, the surface appears to be like a saddle, elevated in both tails of output growth and monetary shock and depressed around the center of the distributions.

Using the M1 money supply, the cross-section diagrams of Figure 18 and 19 based on the cubic and quartic models show that the slopes in the 10th and the 50th percentile of output growth regressions are generally statistically insignificant for most quantiles of

[13] The exception happens when output growth is at the 10th percentiles and the eight-lag specification is used. However, variations in slope diminish when the twelve-lag specification is used instead, thus giving further support that monetary policy stance exerts only weak effects on the coefficient on the first lag of output growth.

monetary shock. At the 90th percentile of output growth, the slopes generally become mostly statistically significant, ranging from a value of -0.30 (Figure 19, Panel B) to -0.36 (Figure 19, Panel A) at the median monetary shock. This observation is also similar when M2 money supply is used, where Figures 20 and 21 show that for the 10th and the 50th percentile of output growth regressions, the slope is generally statistically insignificant for most monetary stances. The 90th percentile of output growth again provides one exception, where the slope is negative and becomes statistically significant when the monetary shock is less than or equal to its median.

The above finding implies that a large increase in the Treasury yield in the previous period is likely to adversely affect the next period's output growth when output growth is high, e.g. at the 90th percentile. Output growth near the center and the left half of the distribution would not generally be affected by lagged changes in the Treasury yield.

Summary

Summarizing our results, we find that

1. The right tail of output growth is generally more sensitive to changes in M1 money supply shocks, while both tails of output growth are more sensitive than the center is to changes in M2 money supply shocks, implying that monetary policy measured by M2 is more effective in bearish and bullish periods of growth.

2. When non-neutral, the influence of monetary policy on output growth is stronger when it is restrictive than expansive, consistent with previous findings on the asymmetric influence of money on output.

3. Contemporaneous output growth responds positively and is more sensitive to lagged output growth when it is located in the left tail.

4. Changes in monetary policy will only weakly affect how lagged output growth influences the quantile of contemporaneous output growth.

5. The right tail of output growth is more likely to be sensitive to past changes in the Treasury yield.

7 Conclusion

Using a newly developed quantile dependence framework, this paper investigates whether two types of nonlinearities are present in the relationship between monetary policy and output growth. First, it examines whether the same quantile of output growth responds differently to changes in monetary policy and finds that whenever monetary policy is effective, the quantile of output growth responds more to restrictive than expansive policy. Second, it investigates which quantiles of output growth are more sensitive to a particular monetary stance. On this issue, the results are dependent on the chosen monetary instrument. When M1 money supply is used, the right tail of output growth distribution is more sensitive to changes in monetary policy than elsewhere, while both left and right tails are the more sensitive than the center of the output growth distribution when M2 money supply is used.

Hence, based on the M1 money supply as the monetary instrument, restrictive monetary policy is useful to slow output growth only when output growth is high. Based on the M2 money supply, restrictive monetary policy influences the entire output growth distribution effectively but is more effective when output growth is located in the tails. Therefore, while monetary policy is asymmetric, the asymmetry becomes more pronounced for specific locations in the output growth distribution. This new observation demonstrates the flexibility of the quantile dependence framework in modeling the relationship between the distributions of monetary shocks and output growth.

As the empirical focus of this paper centers on the reduced-form relationship between monetary policy and output growth, a natural extension would be to examine the quantile dependence between them in a structural model. This model may be built upon a linearized new Keynesian Dynamic Stochastic Equilibrium framework, where the quantile dependence equation is the linearized Euler equation that models output growth as a function of inflation and the interest rate, which in turn are modeled by the new Keynesian Phillips curve and the Taylor rule. Also, a possible extension would be to unify the SVAR model from the application perspective and quantile dependence from the econometric theory perspective.

References

Andrews, D. W. K. (1987) Consistency in nonlinear econometric models: A generic uniform law of large numbers, *Econometrica*, 55 (6), pp. 1465 - 71.

Bofinger, E. (1975) Estimation of a density function using order statistics, *Australian Journal of Statistics*, 17, pp. 1-7.

Boschen, J. F. and Mills, L. O. (1995) The relation between narrative and money market indicators of monetary policy, *Economic Inquiry*, 33(1), pp. 24-44.

Conover, C. M., Jensen, G. R. and Johnson, R. R. (1999) Monetary environments and international stock returns, *Journal of Banking and Finance*, 23(9), pp. 1357-81.

Cover, J. P. (1992) Asymmetric effects of positive and negative money-supply shocks, *Quarterly journal of Economics*, 107, pp. 1261-82.

DeLong, B. and Summers, L. (1988) How does macroeconomic policy affect output, *Brookings Papers on Economic Activity*, 2, pp. 433-94.

Florio, A. (2004) The asymmetric effects of monetary policy, *Journal of Economic Surveys*, 18(3), pp. 409-26.

Florio, A. (2005) Asymmetric monetary policy: Empirical evidence for Italy, *Applied Economics*, 37(7), pp. 751-64.

Hall, P. and Sheather, S. (1987) On the distribution of a studentized quantile, *Journal of Royal Statistical Society B.*, 50(3), pp. 381-91.

He, X. and Shao, Q. M. (2000). On parameters of increasing dimension, *Journal of Multivariate Analysis*, 73(1), pp. 120-35.

Huber, P. J. (1973) Robust regression: Asymptotics, conjectures and Monte Carlo, *Annals of Statistics*, 1(5), pp. 799-821.

Karras, G. (1996) Are the output effects of monetary policy asymmetric? Evidence from a sample of European countries, *Oxford Bulletin of Economics and Statistics*, 58, pp. 267-78.

Koenker, R. and Bassett, G. (1978) Regression quantiles, *Econometrica*, 46(1), pp. 33-50.

Koenker, R. and Park, B.J. (1996) An interior point algorithm for nonlinear quantile regression, *Journal of Econometrics*, 71(1-2), pp. 265-83.

Koenker, R. and Zhao, Q. (1994) L-estimation for linear heteroscedastic models, *Journal of Nonparametric Statistics*, 3(3-4), pp. 223-35.

Koenker, R. (2005) *Quantile Regression*, Econometric Society Monographs 38, Cambridge University Press.

Lee, S. and Su, Z. (2002) The symmetry in the martingale inequality, *Statistics and Probability Letters*, 56(1), pp. 83-91.

Lo, M. C. and Piger, J. (2005) Is the response of output to monetary policy asymmetric? Evidence from a regime-switching coefficients model, *Journal of Money, Credit and Banking*, 37(5), pp. 865-86.

Ma, L. and Koenker, R. (2006) Quantile regression methods for recursive structural equation models, *Journal of Econometrics*, 134(2), pp. 471-506.

Morgan, D. P. (1993) Asymmetric effects of monetary policy, Federal Reserve Bank of Kansas City, *Economic Review*, 78, pp. 21-33.

Newey, W. (1991) Uniform Convergence in Probability and Stochastic Equicontinuity, *Econometrica*, 59, pp. 1161-7.

Newey, W. (1997) Convergence rates and asymptotic normality for series estimators, *Journal of Econometrics*, 79, pp. 147-68.

Newey, W. and McFadden, D. (1994) Large sample estimation and hypothesis testing, *Handbook of Econometrics*, 4, pp. 2111-245.

Parker, R. E. and Rothman, P. (2004) An examination of the asymmetric effects of money supply shocks in the pre-World War I and interwar periods, *Economic Inquiry*, 42(1), pp. 88-100.

Ravn, M. O. and Sola, M. (2004) Asymmetric effects of monetary policy in the United States, Federal Reserve Bank of St. Louis, *Review*, September, pp. 41-60.

Resnik, S. I. (2003) *A probability path.* Birkhauser: Boston.

Rhee and Rich (1995) Inflation and the asymmetric effects of money on output fluctuations, *Journal of Macroeconomics*, 17, pp. 683-702.

Romer, C. D. and Romer, D. H. (2004) A new measure of monetary shocks: Derivation and implications, *American Economic Review*, 94(4), pp. 1055-84.

Senda, T. (2001) Asymmetric effects of money supply shocks and trend inflation, *Journal of Money, Credit and Banking*, 33, pp. 65-89.

Sims, C. A. (1992) Interpreting the macroeconomic time series facts : The effects of monetary policy, *European Economic Review*, 36(5), pp. 975-1000.

Weise, C. L. (2008) Private sector influences on monetary policy in the United States, *Journal of Money, Credit and Banking*, 40(2-3), pp. 449-62.

Zernov, S., Zinde-Walsh, V. and Galbraith, J. W. (2009) Asymptotics for estimation of quantile regressions with truncated infinite-dimensional processes, *Journal of Multivariate Analysis*, 100(3), pp. 497-508.

Appendix

To simplify notation, we will suppress the τ_1 argument and τ_2 subscript in $\boldsymbol{\theta}_{\tau_2}(\tau_1)$ so that the population parameter vector is $\boldsymbol{\theta}$ and the estimated parameter vector is $\hat{\boldsymbol{\theta}}$. For the proofs, define $\|A\| = tr(A'A)^{1/2}$ where tr is the trace operator. In addition, express the objective function as

$$L_T(\tilde{\boldsymbol{\theta}}) = T^{-1} \sum_{t=1}^{T} \left[\rho_{\tau_1}\left(u_t(\tau_1) - ((\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'\hat{\mathbb{X}}_{1,t} + (\hat{\mathbb{X}}_{1,t} - \mathbb{X}_{1,t})'\boldsymbol{\theta} - \Gamma_t' i_3)\right) - \rho_{\tau_1}\left(u_t(\tau_1)\right)\right] \tag{22}$$

The normalization with $\rho_{\tau_1}(u_t(\tau_1))$ is done as matter of convenience for the asymptotic analysis and will not affect the estimation outcome. More importantly, this normalization facilitates using Knight's identity which comes in useful for the proof of consistency (Proposition 1) and uniform law of large numbers (Lemma 1). $\hat{\boldsymbol{\theta}}$ is the minimizer of (22) and the first order condition is

$$\hat{\mathbb{W}}_T(\hat{\boldsymbol{\theta}}) = -T^{-1} \sum_{t=1}^{T} \hat{\mathbb{X}}_{1,t} \psi_{\tau_1}\left(Y_t - \hat{\boldsymbol{\theta}}'\hat{\mathbb{X}}_{1,t}\right)$$

which is equal to zero except for a set of measure zero.

Lemma 1. *(Uniform Law of Large Numbers) The objective function, $L_T(\tilde{\boldsymbol{\theta}})$, defined in* (22)*, satisfies*

$$\sup_{\tilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} |L_T(\tilde{\boldsymbol{\theta}}) - E[L_T(\tilde{\boldsymbol{\theta}})]| \rightarrow 0$$

as $T \rightarrow 0$.

Proof: Using Knight's identity, i.e. $\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_{\tau_1}(u) + \int_0^v I(0 < u \leq s)ds$ and letting $\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}}) = (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'\hat{\mathbb{X}}_{1,t} + \boldsymbol{\theta}'(\hat{\mathbb{X}}_{1,t} - \mathbb{X}_{1,t}) - \Gamma_t' i_3$, we may express

$$\sup_{\tilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} |L_T(\tilde{\boldsymbol{\theta}}) - E[L_T(\tilde{\boldsymbol{\theta}})]|$$

$$\leq \sup_{\tilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} |T^{-1} \sum_{t=1}^{T} (\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})\psi_{\tau_1}(u_t(\tau_1)) - E[\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})\psi_{\tau_1}(u_t(\tau_1))])| \tag{23}$$

$$+ \sup_{\tilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} |T^{-1} \sum_{t=1}^{T} \left(\int_0^{\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})} I(0 < u_t(\tau_1) \leq s) - E[\int_0^{\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})} I(0 < u_t(\tau_1) \leq s)ds]\right)| \tag{24}$$

We now show that (23) is $o(1)$. To do so, we verify assumptions A1, A2 and A3a of Newey (1991). Assumption A1 of Newey (1991) requires compactness of the parameter set, which is A2 of this paper. Assumption A2 of Newey (1991) requires that (23) holds pointwise. Hence, consider some $\tilde{\boldsymbol{\theta}}_l \in \boldsymbol{\Theta}$. Applying Chebyshev inequality and the law of total variance, we have

$$
\begin{aligned}
&P\left(|T^{-1}\sum_{t=1}^{T}(\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}}_l)\psi_{\tau_1}(u_t(\tau_1)) - E[\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}}_l)\psi_{\tau_1}(u_t(\tau_1))])| \geq \delta/2\right) \\
\leq&\frac{4}{\delta^2}T^{-1}E[Var[((\tilde{\boldsymbol{\theta}}_l-\boldsymbol{\theta})'\hat{\mathbb{X}}_{1,t}+\boldsymbol{\theta}'(\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t})-\Gamma_t'i_3)\psi_{\tau_1}(u_t(\tau_1))|\mathbb{X}_{1,t}]] \\
\leq&\frac{4}{\delta^2}T^{-1}E[((\tilde{\boldsymbol{\theta}}_l-\boldsymbol{\theta})'\hat{\mathbb{X}}_{1,t}+\boldsymbol{\theta}'(\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t})-\Gamma_t'i_3)^2Var[\psi_{\tau_1}(u_t(\tau_1))|\mathbb{X}_{1,t}]] \\
\leq&\frac{4\tau_1(1-\tau_1)}{\delta^2}T^{-1}E[(|(\tilde{\boldsymbol{\theta}}_l-\boldsymbol{\theta})'\hat{\mathbb{X}}_{1,t}|+|\boldsymbol{\theta}'(\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t})|+3K\lambda^{-s})^2] \\
\leq&\frac{4\tau_1(1-\tau_1)}{\delta^2}T^{-1}E[(9K\lambda^{-2s}+6K\lambda^{-s}|(\tilde{\boldsymbol{\theta}}_l-\boldsymbol{\theta})'\hat{\mathbb{X}}_{1,t}|+6K\lambda^{-s}|\boldsymbol{\theta}'(\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t})| \\
&+2|(\tilde{\boldsymbol{\theta}}_l-\boldsymbol{\theta})'\hat{\mathbb{X}}_{1,t}||(\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t})'\boldsymbol{\theta}|+|(\tilde{\boldsymbol{\theta}}_l-\boldsymbol{\theta})'\hat{\mathbb{X}_{1,t}}|^2+|(\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t})'\boldsymbol{\theta}|^2] \\
=&O(T^{-1}),
\end{aligned}
$$

which implies that A2 of Newey (1991) is satisfied, where the last line follows from the application of A3, A7 and the Monotone Convergence Theorem. To verify A3a of Newey (1991), consider $|T^{-1}\sum_{t=1}^{T}\psi_{\tau_1}(u_t(\tau_1))(\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})-\hat{\Upsilon}_t(\boldsymbol{\theta})| \leq T^{-1}\sum_{t=1}^{T}|\hat{\mathbb{X}}_{1,t}'(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})| \leq T^{-1}\sum_{t=1}^{T}\max_j|\hat{\mathbb{X}}_{1,t}^{(j)}|\sum_{j=1}^{\lambda}|\tilde{\boldsymbol{\theta}}^{(j)}-\boldsymbol{\theta}^{(j)}|$, where $\tilde{\boldsymbol{\theta}}^{(j)}$ denotes the j element of $\tilde{\boldsymbol{\theta}}$. Now, $\sum_{j=1}^{\lambda}|\tilde{\boldsymbol{\theta}}^{(j)}-\boldsymbol{\theta}^{(j)}| = \|\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}\|_1$ is a Manhattan norm while $T^{-1}\sum_{t=1}^{T}\max_j|\hat{\mathbb{X}}_{1,t}^{(j)}|$ is $O_p(1)$. These two conditions are sufficient for A3a of Newey (1991) to hold and thus (23) is $o(1)$ following Corollary 2.2 of Newey (1991).

Next, we show that (24) is $o(1)$. To do so, we follow Andrews (1987) and verify his assumptions A2b and A3.[14] Assumption A2b requires that $T^{-1}\sum_{t=1}^{T}\int_0^{\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})}(I(0<u_t(\tau_1)\leq s)ds$ satisfies pointwise strong law of large numbers. To do so, we check that $E[\int_0^{\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})}(I(0<u_t(\tau_1)\leq s)ds]<\infty$. Without loss of generality, assume that $\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}}) = (\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})'\hat{\mathbb{X}}_{1,t}+\boldsymbol{\theta}'(\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t})-\Gamma_t'i_3>0$. Also, by the almost sure convergence of $\hat{\mathbb{X}}_{1,t}$ and applying the Dominated Convergence Theorem ensured by A7, we may replace $(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})'\hat{\mathbb{X}}_{1,t}$ with $(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})'\mathbb{X}_{1,t}$ without affecting the end result. Likewise, we may drop $\boldsymbol{\theta}'(\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t})$ since the expectation of this term goes to zero by the Monotone Convergence Theorem. Therefore, considering $|(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})'\mathbb{X}_{1,t}|+|\Gamma_t'i_3| > (\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})'\hat{\mathbb{X}}_{1,t}-\Gamma_t'i_3>0$, we check that $E[\int_0^{|(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})'\mathbb{X}_{1,t}|+|\Gamma_t'i_3|}(I(0<u_t(\tau_1)\leq s)ds] \leq E[\int_0^{3K\lambda^{-s}+|(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})'\mathbb{X}_{1,t}|}E[(I(0<u_t(\tau_1)\leq s)|\mathbb{X}_{1,t}]ds] \leq C_f^{\max}E[\int_0^{3K\lambda^{-s}+|(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})'\mathbb{X}_{1,t}|}sds] = (C_f^{\max}/2)E[(3K\lambda^{-s}+|(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta})'\mathbb{X}_{1,t}|)^2] \leq C_f^{\max}E[3K\lambda^{-s}+\|\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}\|\max_j|\mathbb{X}_{1,t}^{(j)}|] = O(1)$ where $O(1)$ follows from A3 and A7. With bounded first moment, the pointwise strong law of large numbers follows. To verify

[14] Assumption A1 in Andrews (1987) requires that the parameter space be compact, which is A2 in this paper. Assumption A2a of Andrews (1987) , which we rephrase here, imposes that $\int_0^{\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})}(I(0<u_t(\tau_1)\leq s)ds$ is a random variable and for $\tilde{\boldsymbol{\theta}}\in\|\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}\|$ where $\|\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}\|$ is sufficiently small, $\sup_{\tilde{\boldsymbol{\theta}}}\int_0^{\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})}(I(0<u_t(\tau_1)\leq s)ds$ and $\inf_{\tilde{\boldsymbol{\theta}}}\int_0^{\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})}(I(0<u_t(\tau_1)\leq s)ds$ are random variables for all $\tilde{\boldsymbol{\theta}}\in\boldsymbol{\Theta}$.

A3 of Andrews (1987) , we need to show that for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, as $\sup_{\bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} \|\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}\| \to 0$, we have

$$\sup_{t \geq 1} |T^{-1} \sum_{t=1}^{T} (E[\int_0^{\sup_{\bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} \hat{\Upsilon}_t(\bar{\boldsymbol{\theta}})} (I(0 < u_t(\tau_1) \leq s)ds] - E[\int_0^{\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})} (I(0 < u_t(\tau_1) \leq s)ds])| \to 0. \tag{25}$$

We also have to verify the above with inf replacing sup, but the steps are similar once we demonstrate that the condition holds with sup. Arguing as before and considering $(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbb{X}_{1,t} - \Gamma_t' i_3 > 0$, we have

$$\begin{aligned}
& E[\int_0^{\sup_{\bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbb{X}_{1,t} - \Gamma_t' i_3} (I(0 < u_t(\tau_1) \leq s)ds] - E[\int_0^{(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbb{X}_{1,t} - \Gamma_t' i_3} (I(0 < u_t(\tau_1) \leq s)ds] \\
= & E[\int_{(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbb{X}_{1,t} - \Gamma_t' i_3}^{\sup_{\bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbb{X}_{1,t} - \Gamma_t' i_3} (I(0 < u_t(\tau_1) \leq s)ds] \\
\leq & E[\int_{(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbb{X}_{1,t} - 3K\lambda^{-s}}^{\sup_{\bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbb{X}_{1,t} + 3K\lambda^{-s}} E[(I(0 < u_t(\tau_1) \leq s)|\mathcal{F}_{1,t}]ds] \\
\leq & C_f^{\max} E[\int_{(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbb{X}_{1,t} - 3K\lambda^{-s}}^{\sup_{\bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbb{X}_{1,t} + 3K\lambda^{-s}} s ds] \\
= & \frac{C_f^{\max}}{2} E[(\sup_{\bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbb{X}_{1,t} + 3K\lambda^{-s})^2 - ((\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbb{X}_{1,t} - 3K\lambda^{-s})^2] \\
= & \frac{C_f^{\max}}{2} E[((\sup_{\bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} (\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}) + (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}))'\mathbb{X}_{1,t})(6K\lambda^{-s} + \sup_{\bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} (\bar{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})'\mathbb{X}_{1,t})] \\
\leq & \frac{C_f^{\max}}{2} E[2 \sup_{\bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} |(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta})'\mathbb{X}_{1,t}|(6K\lambda^{-s} + \sup_{\bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} |(\bar{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}})'\mathbb{X}_{1,t}|)] \\
\leq & \frac{C_f^{\max}}{2} E[2 \max_j |\mathbb{X}_{1,t}^{(j)}| \sup_{\bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} \|(\bar{\boldsymbol{\theta}} - \boldsymbol{\theta})\|(6K\lambda^{-s} + \max_j |\mathbb{X}_{1,t}^{(j)}| \sup_{\bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} \|\bar{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\|)] \\
\to & 0
\end{aligned}$$

where the last inequality follows from A7, $\sup_{\bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} \|\bar{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\| \leq 2 \sup_{\bar{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} \|\bar{\boldsymbol{\theta}} - \boldsymbol{\theta}\| \to 0$ and an application of the Monotone Convergence Theorem. Thus, we have verified the conditions of Andrews (1987) and the uniform law of large numbers follows. □

Proof of Proposition 1: Define $L(\tilde{\boldsymbol{\theta}}) = E[L_T(\tilde{\boldsymbol{\theta}})]$. Clearly, L is minimized at $\boldsymbol{\theta}$. Following the argument in Theorem 2.1 of Newey and McFadden (1994), we have $L(\hat{\boldsymbol{\theta}}) < L_T(\hat{\boldsymbol{\theta}}) + \delta/3 < L_T(\boldsymbol{\theta}) + 2\delta/3 < L(\boldsymbol{\theta}) + \delta$, where the first and third inequalities follow from the uniform law of large numbers (verified in Lemma 1), and the second inequality is due to $\hat{\boldsymbol{\theta}}$ being a minimizer of L_T. Focusing on the last term, express the summand in $L(\boldsymbol{\theta})$ using Knight's identity as $\rho_{\tau_1}(u_t(\tau_1) - \Gamma_t' i_3) - \rho_{\tau_1}(u_t(\tau_1)) = -\Gamma_t' i_3 \psi_{\tau_1}(u_t(\tau_1)) + \int_0^{\Gamma_t' i_3} \mathbb{I}(u_t(\tau_1) \leq s) - \mathbb{I}(u_t(\tau_1) \leq 0)ds$. Taking conditional expectations and using A3, we can show that $L(\boldsymbol{\theta}) = E[f(\eta \Gamma_t' i_3) i_3' \Gamma_t \Gamma_t' i_3] \leq 9 C_f^{\max} K \lambda^{-2s}$. Since $\lambda \to 0$ as $T \to 0$ such that $L(\boldsymbol{\theta}) \to 0$ also by the continuity of $L(.)$, we have $L(\hat{\boldsymbol{\theta}}) < \delta$ asymptotically for any arbitrary δ. □

Lemma 2. *(Stochastic Equicontinuity) For some $q \geq 1$,*

$$\sup_{\tilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} |\sqrt{T}\left(\mathbb{W}_{T,\tau_1}(\hat{\boldsymbol{\Upsilon}}(\tilde{\boldsymbol{\theta}})) - \mathbb{W}_{T,\tau_1}(0) - E[\mathbb{W}_{T,\tau_1}(\hat{\boldsymbol{\Upsilon}}(\tilde{\boldsymbol{\theta}})) - \mathbb{W}_{T,\tau_1}(0)]\right)| = o_p(\sqrt{\lambda/T}) \tag{31}$$

for $\kappa > 1/2$.

Proof: The proof employs a chaining argument. Consider the cubes $\|\tilde{\boldsymbol{\theta}}_j - \boldsymbol{\theta}\| \leq r_j\varepsilon$, where $r = 0, 1, \cdots, T^{1/2}$, $r_j = 0$ for $j = 0$ and $\varepsilon = (\sqrt{\lambda}/T)\log T^{\kappa}$ for some $\kappa > 0$. Let $\mathcal{G}$ be the collection of nested cubes, $\mathcal{G}^0 \in \mathcal{G}^1 \in \ldots \in \mathcal{G}_j \ldots \in \mathcal{G}$, with cardinality of order $O(T^{1/2})$ and j indexes the cubes. Let the center of cube j be $\boldsymbol{\theta}_j$ so that for every $\boldsymbol{\theta} \in \mathcal{G}_j$, $\|\boldsymbol{\theta} - \boldsymbol{\theta}_j\| < \varepsilon$. For $j = 1$, the center of the cube is $\boldsymbol{\theta}$. Define $s_t(\tilde{\boldsymbol{\theta}}) = -b'\mathbb{X}_{1,t}[\psi_\tau(u_t(\tau_1) - \hat{\boldsymbol{\Upsilon}}_t(\tilde{\boldsymbol{\theta}})) - \psi_\tau(u_t(\tau_1))]$ for each b such that $b'b$ is finite. Hence, (31) holds if

$$\sup_{\tilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}} |\frac{1}{\sqrt{T}}\sum_{t=1}^{T} b'(s_t(\tilde{\boldsymbol{\theta}}) - E[s_t(\tilde{\boldsymbol{\theta}})])| = o_p(1).$$

Now, by definition, $\{s_t, \mathcal{F}_{1,t}\}$ is an adapted stochastic sequence. Since $E[E[s_t(\tilde{\boldsymbol{\theta}})]|\mathcal{F}_{1,t-1}] = E[s_t(\tilde{\boldsymbol{\theta}})|\mathcal{F}_{1,t-1}]$ by smoothing, $s_t(\tilde{\boldsymbol{\theta}}) - E[s_t(\tilde{\boldsymbol{\theta}})]$ is a martingale difference sequence. For $\tilde{\boldsymbol{\theta}} \in \mathcal{G}^1$ and $b = i$, we have

$$\begin{aligned}
&E[s_t^2(\tilde{\boldsymbol{\theta}})|\mathcal{F}_{1,t-1}] \\
=&E[(i'\mathbb{X}_{1,t})^2(\psi_\tau(u_t(\tau_1) - \hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})) - \psi_\tau(u_t(\tau_1)))^2|\mathcal{F}_{1,t-1}] \\
\leq&(i'\mathbb{X}_{1,t})^2 E[(\psi_\tau(u_t(\tau_1) - \hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})) - \psi_\tau(u_t(\tau_1)))|\mathcal{F}_{1,t-1}] \\
\leq&\lambda^2 \max_j |\mathbb{X}_{1,t}^{(j)}|^2 C_f^{\max}(3K\lambda^{-s} + \mathbb{X}_{1,t}'(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) + (\hat{\mathbb{X}}_{1,t} - \mathbb{X}_{1,t})'(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) + (\hat{\mathbb{X}}_{1,t} - \mathbb{X}_{1,t})'\boldsymbol{\theta}) \\
\leq&\lambda^2 \max_j |\mathbb{X}_{1,t}^{(j)}|^2 C_f^{\max}(3K\lambda^{-s} + \mathbb{X}_{1,t}'(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) + (\hat{\mathbb{X}}_{1,t} - \mathbb{X}_{1,t})'(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) + (\hat{\mathbb{X}}_{1,t} - \mathbb{X}_{1,t})'\boldsymbol{\theta}) \\
\leq&\lambda^2 \max_j |\mathbb{X}_{1,t}^{(j)}|^2 C_f^{\max}(3K\lambda^{-s} + (\mathbb{X}_{1,t} + o_p(1))'(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) + O_p(T^{-1/2})) \\
\leq&\lambda^2 \max_j |\mathbb{X}_{1,t}^{(j)}|^2 C_f^{\max}(3K\lambda^{-s} + \max_j |\mathbb{X}_{1,t}^{(j)}|\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\|(1 + o_p(1)) + O_p(T^{-1/2})) \\
:=&p_t
\end{aligned}$$

The first inequality follows from the Cauchy-Schwarz inequality, the second inequality follows from A6 and the third inequality follows from the fact that $\psi_\tau^4 \leq \psi_\tau$. We apply the Hoeffding inequality for martingales proposed by Lee and Su (2002), that is,

$$P(|\sum_{t=1}^{T}(s_t - E[s_t])| > CT) \leq \exp\left(\frac{-C^2T^2}{2\sum_{t=1}^{T} p_t + \frac{2CT}{3}}\right). \tag{26}$$

Applying law of large numbers using A7, we have $T(T^{-1}\sum_{t=1}^{T} p_t) = TO_p(\lambda^{-s+2} + (\lambda^{1/2+2}/T^{1/2})\log T^{\kappa}) \leq O_p(T^{5a/2+1/2}\log T^{\kappa})$ where the inequality follows from setting $s = (1-a)/2a$ and $\lambda = O(T^a)$ as stated in A3. Let $C = \sqrt{2}T^{-b}\log T^{\kappa}$ be the constant and choose $-b = 3a/2 - 1$. Consequently, $P(|\sum_{t=1}^{T}(s_t(\tilde{\boldsymbol{\theta}}) -$

$E[s_t(\tilde{\boldsymbol{\theta}})])| > T^{-b+1}\log T^{\kappa}) \leq \exp(-\log T^{\kappa})$. In other words, $P(T^{-1}|\sum_{t=1}^{T}(s_t(\tilde{\boldsymbol{\theta}}) - E[s_t(\tilde{\boldsymbol{\theta}})])| > T^{-b}\log T^{\kappa}) \leq \exp(-\log T^{\kappa})$. Since $0 < a < 1/2$, we have $1/4 < b < 1$ which implies that $T^{-b}\log T^{\kappa} = o(1)$. Since the cardinality is $T^{1/2}$, stochastic equicontinuity will hold as long as $\kappa > -1/2$. Finally, note that $O(\sqrt{\lambda/T}) = O(T^{(a-1)/2})$ while $T^{-b} = T^{(3a-2)/2}$ dominates the logarithmic term. Since $(a-1)/2 - (3a-2)/2 = (1-2a)/2 > 0$, this implies that $T^{-b}\log T^{\kappa} = o(\sqrt{\lambda/T})$. Hence, the conclusion follows. $\square$

Proof of Proposition 2: Define $\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}}) = (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'\hat{\mathbb{X}}_{1,t} + \boldsymbol{\theta}'(\hat{\mathbb{X}}_{1,t} - \mathbb{X}_{1,t}) - \Gamma_t' i_3$. Without loss of generality, assume that $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\hat{\mathbb{X}}_{1,t} + \boldsymbol{\theta}'(\hat{\mathbb{X}}_{1,t} - \mathbb{X}_{1,t})$ Also, recall that

$$\begin{aligned} Y_{1,t} &= \boldsymbol{\theta}'\mathbb{X}_{1,t} + \Gamma_t' i_3 + u_t(\tau_1) \\ &= \boldsymbol{\theta}'\hat{\mathbb{X}}_{1,t} + \Gamma_t' i_3 + \boldsymbol{\theta}'(\mathbb{X}_{1,t} - \hat{\mathbb{X}}_{1,t}) + u_t(\tau_1) \end{aligned}$$

From now on, let $\boldsymbol{\theta}$ be the population parameter to simplify the notation. Denote the first order condition as

$$\hat{\mathbb{W}}_{T,\tau_1}(\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})) = T^{-1}\sum_{t=1}^{T}\hat{\mathbb{X}}_{1,t}\psi_{\tau_1}\left(u_t(\tau_1) - \hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})\right),$$

where $\hat{\mathbb{W}}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}})) = 0$ except for a finite number of points since $\hat{\boldsymbol{\theta}}$ is the minimizer. In addition, denote

$$\mathbb{W}_{T,\tau_1}(\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})) = T^{-1}\sum_{t=1}^{T}\mathbb{X}_{1,t}\psi_{\tau_1}\left(u_t(\tau_1) - \hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}})\right).$$

That is, the difference between $\hat{\mathbb{W}}_{T,\tau_1}$ and $\mathbb{W}_{T,\tau_1}$ is that the former multiplies ψ_{τ_1} with $\hat{\mathbb{X}}_{1,t}$ while the latter with $\mathbb{X}_{1,t}$. Expand $E[\hat{\mathbb{W}}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}))]$ around $\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}) = 0$:

$$\begin{aligned} &E[\hat{\mathbb{W}}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}))] \\ =&T^{-1}\sum_{t=1}^{T}E[\hat{\mathbb{X}}_{1,t}E[\psi_{\tau_1}(u_t(\tau_1) - \hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}))|\mathbb{X}_{1,t}]] \\ =&T^{-1}\sum_{t=1}^{T}E[\hat{\mathbb{X}}_{1,t}(F(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}})) - F(0))] \\ =&T^{-1}\sum_{t=1}^{T}E[\mathbb{X}_{1,t}(F(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}})) - F(0))] + T^{-1}\sum_{t=1}^{T}E[(\hat{\mathbb{X}}_{1,t} - \mathbb{X}_{1,t})(F(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}})) - F(0))] \end{aligned} \tag{27}$$

Now, $T^{-1}\sum_{t=1}^{T}E[(\hat{\mathbb{X}}_{1,t}^{(j)} - \mathbb{X}_{1,t}^{(j)})(F(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}})) - F(0))] = o(T^{-1/2})$. For the first term in (27), we apply the

Mean Value Theorem, thus obtaining

$$
\begin{aligned}
&E[\hat{\mathbb{W}}^{(j)}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}))] \\
=&T^{-1}\sum_{t=1}^{T}\Big(E[\mathbb{X}^{(j)}_{1,t}(F(-\Gamma' i_3)-F(0))]+E[\mathbb{X}^{(j)}_{1,t}f(\eta\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}))\mathbb{X}_{1,t}{}'](\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})+E[\mathbb{X}^{(j)}_{1,t}f(\eta\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}))\boldsymbol{\theta}](\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t})\Big)+o(T^{-1/2})
\end{aligned}
\tag{28}
$$

where $0<\eta<1$. Now, define

$$
\hat{\mathbb{F}}(\hat{\boldsymbol{\theta}})=\begin{pmatrix} f_1(\eta\hat{\Upsilon}_1(\hat{\boldsymbol{\theta}})) & 0 & \dots & 0 \\ 0 & f_2(\eta\hat{\Upsilon}_2(\hat{\boldsymbol{\theta}})) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & f_T(\eta\hat{\Upsilon}_T(\hat{\boldsymbol{\theta}})) \end{pmatrix}
$$

By A7, since $\hat{\mathbb{X}}_{1,t}$ is a smooth function of $\hat{w}_t$ which converges almost surely to w_t, $\hat{\mathbb{X}}_{1,t}$ also converges almost surely to $\mathbb{X}_{1,t}$. In addition, recall that the i^{th}-diagonal element of $\mathbb{F}(\hat{\boldsymbol{\theta}})$ is $f_i(\eta\Upsilon_i(\hat{\boldsymbol{\theta}}))$ while $\hat{\boldsymbol{\theta}}$ is consistent by Proposition 1. Consider $\hat{\mathbb{F}}(\hat{\boldsymbol{\theta}})=\hat{\mathbb{F}}(\hat{\boldsymbol{\theta}})-\mathbb{F}(\hat{\boldsymbol{\theta}})+\mathbb{F}(\hat{\boldsymbol{\theta}})$. Then, applying the Slutsky Theorem, we have $\hat{\mathbb{F}}(\hat{\boldsymbol{\theta}})-\mathbb{F}(\hat{\boldsymbol{\theta}})\xrightarrow{p}0$ by A7 and $\mathbb{F}(\hat{\boldsymbol{\theta}})\xrightarrow{p}\mathbb{F}(\boldsymbol{\theta})$ by Proposition 1. Hence, $\hat{\mathbb{F}}(\hat{\boldsymbol{\theta}})=\mathbb{F}(\boldsymbol{\theta})+o_p(1)$. Recall that $Q_1(\boldsymbol{\theta})=E[\mathbb{X}_1'\mathbb{F}(\boldsymbol{\theta})\mathbb{X}_1/T]$. To simplify the notation further, let Q_1 and $\mathbb{F}$ correspond to the values where the population parameter $\boldsymbol{\theta}$ is in the argument. We may then rewrite (28) as

$$
\begin{aligned}
E[\hat{\mathbb{W}}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}))]=&Q_{1,T}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})-T^{-1}\sum_{t=1}^{T}E[f_t(\eta\Upsilon_t(\boldsymbol{\theta}))\mathbb{X}_{1,t}]\Gamma_t' i_3+T^{-1}\sum_{t=1}^{T}E[f_t(\eta\Upsilon_t(\boldsymbol{\theta}))\mathbb{X}_{1,t}\boldsymbol{\theta}'](\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t}) \\
&+o(T^{-1/2}).
\end{aligned}
\tag{29}
$$

Hence, we rearrange (29) to obtain

$$
\begin{aligned}
&\hat{\boldsymbol{\theta}}-\boldsymbol{\theta} \\
=&Q_1^{-1}E[\hat{\mathbb{W}}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}))]+Q_1^{-1}T^{-1}\sum_{t=1}^{T}E[f_t(\eta\Upsilon_t(\boldsymbol{\theta}))\mathbb{X}_{1,t}]\Gamma_t' i_3-Q_1^{-1}T^{-1}\sum_{t=1}^{T}E[f_t(\eta\Upsilon_t(\boldsymbol{\theta}))\mathbb{X}_{1,t}\boldsymbol{\theta}'](\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t}) \\
&+o(T^{-1/2}) \\
=&Q_1^{-1}\mathbb{W}_{1,T}(0)+Q_1^{-1}T^{-1}\sum_{t=1}^{T}E[f_t(\eta\Upsilon_t(\boldsymbol{\theta}))\mathbb{X}_{1,t}]\Gamma_t' i_3-Q_1^{-1}T^{-1}\sum_{t=1}^{T}E[f_t(\eta\Upsilon_t(\boldsymbol{\theta}))\mathbb{X}_{1,t}\boldsymbol{\theta}'](\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t}) \\
&+\underbrace{Q_1^{-1}(\mathbb{W}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}))-\mathbb{W}_{T,\tau_1}(0)-E[\mathbb{W}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}))])}_{A}+\underbrace{Q_1^{-1}(\hat{\mathbb{W}}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}))-\mathbb{W}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}})))}_{B} \\
&+o(T^{-1/2})
\end{aligned}
\tag{30}
$$

Following stochastic equicontinuity established by Lemma 2,the rate for (A) is $o_p(\sqrt{\lambda}/T)$, since

$$\sup_{\tilde{\boldsymbol{\theta}}\in\boldsymbol{\Theta}}\|\sqrt{T}\left(\mathbb{W}_{T,\tau_1}(\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}}))-\mathbb{W}_{T,\tau_1}(0)-E[\mathbb{W}_{T,\tau_1}(\hat{\Upsilon}_t(\tilde{\boldsymbol{\theta}}))-\mathbb{W}_{T,\tau_1}(0)]\right)\|=o_p(\sqrt{\lambda/T}), \tag{31}$$

where we have used the fact that $E[\mathbb{W}_{T,\tau_1}(0)]=0$. To establish the rate for (B), observe that $\hat{\mathbb{W}}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}))-\mathbb{W}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}))\leq\frac{1}{T}\sum_{t=1}^T(\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t})\psi_{\tau_1}(u_t(\tau_1)-\hat{\boldsymbol{\Upsilon}}_t(\hat{\boldsymbol{\theta}}))\leq\max_t\|\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t}\|\frac{1}{T}\sum_{t=1}^T\psi_{\tau_1}(u_t(\tau_1)-\hat{\boldsymbol{\Upsilon}}_t(\hat{\boldsymbol{\theta}}))$, where $\max_t\|\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t}\|=o_p(1)$ by A7.[15]

To apply Chebyshev inequality, check that $Var\left(\frac{1}{T}\sum_{t=1}^T\psi_{\tau_1}(u_t(\tau_1)-\hat{\boldsymbol{\Upsilon}}_t(\hat{\boldsymbol{\theta}}))\right)\leq T^{-2}\sum_{t=1}^T[f(\eta\hat{\boldsymbol{\Upsilon}}_t(\hat{\boldsymbol{\theta}}))((\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\mathbb{X}_{1,t}+\boldsymbol{\theta}'(\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t})-\Gamma_t'i_3)]\leq T^{-2}\sum_{t=1}^T C_f^{\max}(\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|\|\mathbb{X}_{1,t}\|+\|\boldsymbol{\theta}\|\|\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t}\|+K\lambda^{-s}\|=o_p(T^{-1})$, where the second inequality follows from A3 and A4 and the last equality follows from A6 and A7. Collecting the results, we may conclude that $\hat{\mathbb{W}}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}))-\mathbb{W}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\boldsymbol{\theta}}))=o_p(T^{-1})$.

We will now establish the rate of convergence for $\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|$. First consider

$$\begin{aligned}\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}=&-Q_1^{-1}T^{-1}\sum_{t=1}^T\mathbb{X}_{1,t}\psi_{\tau_1}(u_t(\tau_1))+Q_1^{-1}T^{-1}\sum_{t=1}^T E[f_t(\eta\Upsilon_t(\boldsymbol{\theta}))\mathbb{X}_{1,t}]\Gamma_t'i_3\\&-Q_1^{-1}T^{-1}\sum_{t=1}^T E[f_t(\eta\Upsilon_t(\boldsymbol{\theta}))\mathbb{X}_{1,t}\boldsymbol{\theta}'](\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t})+o_p(\sqrt{\lambda/T}).\end{aligned} \tag{32}$$

Recalling that $\Gamma=[\Gamma_0\ \Gamma_1\ \Gamma_2]$, we may express $Q_1^{-1}T^{-1}\sum_{t=1}^T E[f_t(\eta\Upsilon_t(\boldsymbol{\theta}))\mathbb{X}_{1,t}]\Gamma_{0,t}=Q_1^{-1}T^{-1}E[\mathbb{X}_1{}'\mathbb{F}]\Gamma_0$ using A3. Using the fact that $\|\Gamma_0\|=(\Gamma_0{}'\Gamma_0)^{1/2}\leq K\lambda^{1/2-s}$, consider $\|Q_1^{-1}T^{-1}E[\mathbb{X}_1{}'\mathbb{F}]\Gamma_0\|\leq\|Q_1^{-1}T^{-1}E[\mathbb{X}_1{}'\mathbb{F}]\|\|\Gamma_0\|$. In addition, since Jensen's inequality implies $\|Q_1^{-1}T^{-1}E[\mathbb{X}_1{}'\mathbb{F}]\|^2\leq E\|Q_1^{-1}T^{-1}\mathbb{X}_1{}'\mathbb{F}\|^2$, consider

$$\begin{aligned}\|Q_1^{-1}T^{-1}\mathbb{X}_1{}'\mathbb{F}\|^2=&T^{-2}tr(\mathbb{F}\mathbb{X}_1Q_1^{-1}Q_1^{-1}\mathbb{X}_1{}'\mathbb{F})\\\leq&C_f^{\max}K_{\min}(Q_1)^{-2}T^{-1}tr(\mathbb{X}_1{}'\mathbb{F}\mathbb{X}_1/T)\\\leq&C_f^{\max}K_{\min}(Q_1)^{-2}T^{-1}K_{\max}(Q_{1,T})tr(I_\lambda)\\=&O_p(\lambda/T),\end{aligned}$$

which follows from the assumption that $K_{\max}(Q_{1,T})=O_p(1)$. Therefore, $\|Q_1^{-1}T^{-1}\sum_{t=1}^T E[f_t(\eta\Upsilon_t(\boldsymbol{\theta}))\mathbb{X}_{1,t}]\Gamma_{0,t}\|=O(\sqrt{\lambda^{2-2s}/T})$. Repeating with Γ_1 and Γ_2, we have $\|Q_1^{-1}T^{-1}\sum_{t=1}^T E[f_t(\eta\Upsilon_t(\boldsymbol{\theta}))\mathbb{X}_{1,t}]\Gamma_t{}'i_3\|=O(\sqrt{\lambda^{2-2s}/T})$.

[15] Intuitively, letting j denote the element of the design matrix, $\max_t\|\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t}\|\leq\lambda\max_t\max_j|\hat{\mathbb{X}}_{1,t}^j-\mathbb{X}_{1,t}^j|=\lambda o_p(T^{-1/2})$. But $\lambda\leq O_p(T^{2/5})$ by A3. Hence, $\max_t\|\hat{\mathbb{X}}_{1,t}-\mathbb{X}_{1,t}\|=o_p(1)$.

Next, consider the fact that $Q_1^{-1}T^{-1}\sum_{t=1}^{T}\mathbb{X}_{1,t}\psi_{\tau_1}(u_t(\tau_1)) \leq Q_1^{-1}T^{-1}\sum_{t=1}^{T}\mathbb{X}_{1,t}$. Since

$$\begin{aligned}\|Q_1^{-1}T^{-1}\sum_{t=1}^{T}\mathbb{X}_{1,t}\|^2 =& T^{-2}tr(Q_1^{-1}\sum_{t=1}^{T}\mathbb{X}_{1,t}\mathbb{X}_{1,t}'Q_1^{-1})\\ \leq& T^{-1}K_{\min}(Q_1)^{-2}tr(\sum_{t=1}^{T}\mathbb{X}_{1,t}(\mathbb{X}_{1,t}'/T)\\ =& T^{-1}K_{\min}(Q_1)^{-2}K_{\max}(Q_{1,T})tr(I_\lambda)\\ =& O_p(\lambda/T),\end{aligned}$$

this implies that $\|Q_1^{-1}T^{-1}\sum_{t=1}^{T}\mathbb{X}_{1,t}\psi_{\tau_1}(u_t(\tau_1))\| = O_p(\sqrt{\lambda/T})$.

Finally, consider $Q_1^{-1}T^{-1}\sum_{t=1}^{T}E[f_t(\eta\Upsilon_t(\boldsymbol{\theta}))\mathbb{X}_{1,t}](\hat{\mathbb{X}}_{1,t}'-\mathbb{X}_{1,t}')\boldsymbol{\theta} \leq \|Q_1^{-1}T^{-1}\sum_{t=1}^{T}E[f_t(\eta\Upsilon_t(\boldsymbol{\theta}))\mathbb{X}_{1,t}](\hat{\mathbb{X}}_{1,t}'-\mathbb{X}_{1,t}')\|\|\boldsymbol{\theta}\|$. Now, $\|\boldsymbol{\theta}\| \leq \max_j \theta^{(j)}(i_\lambda' i_\lambda)^{1/2} = O(\sqrt{\lambda})$ while $\|Q_1^{-1}T^{-1}\sum_{t=1}^{T}E[f_t(\eta\Upsilon_t(\boldsymbol{\theta}))\mathbb{X}_{1,t}](\hat{\mathbb{X}}_{1,t}'-\mathbb{X}_{1,t}')\| = O_p(T^{-1/2})$, thus $\|Q_1^{-1}T^{-1}\sum_{t=1}^{T}E[f_t(\eta\Upsilon_t(\boldsymbol{\theta})|)\mathbb{X}_{1,t}](\hat{\mathbb{X}}_{1,t}'-\mathbb{X}_{1,t}')\boldsymbol{\theta}\| = O_p(\sqrt{\lambda/T})$. This result, combining with the above, and applying the triangular inequality establishes the proposition. □

Lemma 3. *Under A1-A7,* $T^{-1/2}\sum_{t=1}^{T}\mathbb{X}_{1,t}\psi_{\tau_1}(Y_t-\boldsymbol{\theta}'\mathbb{X}_{1,t}) \Rightarrow N(0,\tau_1(1-\tau_2)D_1)$.

Proof: Let c be a fixed vector of unit length and consider $T^{-1/2}\sum_{t=1}^{T}c'\mathbb{X}_{1,t}\psi_{\tau_1}(Y_t-\boldsymbol{\theta}'\mathbb{X}_{1,t})$. Consider the sum of the variance $\Psi_T^2 = \sum_{t=1}^{T}Var(c'\mathbb{X}_{1,t}\psi_{\tau_1}(Y_t-\boldsymbol{\theta}'\mathbb{X}_{1,t}))$. Now,

$$\begin{aligned}&Var\left(c'\mathbb{X}_{1,t}\psi_{\tau_1}(Y_t-\boldsymbol{\theta}'\mathbb{X}_{1,t})\right)\\ =&E[|c'\mathbb{X}_{1,t}|^2\psi_{\tau_1}(Y_t-\boldsymbol{\theta}'\mathbb{X}_{1,t})^2]-(E[c'\mathbb{X}_{1,t}\psi_{\tau_1}(Y_t-\boldsymbol{\theta}'\mathbb{X}_{1,t})])^2\\ \geq&E[|c'\mathbb{X}_{1,t}|^2E[\psi_{\tau_1}(Y_t-\boldsymbol{\theta}'\mathbb{X}_{1,t})^2|\mathbb{X}_{1,t}]]-E[|c'\mathbb{X}_{1,t}|^2|F_t(\Gamma_t'i_3)-\tau_1|^2]\\ =&E[|c'\mathbb{X}_{1,t}|^2F_t(\Gamma_t'i_3)(1-F_t(\Gamma_t'i_3))],\end{aligned}$$

where the second last line follows from Jensen's inequality and the law of iterated expectations. Now, using Minkowski's inequality and A6, we can show that $E|c'\mathbb{X}_{1,t}|^2 = O(\lambda^2)$. In addition, since $E|c'\mathbb{X}_{1,t}|^2 > 0$, there is a positive constant L such that $E|c'\mathbb{X}_{1,t}|^2 > L\lambda^2$. Hence, $\Psi_T^2 \geq \min_t TE[|c'\mathbb{X}_{1,t}|^2F_t(\Gamma_t'i_3)(1-F_t(\Gamma_t'i_3))] \geq LT\lambda^2$, so that

$$\begin{aligned}\sum_{t=1}^{T}E\left[\frac{|c'\mathbb{X}_{1,t}\psi_{\tau_1}(Y_t-\boldsymbol{\theta}'\mathbb{X}_{1,t})|^2}{\Psi_T^2}\mathbb{I}\left(\left|\frac{c'\mathbb{X}_{1,t}\psi_{\tau_1}(Y_t-\boldsymbol{\theta}'\mathbb{X}_{1,t})}{\Psi_T}\right|>\epsilon\right)\right] \leq& \sum_{t=1}^{T}E\left[\frac{|c'\mathbb{X}_{1,t}|^{2+\delta}}{L^{1+\delta}T^{1+\delta}\lambda^{2+\delta}\epsilon^{1+\delta}}\right]\\ \leq& T\frac{\lambda^{2+\delta}\max_j c_j^{2+\delta}\Delta}{L^{1+\delta}T^{1+\delta}\lambda^{2+\delta}\epsilon^{1+\delta}} \to 0\end{aligned}$$

since $\delta > 0$. Therefore, by the Lindeberg-Feller Central Limit Theorem $T^{-1/2}\sum_{t=1}^{T}c'\mathbb{X}_{1,t}\psi_{\tau_1}(Y_t-\boldsymbol{\theta}'\mathbb{X}_{1,t})$ converges to a normal distribution and so does $T^{-1/2}\sum_{t=1}^{T}c\mathbb{X}_{1,t}\psi_{\tau_1}(Y_t-\boldsymbol{\theta}'\mathbb{X}_{1,t})$ by the Cramer-Wold device.

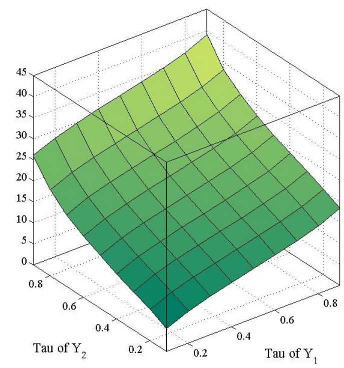
□

Proof of Proposition 3: This follows from Lemma 3, the assumption that $Q_{3,T}$ converges to a full column rank matrix Q_3, and the fact that $\sqrt{T}(\hat{\gamma}(\tau_2) - \gamma(\tau_2))$ is asymptotically normal. □

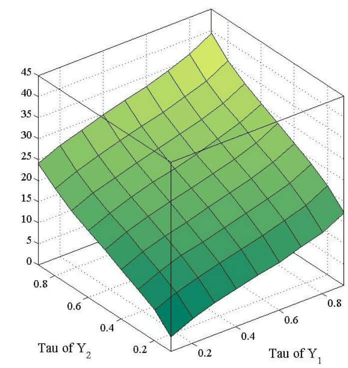
Figure 1

Estimated α_2 in Case 1.

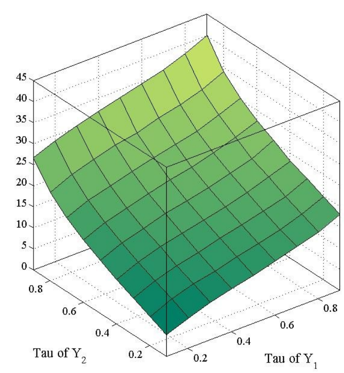
This figure shows the estimates of $\hat{\alpha}_2$ using the regression function of $H = \alpha_0 + \alpha_1 X_1 + \alpha_2 Y_2 + \sum_{i=1}^{I} \varphi_i \frac{\hat{w}(\tau_2)^i}{I!} Y_2$ when the true data generating process is $Y_1 = \alpha_0 + \alpha_1 X_1 + (\alpha_2 + \delta(\lambda e^{w} + u))Y_2$, where $(\alpha_0, \alpha_1, \alpha_2, \delta, \lambda) = (3, 4, 4, 5, 3)$, $(\beta_1, \beta_2, \beta_3) = (1, 2, 3)$, $X_1 \sim t_3$, $X_2 \sim N(15, 2)$, $w \sim N(0, 0.5)$ and $u \sim N(0, 1)$. Panel A shows the true parameter value $\alpha_2 + \delta(\lambda e^{F_w^{-1}(\tau_2)} + F_u^{-1}(\tau_1))$.



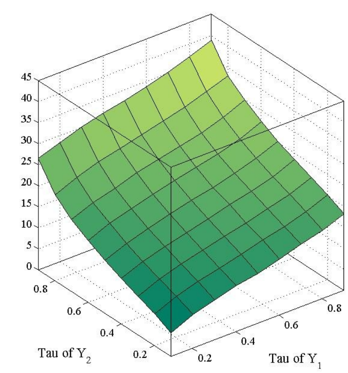
A. True



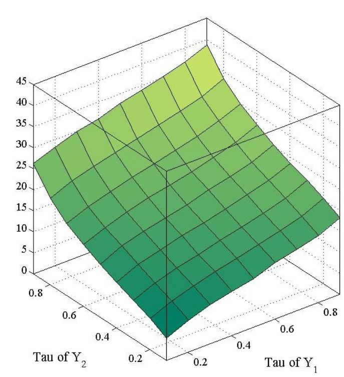
B. Linear Expansion



C. Quadratic Expansion



D. Cubic Expansion

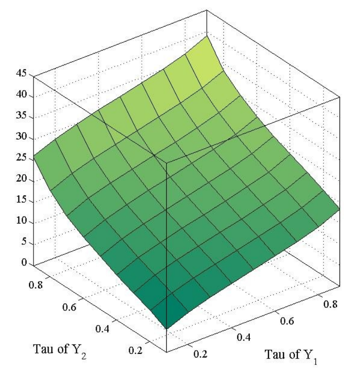


E. Quartic Expansion

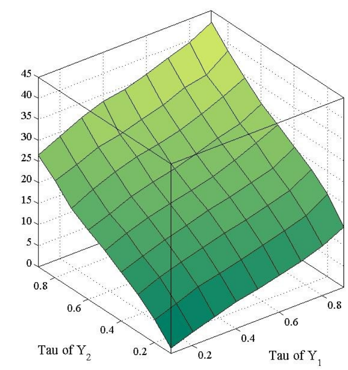
Figure 2

Estimated α_2 in Case 2.

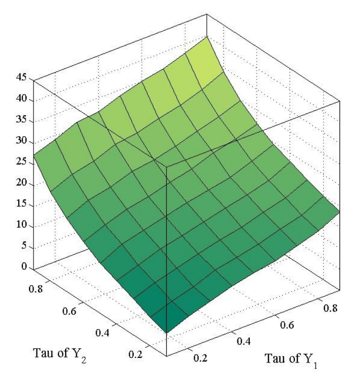
This figure shows the estimates of $\hat{\alpha}_2$ using the regression function of $H = \alpha_0 + \alpha_1 X_1 + \alpha_2 Y_2 + \sum_{i=1}^{I} \varphi_{i,0} \frac{\hat{w}(\tau_2)^i}{I!} + \sum_{i=1}^{I} \varphi_{i,1} \frac{\hat{w}(\tau_2)^i}{I!} X_1 + \sum_{i=1}^{I} \varphi_{i,2} \frac{\hat{w}(\tau_2)^i}{I!} Y_2$ when the true data generating process is $Y_1 = \alpha_0 + \alpha_1 X_1 + (\alpha_2 + \delta(\lambda e^{w} + u))Y_2$, where $(\alpha_0, \alpha_1, \alpha_2, \delta, \lambda) = (3, 4, 4, 5, 3)$, $(\beta_1, \beta_2, \beta_3) = (1, 2, 3)$, $X_1 \sim t_3$, $X_2 \sim N(15, 2)$, $w \sim N(0, 0.5)$ and $u \sim N(0, 1)$. Panel A shows the true parameter value $\alpha_2 + \delta(\lambda e^{F_w^{-1}(\tau_2)} + F_u^{-1}(\tau_1))$.



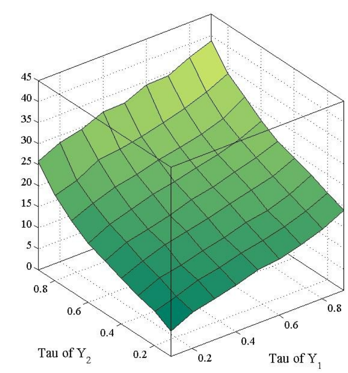
A. True



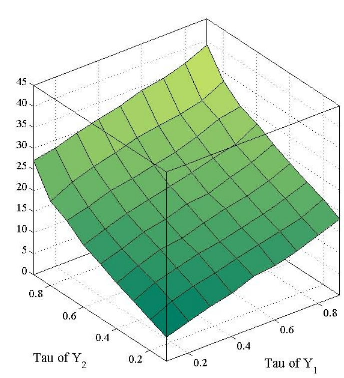
B. Linear Expansion



C. Quadratic Expansion



D. Cubic Expansion

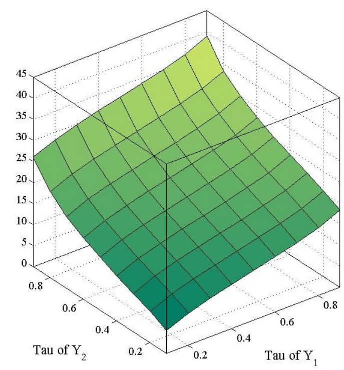


E. Quartic Expansion

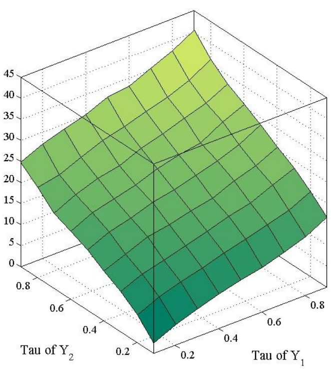
Figure 3

Estimated α_2 in Case 3.

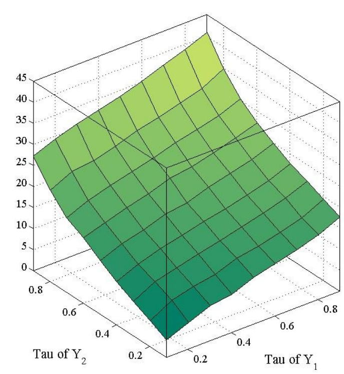
This figure shows the estimates of $\hat{\alpha}_2$ using the regression function of $H = \alpha_0 + \alpha_1 X_1 + \alpha_2 Y_2 + \sum_{i=1}^{I} \varphi_i \frac{\hat{w}(\tau_2)^i}{I!} Y_2$ when the true data generating process is $Y_1 = (\alpha_0 + \tilde{\delta}(\tilde{\lambda} e^w + u)) + (\alpha_1 + \tilde{\delta}(\tilde{\lambda} e^w + u))X_1 + (\alpha_2 + \delta(\lambda e^w + u))Y_2$, where $(\alpha_0, \alpha_1, \alpha_2, \delta, \lambda, \tilde{\delta}, \tilde{\lambda}) = (3, 4, 4, 5, 3, 10, 5)$, $(\beta_1, \beta_2, \beta_3) = (1, 2, 3)$, $X_1 \sim t_3$, $X_2 \sim N(15, 2)$, $w \sim N(0, 0.5)$ and $u \sim N(0, 1)$. Panel A shows the true parameter value $\alpha_2 + \delta(\lambda e^{F_w^{-1}(\tau_2)} + F_u^{-1}(\tau_1))$.



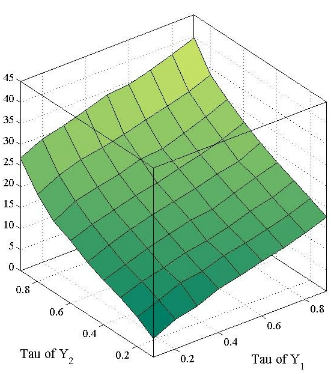
A. True



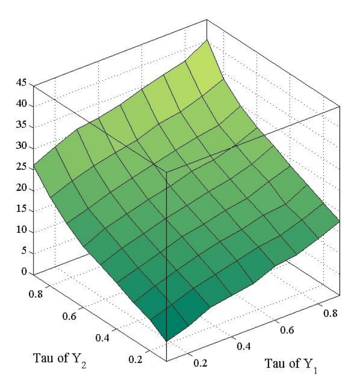
B. Linear Expansion



C. Quadratic Expansion



D. Cubic Expansion

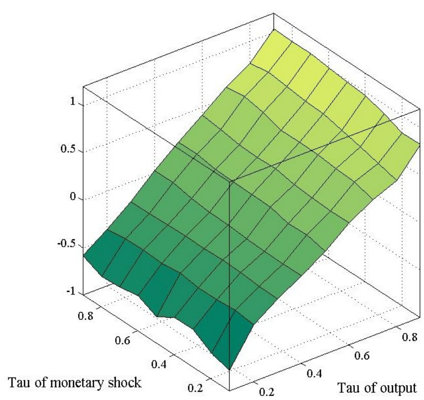


E. Quartic Expansion

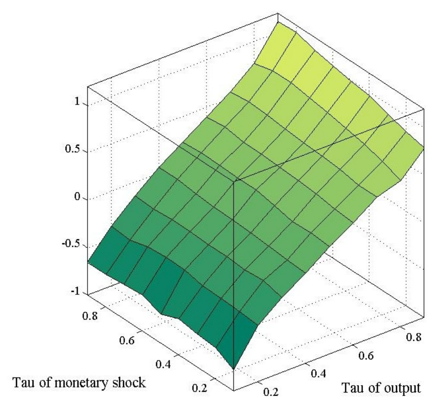
Figure 4

Intercept Term - M1 Money Supply.

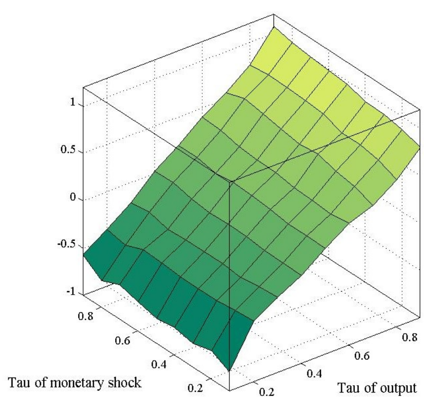
This figure shows the estimated intercept in the output process equation based on a cubic (Panels A and B) or quartic (Panels C and D) regression model. The monetary instrument is M1 money supply. The cubic or quartic expansion involves the intercept term, the first four lags of the output growth and the lag of the first difference Treasury yield. "Lags" refers to the number of output lags used in the output process equation.



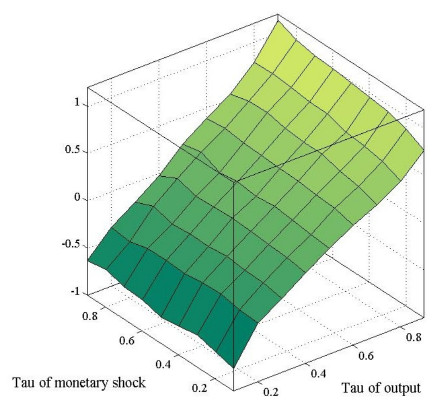
A. Cubic, 8 Lags



B. Cubic, 12 Lags



C. Quartic, 8 Lags

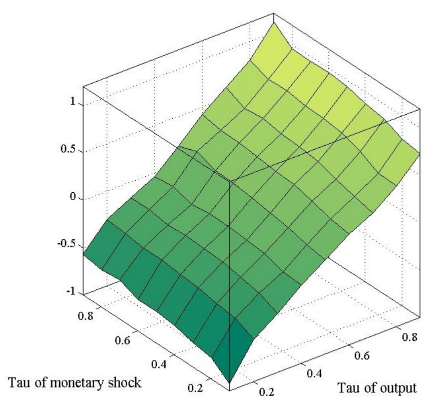


D. Quartic, 12 Lags

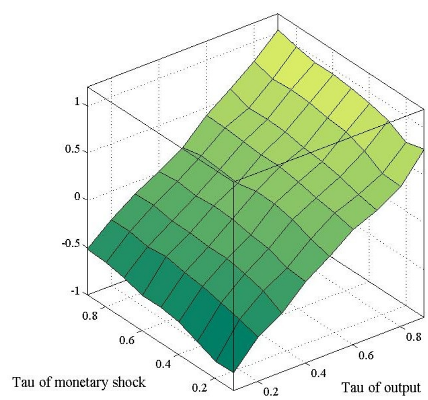
Figure 5

Intercept Term - M2 Money Supply.

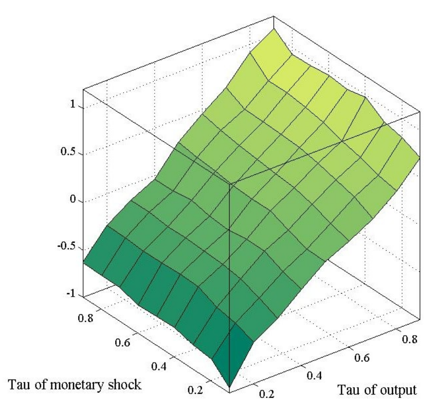
This figure shows the estimated intercept in the output process equation based on a cubic (Panels A and B) or quartic (Panels C and D) regression model. The monetary instrument is M1 money supply. The cubic or quartic expansion involves the intercept term, the first four lags of the output growth and the lag of the first difference Treasury yield. “Lags” refers to the number of output lags used in the output process equation.



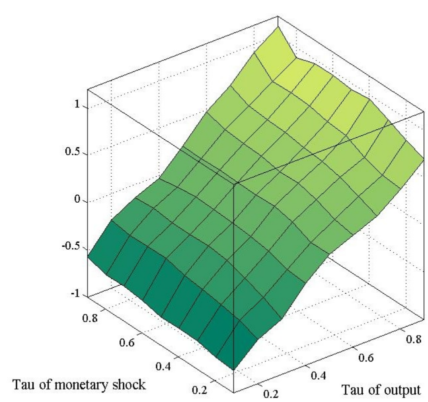
A. Cubic, 8 Lags



B. Cubic, 12 Lags



C. Quartic, 8 Lags

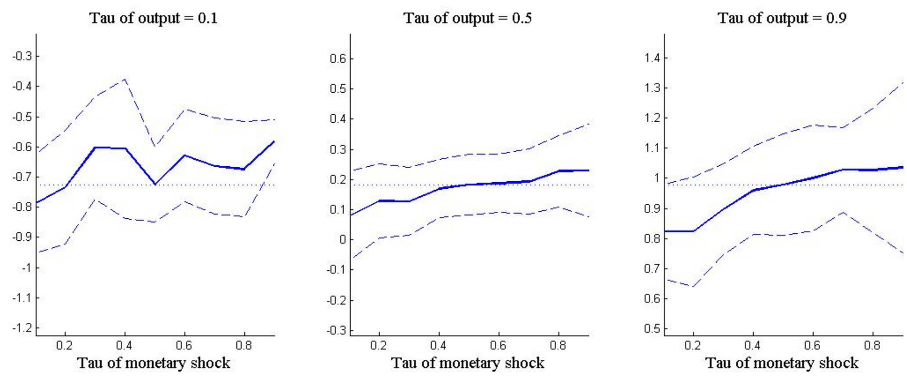


D. Quartic, 12 Lags

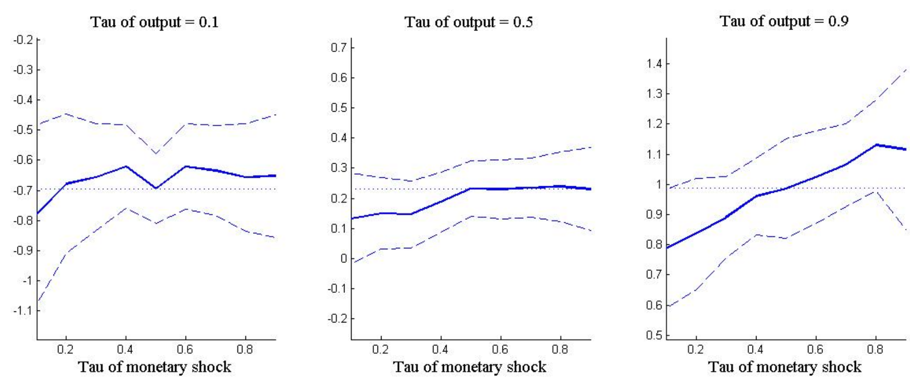
Figure 6

Cross-Section of Intercept Surface - Cubic Regression & M1 Money Supply.

This figure plots the cross-section of Panels A and B in Figure 4 for the 10th, 50th and 90th percentile output sections. The horizontal line plots the median of the monetary shock. The dash lines are the two standard deviation bands. The estimated values are reported from the cubic regression model.



A. 8 Lags

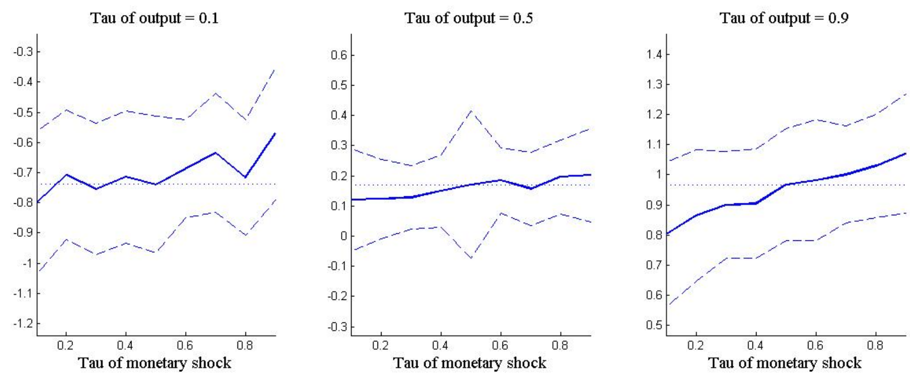


B. 12 Lags

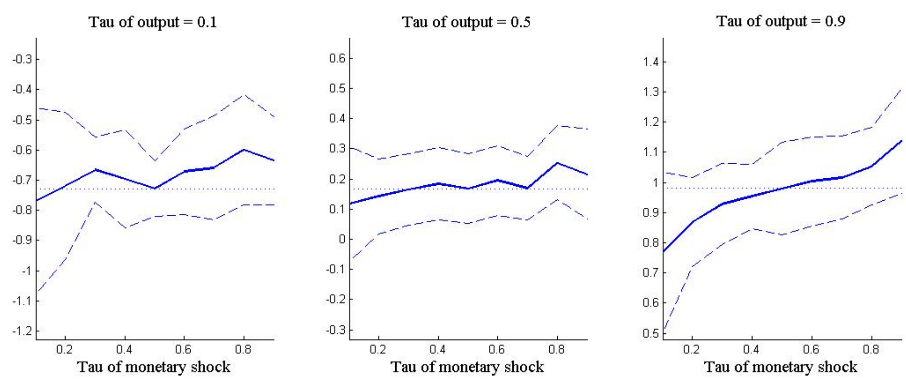
Figure 7

Cross-Section of Intercept Surface - Quartic Regression & M1 Money Supply.

This figure plots the cross-section of Panels C and D in Figure 4 for the 10th, 50th and 90th percentile output sections. The horizontal line plots the median of the monetary shock. The dash lines are the two standard deviation bands. The estimated values are reported from the quartic regression model.



A. 8 Lags

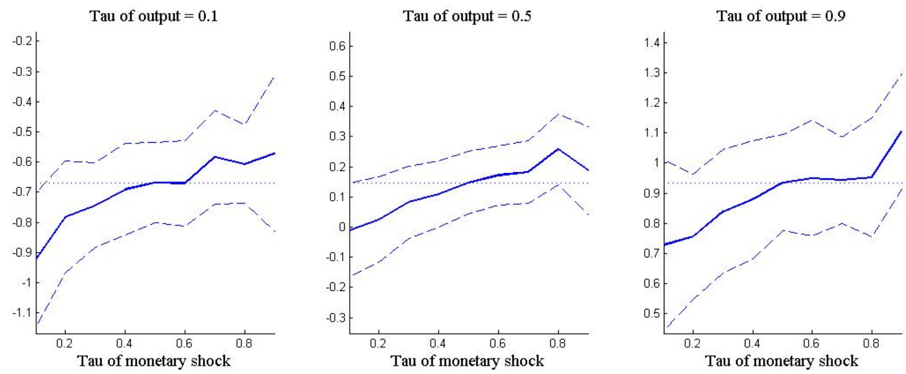


B. 12 Lags

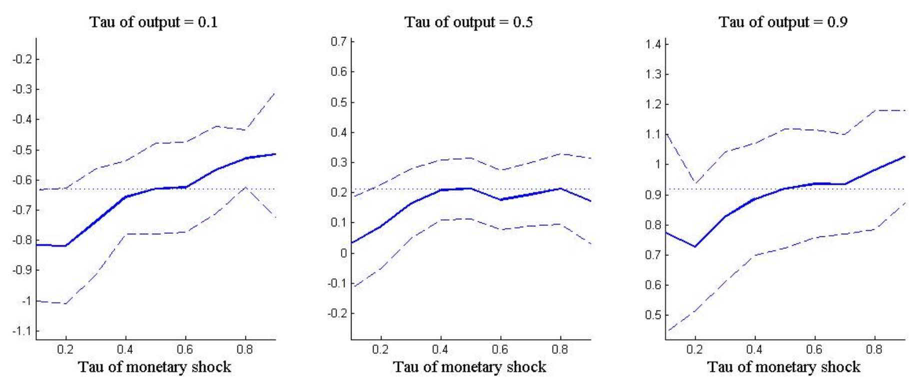
Figure 8

Cross-Section of Intercept Surface - Cubic Regression & M2 Money Supply.

This figure plots the cross-section of Panels A and B in Figure 5 for the 10th, 50th and 90th percentile output sections. The horizontal line plots the median of the monetary shock. The dash lines are the two standard deviation bands. The estimated values are reported from the cubic regression model.



A. 8 Lags

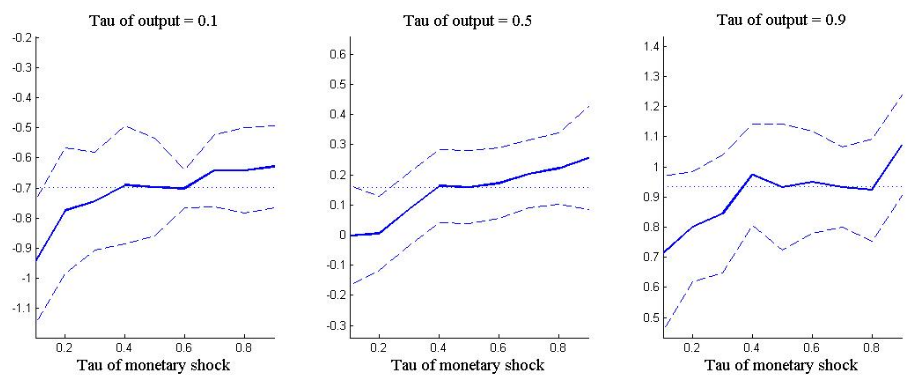


B. 12 Lags

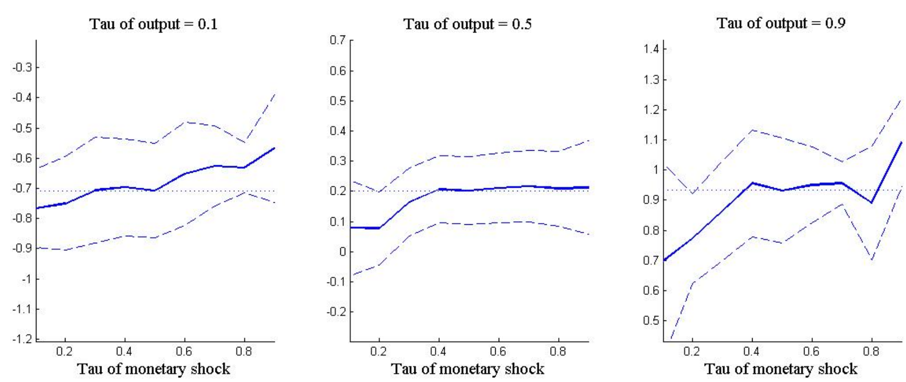
Figure 9

Cross-Section of Intercept Surface - Quartic Regression & M2 Money Supply.

This figure plots the cross-section of Panels C and D in Figure 5 for the 10th, 50th and 90th percentile output sections. The horizontal line plots the median of the monetary shock. The dash lines are the two standard deviation bands. The estimated values are reported from the quartic regression model.



A. 8 Lags

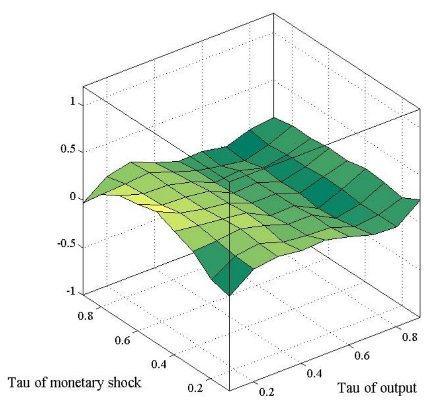


B. 12 Lags

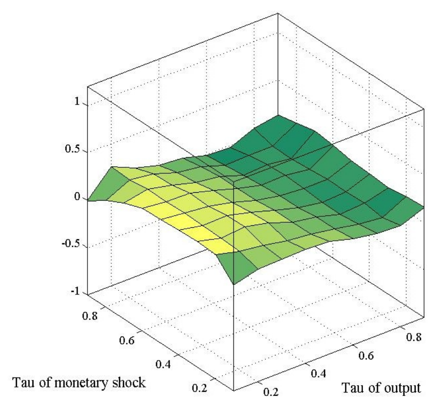
Figure 10

Slope of the First Lag of Output Growth - M1 Money Supply.

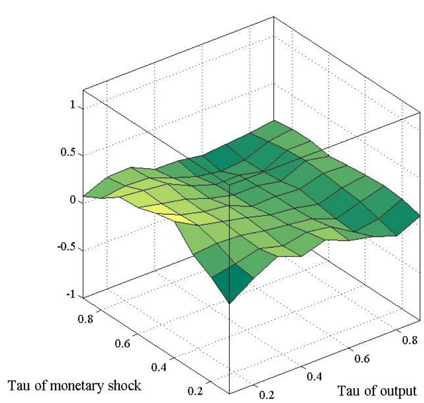
This figure shows the estimated slope of the first lag of output growth in the output process equation based on a cubic (Panels A and B) or quartic (Panels C and D) regression model. The monetary instrument is M1 money supply. The cubic or quartic expansion involves the intercept term, the first four lags of the output growth and the lag of the first difference Treasury yield. "Lags" refers to the number of output lags used in the output process equation.



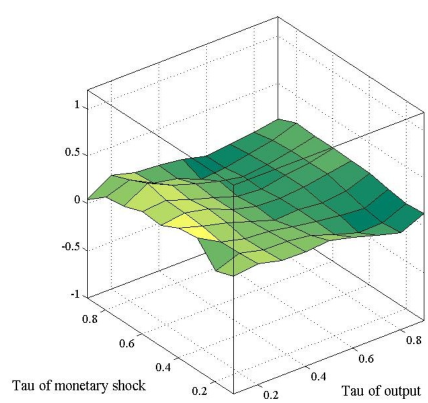
A. Cubic, 8 Lags



B. Cubic, 12 Lags



C. Quartic, 8 Lags

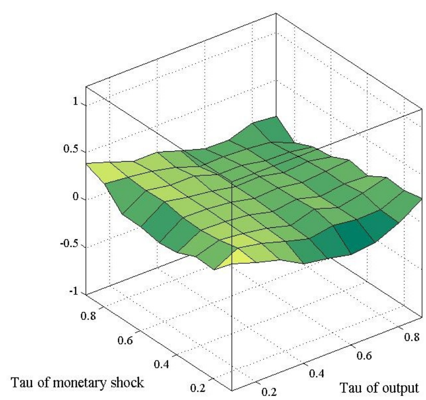


D. Quartic, 12 Lags

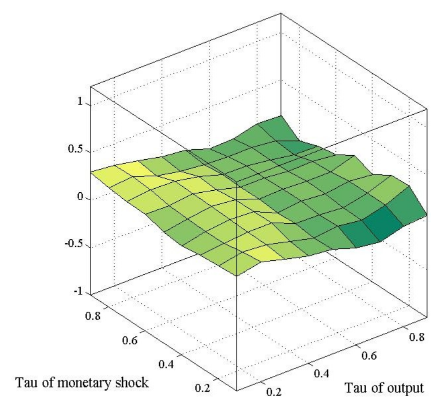
Figure 11

Slope of the First Lag of Output Growth - M2 Money Supply.

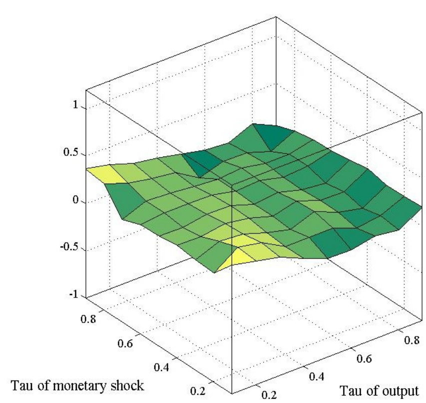
This figure shows the estimated slope of the first lag of output growth in the output process equation based on a cubic (Panels A and B) or quartic (Panels C and D) regression model. The monetary instrument is M2 money supply. The cubic or quartic expansion involves the intercept term, the first four lags of the output growth and the lag of the first difference Treasury yield. "Lags" refers to the number of output lags used in the output process equation.



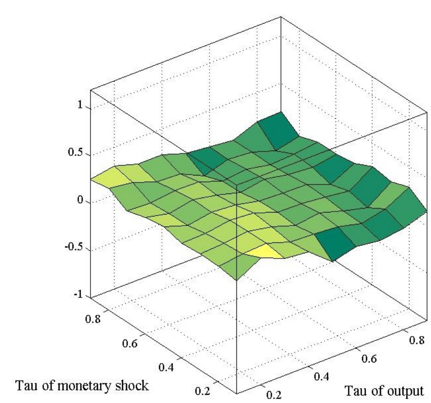
A. Cubic, 8 Lags



B. Cubic, 12 Lags



C. Quartic, 8 Lags

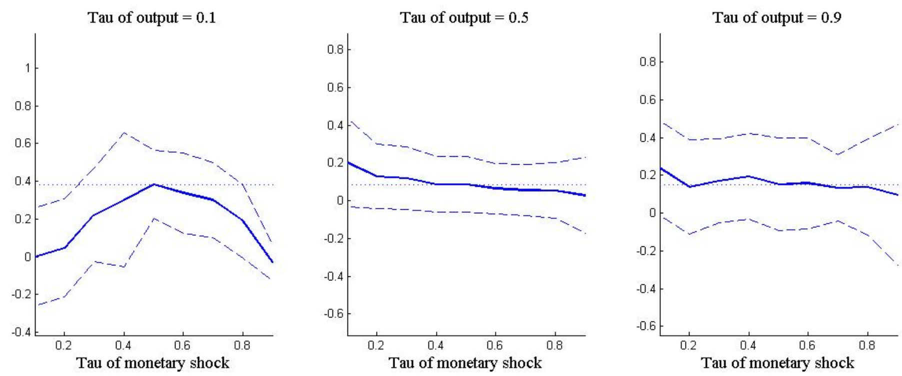


D. Quartic, 12 Lags

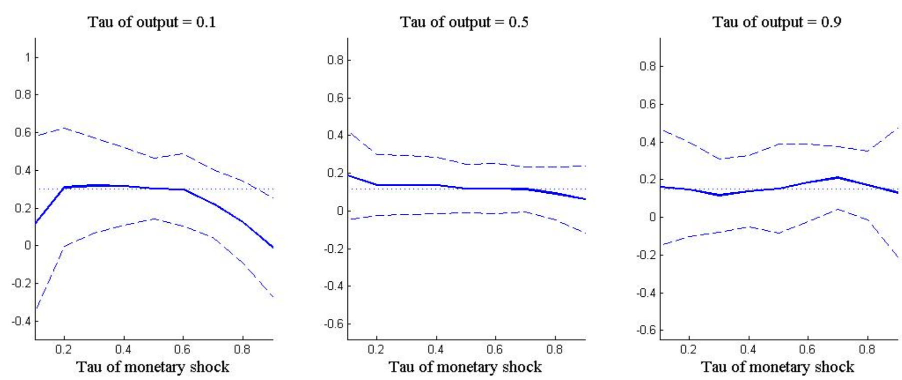
Figure 12

Cross-Section of the First Lag of Output Growth Slope Surface - Cubic Regression & M1 Money Supply.

This figure plots the cross-section of Panels A and B in Figure 10 for the 10th, 50th and 90th percentile output sections. The horizontal line plots the median of the monetary shock. The dash lines are the two standard deviation bands. The estimated values are reported from the cubic regression model.



A. 8 Lags

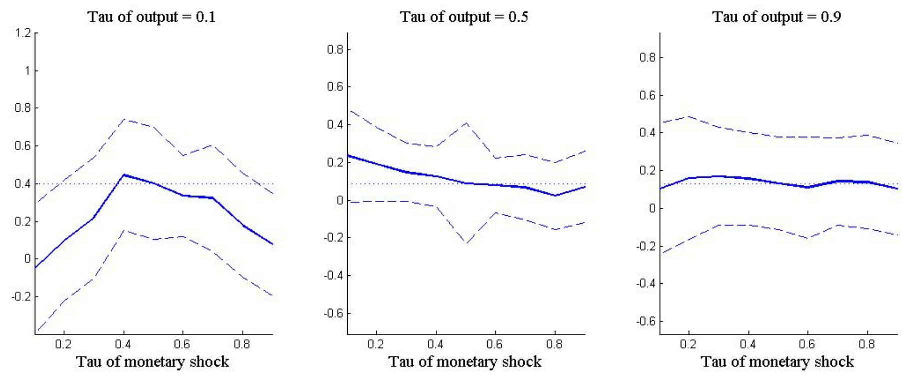


B. 12 Lags

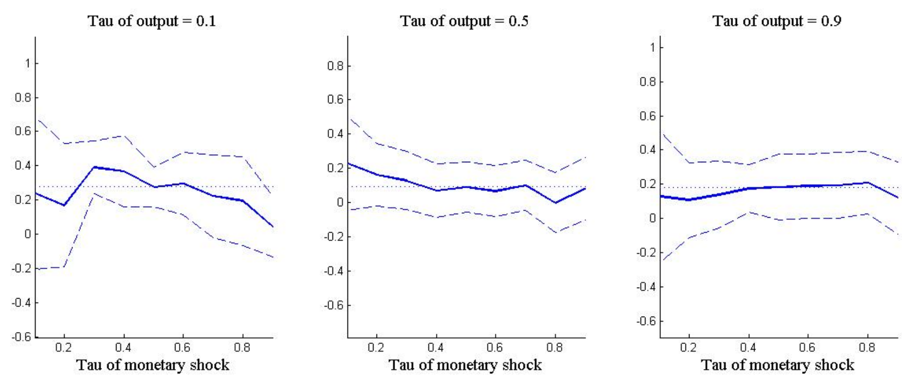
Figure 13

Cross-Section of the First Lag of Output Growth Slope Surface - Quartic Regression & M1 Money Supply.

This figure plots the cross-section of Panels C and D in Figure 10 for the 10th, 50th and 90th percentile output sections. The horizontal line plots the median of the monetary shock. The dash lines are the two standard deviation bands. The estimated values are reported from the quartic regression model.



A. 8 Lags

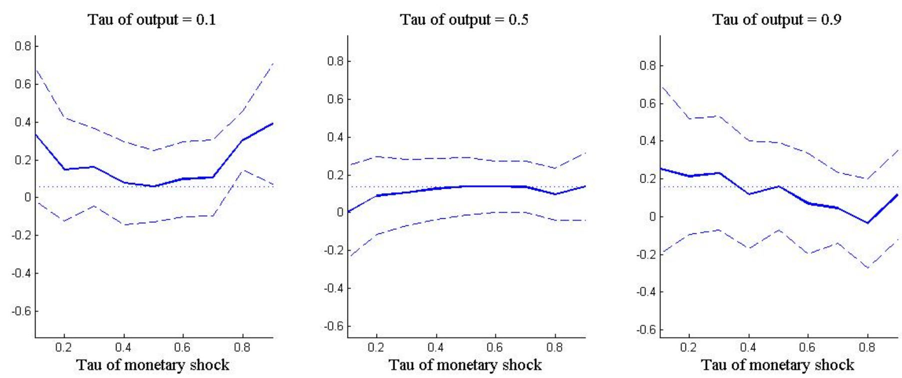


B. 12 Lags

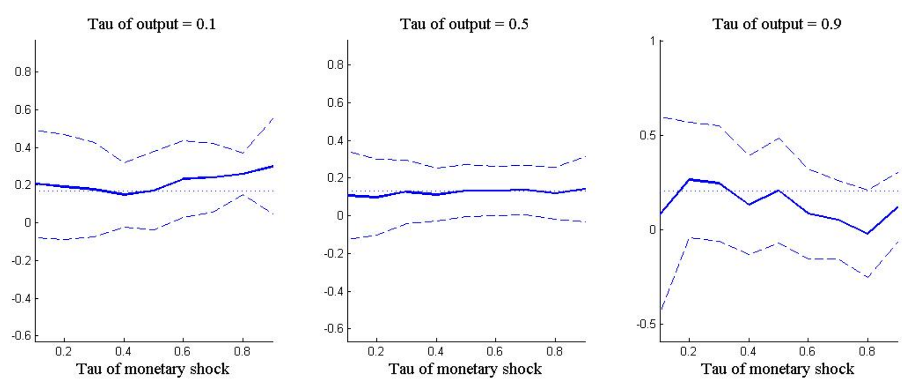
Figure 14

Cross-Section of the First Lag of Output Growth Slope Surface - Cubic Regression & M2 Money Supply.

This figure plots the cross-section of Panels A and B in Figure 11 for the 10th, 50th and 90th percentile output sections. The horizontal line plots the median of the monetary shock. The dash lines are the two standard deviation bands. The estimated values are reported from the cubic regression model.



A. 8 Lags

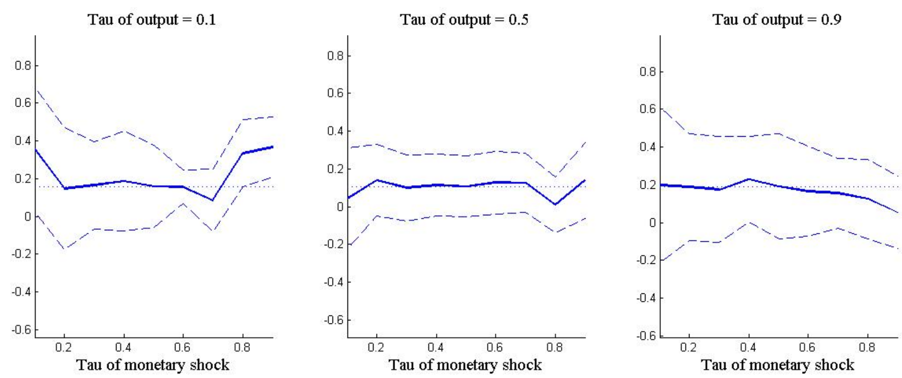


B. 12 Lags

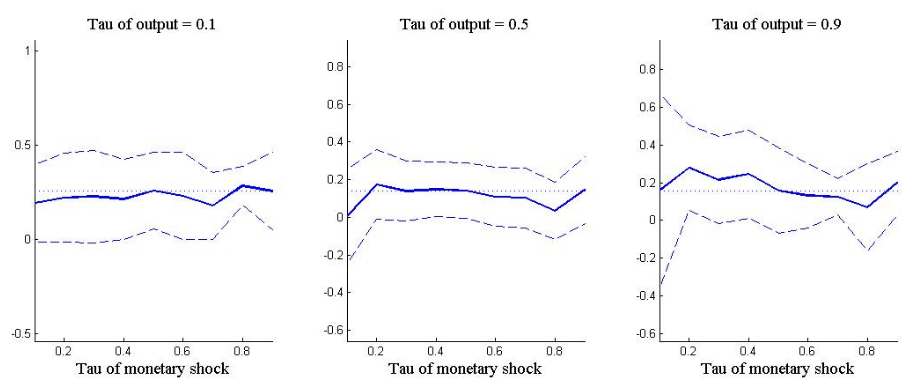
Figure 15

Cross-Section of the First Lag of Output Growth Slope Surface - Quartic Regression & M2 Money Supply.

This figure plots the cross-section of Panels C and D in Figure 11 for the 10th, 50th and 90th percentile output sections. The horizontal line plots the median of the monetary shock. The dash lines are the two standard deviation bands. The estimated values are reported from the quartic regression model.



A. 8 Lags

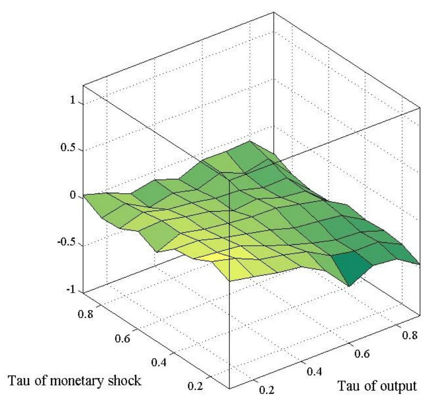


B. 12 Lags

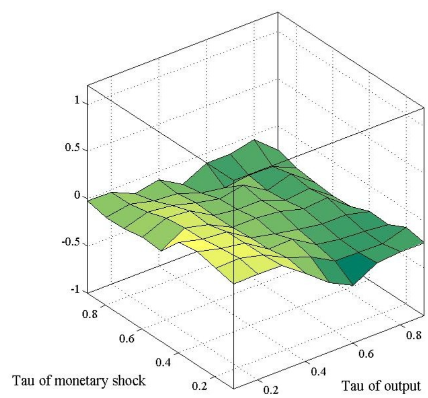
Figure 16

Slope of the First Lag of First Differenced Treasury Yield- M1 Money Supply.

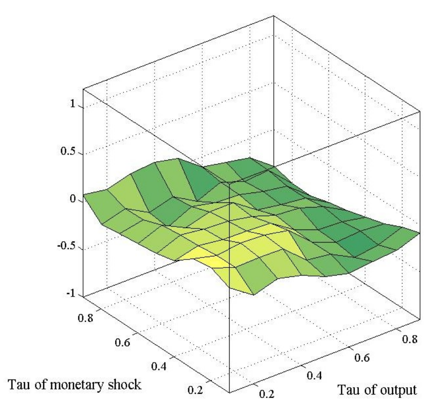
This figure shows the estimated slope of the lagged first differenced Treasury yield in the output process equation based on a cubic (Panels A and B) or quartic (Panels C and D) regression model. The monetary instrument is M1 money supply. The cubic or quartic expansion involves the intercept term, the first four lags of the output growth and the lag of the first difference Treasury yield. "Lags" refers to the number of output lags used in the output process equation.



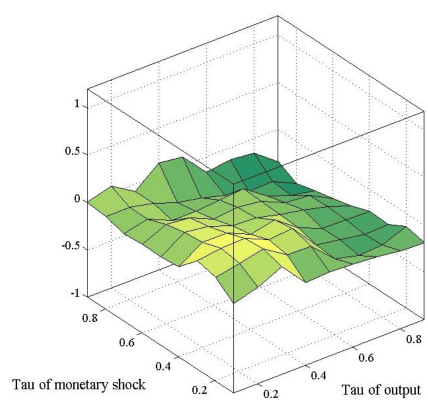
A. Cubic, 8 Lags



B. Cubic, 12 Lags



C. Quartic, 8 Lags

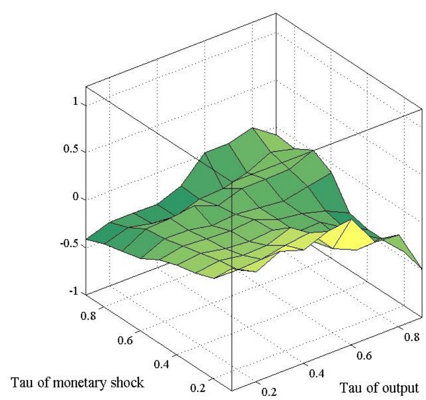


D. Quartic, 12 Lags

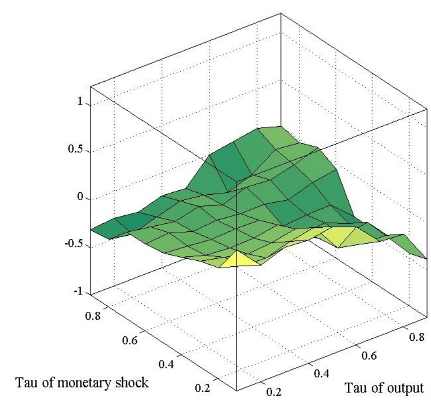
Figure 17

Slope of the First Lag of First Differenced Treasury Yield - M2 Money Supply.

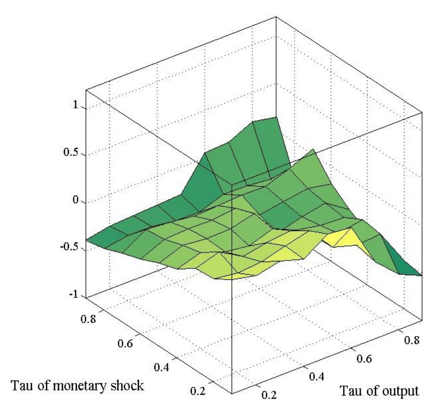
This figure shows the estimated slope of the first lag of first difference Treasury yield in the output process equation based on a cubic (Panels A and B) or quartic (Panels C and D) regression model. The monetary instrument is M2 money supply. The cubic or quartic expansion involves the intercept term, the first four lags of the output growth and the lag of the first difference Treasury yield. “Lags” refers to the number of output lags used in the output process equation.



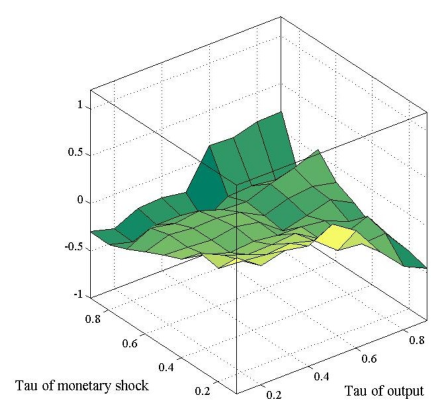
A. Cubic, 8 Lags



B. Cubic, 12 Lags



C. Quartic, 8 Lags

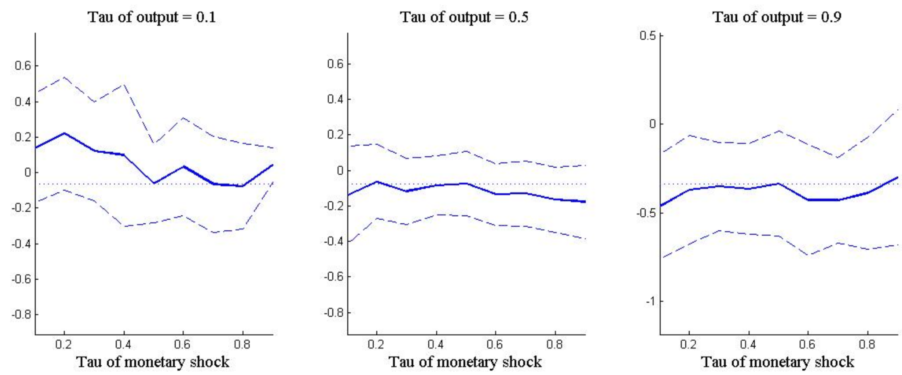


D. Quartic, 12 Lags

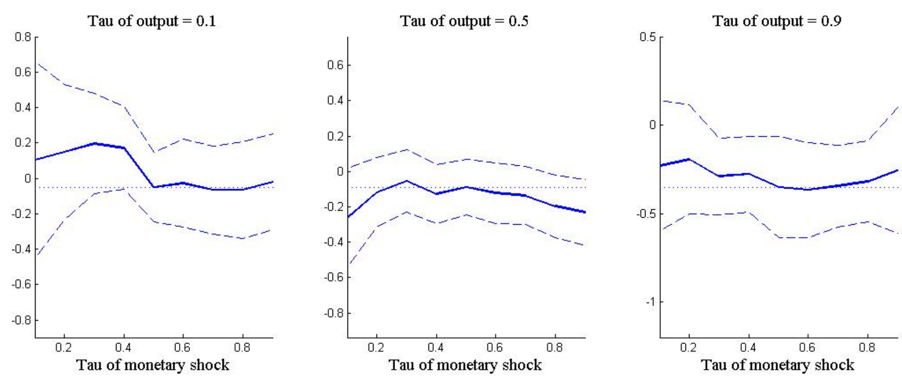
Figure 18

Cross-Section of the First Lag of First Differenced Treasury Yield Slope Surface - Cubic Regression & M1 Money Supply.

This figure plots the cross-section of Panels A and B in Figure 16 for the 10th, 50th and 90th percentile output sections. The horizontal line plots the median of the monetary shock. The dash lines are the two standard deviation bands. The estimated values are reported from the cubic regression model.



A. 8 Lags

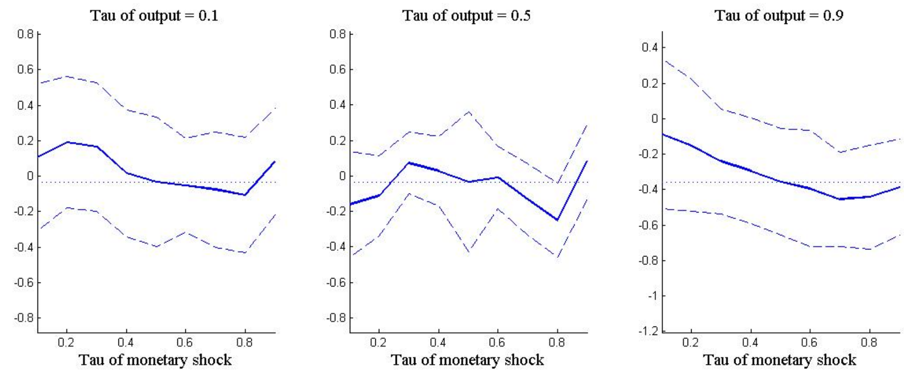


B. 12 Lags

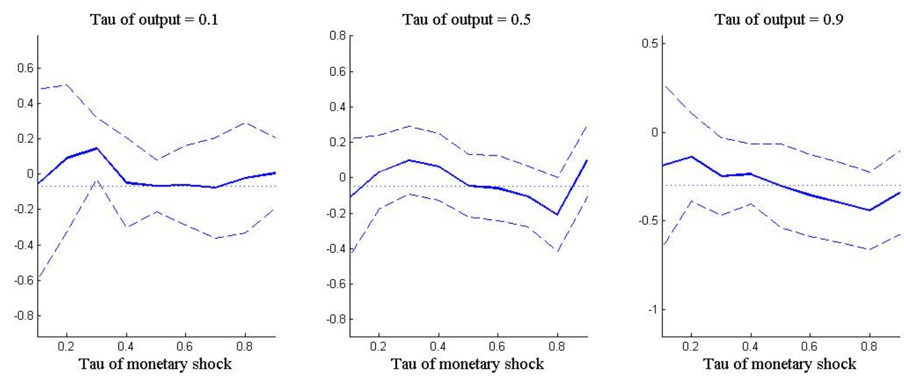
Figure 19

Cross-Section of the First Lag of First Differenced Treasury Yield Slope Surface - Quartic Regression & M1 Money Supply.

This figure plots the cross-section of Panels A and B in Figure 16 for the 10th, 50th and 90th percentile output sections. The horizontal line plots the median of the monetary shock. The dash lines are the two standard deviation bands. The estimated values are reported from the cubic regression model.



A. 8 Lags

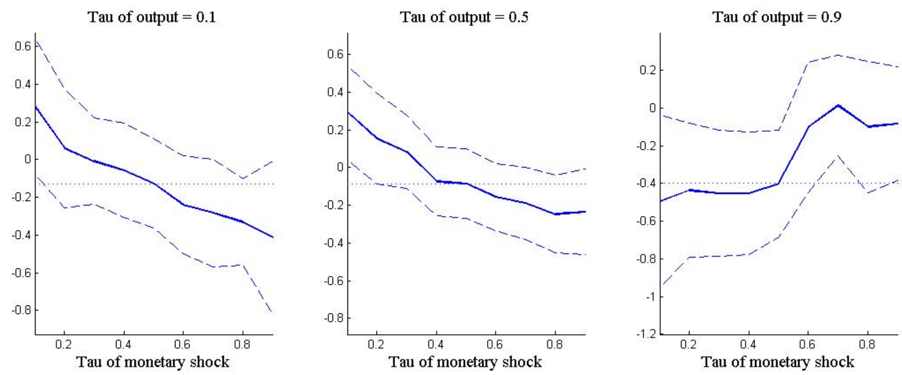


B. 12 Lags

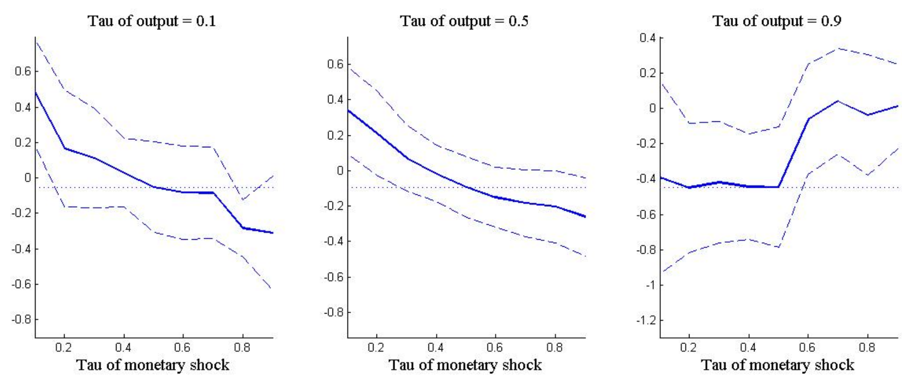
Figure 20

Cross-Section of the First Lag of First Differenced Treasury Yield Slope Surface - Cubic Regression & M2 Money Supply.

This figure plots the cross-section of Panels A and B in Figure 17 for the 10th, 50th and 90th percentile output sections. The horizontal line plots the median of the monetary shock. The dash lines are the two standard deviation bands. The estimated values are reported from the cubic regression model.



A. 8 Lags

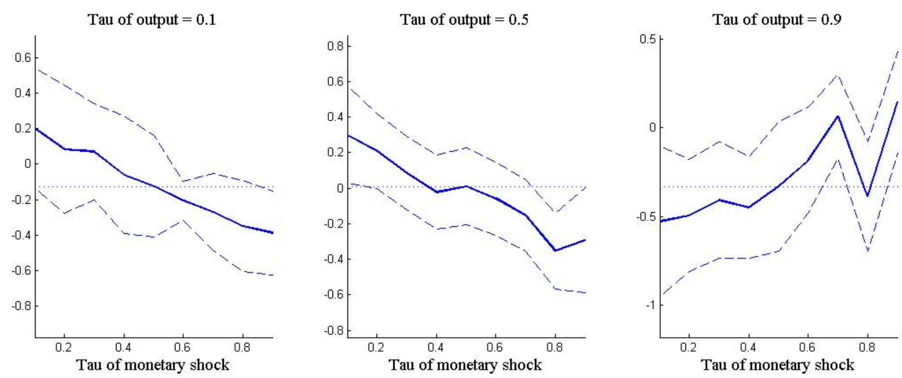


B. 12 Lags

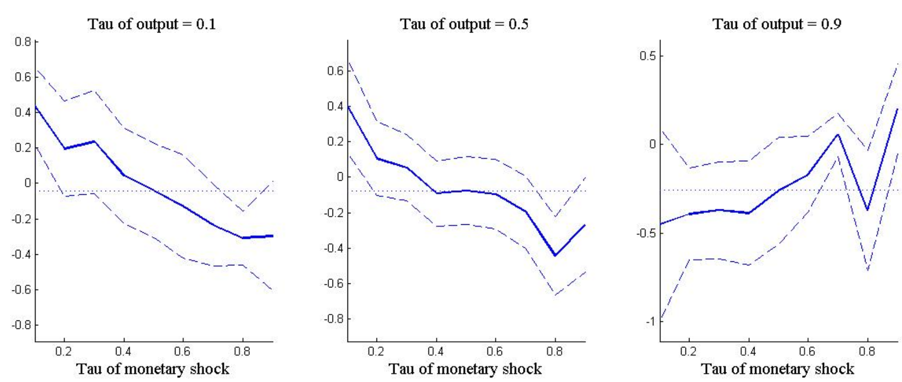
Figure 21

Cross-Section of the First Lag of First Differenced Treasury Yield Slope Surface - Quartic Regression & M2 Money Supply.

This figure plots the cross-section of Panels C and D in Figure 17 for the 10th, 50th and 90th percentile output sections. The horizontal line plots the median of the monetary shock. The dash lines are the two standard deviation bands. The estimated values are reported from the quartic regression model.



A. 8 Lags



B. 12 Lags

Table 1

Monte Carlo Simulation of the Estimator's Properties in Case 1

This table estimates $\hat{\alpha}_2$ using the regression function of $H = \alpha_0 + \alpha_1 X_1 + \alpha_2 Y_2 + \sum_{i=1}^{I} \varphi_i \frac{\hat{w}(\tau_2)^i}{I!} Y_2$ when the true data generating process is $Y_1 = \alpha_0 + \alpha_1 X_1 + (\alpha_2 + \delta(\lambda e^w + u))Y_2$, where $(\alpha_0, \alpha_1, \alpha_2, \delta, \lambda) = (3, 4, 4, 5, 3)$, $(\beta_1, \beta_2, \beta_3) = (1, 2, 3)$, $X_1 \sim t_3$, $X_2 \sim N(15, 2)$, $w \sim N(0, 0.5)$ and $u \sim N(0, 1)$. I represents the order of the polynomial and the maximum order considered is 4. The figures reported corresponds to $\tau_1 = \tau_2 = \tau$.

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
true	5.4955	9.6396	12.9183	15.8271	18.9856	22.2924	25.8738	30.9784	38.87705
$I = 1$									
estimated	3.5521	9.4415	13.6079	17.3263	20.7860	24.0118	27.8808	32.5694	39.1009
bias	-1.9434	-0.1981	0.6896	1.3778	1.7860	1.7194	1.7620	1.5134	0.2239
rmse	4.1069	3.2517	2.9541	3.0795	3.4170	3.4321	3.6242	4.0522	4.8128
$I = 2$									
estimated	5.0154	8.9553	12.4897	15.7282	18.9529	22.3514	26.4806	31.7986	39.9921
bias	-0.4801	-0.6843	-0.4286	-0.2204	-0.0471	0.0589	0.3618	0.7426	1.1151
rmse	3.9834	3.2995	3.0643	2.7649	2.7667	2.8626	2.9577	3.3835	4.1471
$I = 3$									
estimated	5.8375	9.8298	12.9214	15.9167	18.9769	22.0809	25.9667	30.9533	39.2663
bias	0.3420	0.1902	0.0031	-0.0319	-0.0231	-0.2115	-0.1521	-0.1027	0.3893
rmse	3.7537	3.2860	3.1333	2.9173	2.8474	2.8916	3.1235	3.2884	4.0012
$I = 4$									
estimated	5.6320	9.7275	12.9226	16.0770	18.8716	22.1796	26.0498	30.9736	39.0331
bias	0.1365	0.0878	0.0043	0.1284	-0.1284	-0.1128	-0.0690	-0.0824	0.1561
rmse	4.0254	3.2389	2.9776	3.0358	2.8696	2.8846	3.0703	3.4039	3.8870

Table 2

Monte Carlo Simulation of the Estimator's Properties in Case 2

This table estimates $\hat{\alpha}_2$ using the regression function of $H = \alpha_0 + \alpha_1 X_1 + \alpha_2 Y_2 + \sum_{i=1}^{I} \varphi_{i,0} \frac{\hat{w}(\tau_2)^i}{I!} + \sum_{i=1}^{I} \varphi_{i,1} \frac{\hat{w}(\tau_2)^i}{I!} X_1 + \sum_{i=1}^{I} \varphi_{i,2} \frac{\hat{w}(\tau_2)^i}{I!} Y_2$ when the true data generating process is $Y_1 = \alpha_0 + \alpha_1 X_1 + (\alpha_2 + \delta(\lambda e^{w} + u))Y_2 4$, where $(\alpha_0, \alpha_1, \alpha_2, \delta, \lambda) = (3, 4, 4, 5, 3)$, $(\beta_1, \beta_2, \beta_3) = (1, 2, 3)$, $X_1 \sim t_3$, $X_2 \sim N(15, 2)$, $w \sim N(0, 0.5)$ and $u \sim N(0, 1)$. I represents the order of the polynomial and the maximum order considered is 4. The figures reported corresponds to $\tau_1 = \tau_2 = \tau$.

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
true	5.4955	9.6396	12.9183	15.8271	18.9856	22.2924	25.8738	30.9784	38.87705
$I = 1$									
estimated	1.3907	7.4389	12.6760	16.8992	20.5550	24.7098	29.5900	34.3608	42.2158
bias	-4.1048	-2.2007	-0.2424	0.9506	1.5550	2.4173	3.4712	3.3048	3.3388
rmse	6.1639	3.7894	2.9829	2.9556	3.3773	4.0531	5.1396	5.6382	7.6920
$I = 2$									
estimated	5.5257	8.7284	11.9460	15.0544	18.4307	21.9027	26.2857	31.6772	39.6226
bias	0.0302	-0.9112	-0.9723	-0.8942	-0.5693	-0.3897	0.1669	0.6212	0.7455
rmse	3.8759	2.9657	2.9725	2.9031	2.8556	3.4483	3.7388	4.5569	6.6199
$I = 3$									
estimated	6.0621	9.8282	12.7148	16.0777	19.2028	22.1502	25.6016	31.1070	38.6632
bias	0.5666	0.1886	-0.2036	0.1291	0.2028	-0.1422	-0.5172	0.0510	-0.2138
rmse	3.9763	3.1133	2.9089	2.4963	3.0832	3.1441	3.8215	4.3669	7.3487
$I = 4$									
estimated	5.7903	9.7579	13.0894	15.7274	19.2629	22.1243	26.2222	30.1782	38.7863
bias	0.2948	0.1183	0.1711	-0.2211	0.2629	-0.1681	0.1034	-0.8778	-0.0907
rmse	4.4540	3.1824	2.7624	3.0286	2.8153	3.2167	3.6239	4.7077	7.3832

Table 3

Monte Carlo Simulation of the Estimator's Properties in Case 3

This table estimates $\hat{\alpha}_2$ using the regression function of $H = \alpha_0 + \alpha_1 X_1 + \alpha_2 Y_2 + \sum_{i=1}^{I} \varphi_i \frac{\hat{w}(\tau_2)^i}{I!} Y_2$ when the true data generating process is $Y_1 = (\alpha_0 + \tilde{\delta}(\tilde{\lambda} e^w + u)) + (\alpha_1 + \tilde{\delta}(\tilde{\lambda} e^w + u))X_1 + (\alpha_2 + \delta(\lambda e^w + u))Y_2$, where $(\alpha_0, \alpha_1, \alpha_2, \delta, \lambda, \tilde{\delta}, \tilde{\lambda}) = (3, 4, 4, 5, 3, 10, 5)$, $(\beta_1, \beta_2, \beta_3) = (1, 2, 3)$, $X_1 \sim t_3$, $X_2 \sim N(15, 2)$, $w \sim N(0, 0.5)$ and $u \sim N(0, 1)$. I represents the order of the polynomial and the maximum order considered is 4. The figures reported corresponds to $\tau_1 = \tau_2 = \tau$.

τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
true	5.4955	9.6396	12.9183	15.8271	18.9856	22.2924	25.8738	30.9784	38.87705
$I = 1$									
estimated	2.5477	8.9765	13.1274	16.9606	20.7226	24.4943	28.6231	33.0332	40.3293
bias	-2.9478	-0.6631	0.2091	1.0120	1.7226	2.2019	2.5043	1.9772	1.4522
rmse	4.9030	3.0782	3.1211	3.0488	3.4170	3.9555	4.4128	4.5147	5.2044
$I = 2$									
estimated	4.2612	9.1409	12.2728	15.3445	18.6133	22.6044	26.8681	32.0176	40.8277
bias	-1.2343	-0.4987	-0.6456	-0.6041	-0.3867	0.3120	0.7493	0.9615	1.9507
rmse	4.5046	3.5419	3.4319	2.9921	3.1263	2.8346	2.9027	3.4657	4.4431
$I = 3$									
estimated	4.4812	9.1210	12.3174	15.7622	18.8094	22.1458	26.0107	31.3768	39.4584
bias	-1.0143	-0.5186	-0.6009	-0.1864	-0.1906	-0.1466	-0.1081	0.3208	0.5813
rmse	4.4186	3.4230	3.2645	2.9386	2.8749	2.9271	2.9006	3.3398	4.2343
$I = 4$									
estimated	4.8134	9.0978	12.6131	16.1044	19.0015	22.5886	26.1960	31.3030	40.1433
bias	-0.6821	-0.5418	-0.3052	0.1558	0.0015	0.2962	0.0772	0.2469	1.2662
rmse	4.0521	3.6663	3.5244	3.1108	3.2189	3.2354	3.1364	3.4301	4.4341