Modeling Quantile Dependence: A New Look at the Money-Output Relationship

Nicholas C.S. Sim
MODELING QUANTILE DEPENDENCE: A NEW LOOK AT THE MONEY-OUTPUT RELATIONSHIP †

NICHOLAS C.S. SIM∗
School of Economics
University of Adelaide

ABSTRACT

Do money supply shocks influence output growth asymmetrically? At different levels of output growth, would the influence of the same monetary policy stance vary? To address these questions, we propose a series-estimation method that models the quantile of output growth on the quantile of money supply shock, where restrictive (expansive) policies are represented by the left (right) tail of the shock’s distribution. Generally, we find that each quantile of output growth responds more to restrictive than expansive money supply shocks. For M2 money supply, both restrictive and expansive shocks become even more effective when applied to output growth in its tails.

JEL Classification: E50, C50.
Key Words: Monetary Policy, Output Growth, Quantile Regression, Quantile Dependence, Series Estimation.

†I am indebted to Christopher Baum, Fabio Schiantarelli and Zhijie Xiao for their support. I am also grateful to Seungmoon Choi and Hongtao Zhou for their very helpful comments and suggestions.

∗Corresponding address: School of Economics, University of Adelaide, Adelaide, SA 5000 Australia. Tel: +61-08-8303-4927, Fax: +61-08-8303-1460, Email: nicholas.sim@adelaide.edu.au
1 Introduction

The relationship between output growth and monetary policy has been a topic of considerable debate in the past decades. Early attempts to identify unanticipated changes in the monetary policy focused on the innovations in money supply itself. Starting from Cover (1992), it is now well-known that the money-output relationship is nonlinear, where output growth declines more strongly following a negative money supply shock than it rises following a positive money supply shock of the same magnitude. The econometric methodology employed by Cover and similar variations by other researchers involved 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 affect output growth asymmetrically 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 conditional mean, which is necessary for identifying episodes of monetary contraction and expansion. To see this, suppose the data-generating process is

\[ m_t = b_0 + b_1 m_{t-1} + w_t \]

where \( m_t \) is the money supply growth at time \( t \) and \( w_t \) is the money supply shock. We may express the process as

\[ m_t = b_0 + E[w_t|m_{t-1}] + b_1 m_{t-1} + w_t - E[w_t|m_{t-1}] = b_0 + E[w_t|m_{t-1}] + b_1 m_{t-1} + w^*_t \]

so that \( w^*_t \) is guaranteed to have a zero conditional mean. In the regression, \( b_0 \) and \( E[w_t|m_{t-1}] \) are not separately identifiable and unless \( E[w_t|m_{t-1}] = 0 \) holds, the estimated “shocks” will be estimates of \( w^*_t \) but not \( w_t \). Therefore, suppose if \( E[w_t|m_{t-1}] < 0 \), then positive estimates of \( w^*_t \) could in fact be obtained for some negative \( w_t \), implying that some estimated episodes of monetary expansions could in fact be monetary contractions.\(^2\)

Following the tradition of Cover by adopting the money supply as a policy instrument, this paper

---


\(^2\)Suppose the money supply shock indeed has a zero conditional mean. Observing that -1.117 is one of the statistically significant estimates of the intercept term in Cover’s M1 money supply process regressions, this implies that the Federal Reserve might have systematically contracted M1 money supply by more than 4% per annum all else being equal during the sample period. Attempts to reconcile this observation inevitably lead to questioning the plausibility that the money supply shock truly has a zero conditional mean.
investigates the relationship between monetary policy and output growth by proposing a new quantile regression methodology. This methodology relaxes Cover’s assumption while still enabling us to uncover any potential asymmetric influence that money supply shock has on output growth. The insight comes from observing that the quantile of money supply shock contains information about the stance of monetary policy. Hence, even though we no longer rely on the signs of the estimated shocks to distinguish between monetary expansions and contractions, our methodology 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) using the quantile of money supply shock itself.

To further elucidate the idea of ranking the policy environments, one may assert that relative to the median, a 10th percentile money supply shock reflects a more restrictive monetary policy stance while a 90th percentile shock reflects a more expansive one. Such interpretation motivates constructing an econometric model of output growth as a function of the quantile of money supply shock. Another potential dimension of nonlinearity is to allow output growth react differently to a certain monetary policy stance contingent on whether output growth is high or low. This will enable us to ascertain if that particular monetary policy objective, as indexed by some quantile of money supply shock, will be more effective in some states of economic growth than others. Therefore, a unified econometric framework that can simultaneously accommodate these two dimensions of nonlinearities will be one that models the quantile of output growth as a function of the quantile of money supply shock.

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,3 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 without any further identifying assumption, the 20th percentile shock is expansive relative to the 10th percentile shock, as the 10th percentile shock is purported to restrict

---

3It should be noted that Conover et al. (1999) had first used the terms expansive and restrictive to describe the monetary environment. However, the interpretation here differs from theirs.
output growth more aggressively than does the 20th percentile shock.

In short, the $\tau^{th}$ quantile of money supply shock reflects some $\tau^{th}$ quantile monetary policy position. However, further assertion that the $\tau^{th}$ quantile of money supply shock is expansive or restrictive can only be made in reference to another 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, e.g. the quantile of output growth, conditioned on the quantile of a regressor, e.g. the quantile of money supply shock. The quantile of the regressor will then be separately modeled in another regression.

We consider a linear system of equations to align our empirical work more closely with Cover’s methodology. As we will see, a more general version of Cover’s model can be constructed where the coefficients are nonparametric functions of the unobservables such as shocks to the money supply and output growth. To estimate this semiparametric model, we propose a power series expansion of the nonparametric coefficients, employing truncation arguments similar to Andrews (1991), Newey (1997) and Lee (2007) while allowing the truncation parameter to grow with the sample size. In the Monte Carlo exercise, we demonstrate the model’s properties of bias and root-mean-squared error based on the various orders of expansion.

Previewing our results, we find that restrictive money supply shocks are more effective in contracting output growth than expansive shocks are in expanding it. At the median output growth, the influence of the expansive shock is very weak, reminiscent of “pushing on the string”. However, for output growth located in both the tails, expansive shocks become increasingly effective even while the

\footnote{This is similar to a model with a finite-dimensional 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 for infinite dimensional quantile regressions. Their paper is similar as they also employed a truncation argument in their analysis.}
asymmetric effects of money supply shocks on output growth remain. This implies that the impact of the same restrictive or expansive stance may vary when applied to different levels of output growth.

The rest of the paper is organized as follows. As a new econometric framework is proposed for the empirical work, the next two sections will focus almost exclusively on the technique itself. Section 2 presents the empirical framework and considers the class of linear semiparametric QQ models. Section 3 discusses the method of estimation, derives the asymptotic distribution, and presents some Monte Carlo evidence on the performance of the series regression. The empirical results are given in Section 4 and Section 5 concludes.

2 The Model

This section will first present the empirical model to be estimated in this paper, followed by its generalization to a linear semiparametric QQ model. A brief review on the standard linear quantile regression model, which contains some background information for the technical content that follows, can be found in Appendix A.

2.1 The Money-Output Framework

It is widely recognized that the money-output relationship is akin to “pushing on the string” as monetary policy appears to be a more effective tool for contracting than expanding the economy. Empirical support for this phenomenon can be drawn from Cover (1992), whose model will be followed closely in this paper. In his model, the monetary policy shocks are first identified as residuals from estimating a monetary process equation of

\[ m_t = b_0 + \sum_{i=1}^{K_m} b_{m,i} m_{t-i} + \sum_{i=1}^{K_x} b'_{x,i} x_{t-i} + w_t, \]  

(1)

where \( m_t \) is a monetary instrument, \( x_t \) is a vector of other information variables and \( w_t \) is the money supply shock. 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 this relationship during the pre-war and pre-Depression
periods.\(^5\) Having obtained the money supply shocks from (1), output growth is then regressed on the positive and negative residuals, \(\hat{w}_t^*+\) and \(\hat{w}_t^*-\), replacing the positive and negative shocks, \(w_t^+\) and \(w_t^-\), in

\[
y_t = a_0 + \sum_{i=1}^{K_y} a_{y,i}y_{t-i} + \sum_{i=0}^{K_r} a_{r,i} \Delta r_{t-i} + \sum_{i=0}^{K_u} (a_{w,i}^+ w_{t-i}^+ + a_{w,i}^- w_{t-i}^-) + u_t,
\]

(2)

where \(y_t\) is output growth, \(\Delta r_t\) is the change in the three-month Treasury yield, i.e. \(\Delta r_t = r_t - r_{t-1}\), and \(u_t\) is the innovation in output growth. As explained earlier, the residual is an estimator of \(w_t^* = w_t - E[w_t|\mathcal{G}_{t-1}]\) but not \(w_t\), where \(\mathcal{G}_{t-1}\) denotes the \textit{ex-ante} information set at time \(t\). However, there is no guarantee that \(w_t\) has a zero conditional mean, which is necessary for the residual to consistently estimate \(w_t\).

In his benchmark model, Cover specified a single lag for output growth and the change in Treasury yield and included only contemporaneous positive and negative money supply shocks. The extensions of this benchmark model included lagged money supply shocks. Interestingly, the lagged change in Treasury yield was statistically significant in all of Cover’s regressions. In addition, he found that the contemporaneous negative shock was typically most influential and statistically most significant among the negative shocks when lagged shocks were included. The positive shocks usually were either statistically insignificant or estimated with the wrong signs if they were statistically significant. Hence Cover’s finding motivates a parsimonious setup similar to his benchmark model, which under the null of symmetry can be expressed as

\[
y_t = \alpha_I(w_t, u_t) + \sum_{i=1}^{K_y} a_{y,i}y_{t-i} + \sum_{i=1}^{K_r} a_{r,i} \Delta r_{t-i}. \tag{3}
\]

In (3), we incorporate \(w_t\) into \(\alpha_I\) along with \(u_t\), thus producing a random intercept term. As discussed in Appendix A, the \(\tau^{th}\) conditional quantile of \(y_t\) corresponds to the \(\tau^{th}\) quantile of \(u_t\). Let

\(^5\)Other non-money measures that could be used include the Federal funds rate (Morgan, 1993) and the short-term interbank offer rates (Florio, 2005).
$F_u(\cdot)$ be the distribution function of $u_t$ and $F_u^{-1}(\tau)$ be its $\tau^{th}$ quantile. Similar notations will be used for the counterparts of $w_t$. Therefore, the $\tau_1^{th}$ quantile of output growth conditioning on the $\tau_2^{th}$ quantile of money supply shock can be expressed as

$$Q_{yt}(\tau_1|S_{t-1}, F_{w}^{-1}(\tau_2)) = \alpha_I(F_{w}^{-1}(\tau_2), F_u^{-1}(\tau_1)) + \sum_{i=1}^{K_y} a_{y,i}y_{t-i} + \sum_{i=1}^{K_r} a_{r,i}\Delta r_{t-i}$$

$$\equiv \alpha_I(\tau_2, \tau_1) + \sum_{i=1}^{K_y} a_{y,i}y_{t-i} + \sum_{i=1}^{K_r} a_{r,i}\Delta r_{t-i}. \tag{4}$$

The quantile of money supply shock influences the quantile of output growth through the $\tau_2$-argument in $\alpha_I(\tau_2, \tau_1)$. In fact, it will influence the $\tau_1^{th}$ quantile of output growth asymmetrically if $\alpha_I(\tau_2, \tau_1) \neq \alpha_I(1-\tau_2, \tau_1)$ for $\tau_2 \neq 0.5$. Notice that (3) is a pure location shift model which can be extended to exhibit both location and scale shift by allowing $w_t$ to affect $y_t$ through the slope parameters as well.\(^6\) This extension can be written as

$$y_t = \alpha_I(w_t, u_t) + \sum_{i=1}^{K_y} \alpha_{y,i}(w_t, u_t)y_{t-i} + \sum_{i=1}^{K_r} \alpha_{r,i}(w_t, u_t)\Delta r_{t-i}, \tag{5}$$

assuming $y_t$ is monotonic in $w_t$ and $u_t$, so that the $\tau_1^{th}$ quantile of output growth conditioning on the $\tau_2^{th}$ quantile of the money supply shock may now be written as

$$Q_{yt}(\tau_1|S_{t-1}, F_{w}^{-1}(\tau_2))$$

$$= \alpha_I(F_{w}^{-1}(\tau_2), F_u^{-1}(\tau_1)) + \sum_{i=1}^{K_y} a_{y,i}(F_{w}^{-1}(\tau_2), F_u^{-1}(\tau_1))y_{t-i} + \sum_{i=1}^{K_r} a_{r,i}(F_{w}^{-1}(\tau_2), F_u^{-1}(\tau_1))\Delta r_{t-i}$$

$$\equiv \alpha_I(\tau_2, \tau_1) + \sum_{i=1}^{K_y} a_{y,i}(\tau_2, \tau_1)y_{t-i} + \sum_{i=1}^{K_r} a_{r,i}(\tau_2, \tau_1)\Delta r_{t-i}. \tag{6}$$

To generalize (5), we allow the parameters to be nonlinear in both $w_t$ and $u_t$. This contrasts (3), where the intercept is linear in $w_t$ and $u_t$. In addition, we will consider a more flexible case where the parameters are also unknown functions of $w_t$ and $u_t$, thus extending (5) to a semiparametric model.

The objective of the paper is to estimate (6), the QQ counterpart, which belongs to the class of linear semiparametric QQ model that we proposed in this paper.

\(^6\)A discussion of location and scale shift in a linear model can be found in Appendix A.
2.2 The General Framework

The QQ framework extends the standard linear quantile regression model by allowing the regressor to be itself a conditional quantile. We begin with a general linear framework where the coefficients are unknown functions of the model’s innovation terms. For the standard quantile regression model, Appendix A explains how this can be expressed as a random coefficients model where the coefficients are functions of a single innovation term. Using this random coefficient interpretation, a more general framework to express the relationship between the quantiles of $Y_{1,t}$ and $Y_{2,t}$ can first be written as

$$Y_{1,t} = \alpha_0(w_t, u_t) + \alpha_1(w_t, u_t)' X_{1,t} + \alpha_2(w_t, u_t)Y_{2,t}$$  \hspace{1cm} (7)$$

and

$$Y_{2,t} = \beta_0 + \beta_1' X_{2,t} + w_t,$$  \hspace{1cm} (8)$$

where $X_{1,t}$ and $X_{2,t}$ consist of exogenous variables satisfying the standard exclusionary restriction; $Y_{1,t}$ is monotonic in the unobservables, of which the parameters in (7) are unknown functions; $w_t$ is a homoskedastic innovation term, thus the conditional quantile function of $Y_{2,t}$ exhibits only location shift. Equation (5) is a special case of (7) where $Y_{2,t}$ is absent.

Assume that $w_t$ and $u_t$ are conditionally independent, where $u_t$ is the innovation in $Y_{1,t}$. Conditioning on $X_{1,t}$ and $Y_{2,t}$, the $\tau_1^{th}$ quantile of $Y_{1,t}$ is obtained when $u_t$ in (7) corresponds to its $\tau_1^{th}$ quantile, $F_u^{-1}(\tau_1)$:

$$Q_{Y_{1,t}}(\tau_1|X_{1,t}, Y_{2,t}) = \alpha_0(w_t, F_u^{-1}(\tau_1)) + \alpha_1(w_t, F_u^{-1}(\tau_1))' X_{1,t} + \alpha_2(w_t, F_u^{-1}(\tau_1))Y_{2,t}. \hspace{1cm} (9)$$

The next step is to obtain the dependence between the quantiles which (9) has yet to express. In a similar way, the $\tau_2^{th}$ conditional quantile of $Y_{2,t}$ is obtained when $w_t$ in (8) corresponds to its $\tau_2^{th}$ quantile, $F_w^{-1}(\tau_2)$. Therefore, conditioning recursively on $Q_{Y_{1,t}}(\tau_2|X_{2,t})$, which is also consistent
with setting \( w_t \) in (9) to \( F^{-1}_w(\tau_2) \), yields

\[
Q_{Y_1,t}(\tau_1|X_{1,t}, Q_{Y_2,t}(\tau_2|X_{2,t}))
= \alpha_0(F^{-1}_w(\tau_2), F^{-1}_u(\tau_1)) + \alpha_1(F^{-1}_w(\tau_2), F^{-1}_u(\tau_1))'X_{1,t} + \alpha_2(F^{-1}_w(\tau_2), F^{-1}_u(\tau_1))Q_{Y_2,t}(\tau_2|X_{2,t})
\equiv \alpha_0(\tau_2, \tau_1) + \alpha_1(\tau_2, \tau_1)'X_{1,t} + \alpha_2(\tau_2, \tau_1)Q_{Y_2,t}(\tau_2|X_{2,t}),
\]

(10)

where \( \alpha_i(\tau_2, \tau_1) \equiv \alpha_i(F^{-1}_w(\tau_2), F^{-1}_u(\tau_1)) \).

To obtain the QQ model represented by (10), we have specified a recursive system (7) and (8) so that \( Y_{2,t} \) may influence \( Y_{1,t} \) but not vice-versa. This setup is similar to Ma and Koenker (2006) with two differences. First, Ma and Koenker considered a system of nonlinear equations while we specialize it to the linear case. Second, Ma and Koenker considered a fully parametric setup involving both regressors and innovation terms. For the linear model, following Ma and Koenker would entail specifying how \( u_t \) and \( w_t \) enter the \( \alpha \) parameters, which is avoided here.

Two observations may be drawn from the QQ model in general. First, (10) suggests that the influence by the quantile regressor may also come indirectly from \( \alpha_0 \) and \( \alpha_1 \) as these parameters may be functions of \( w_t \) as well. Second, it turns out that 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 \( Q_{Y_2,t}(\tau_2|X_{2,t}) \). This will be explained in Section 3 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.

In a special case, the QQ model may also arise from specific assumptions about heteroskedasticity of the linear form. Assuming that the dimension of \( X_{1,t} \) is one for simplicity, consider specializing (7) to

\[
Y_{1,t} = a_0 + a_1X_{1,t} + a_2Y_{2,t} + \tilde{u}_t,
\]

(11)

where \( \tilde{u}_t = \delta_{w,0}w_t + \delta_{u,0}u_t + (\delta_{w,1}w_t + \delta_{u,1}u_t)X_{1,t} + (\delta_{w,2}w_t + \delta_{u,2}u_t)Y_{2,t} \), \( w_t \) and \( u_t \) are i.i.d. processes, and \( w_t \) is a conditionally homoskedastic innovation component in \( Y_{2,t} \) as before. Note that (11) contains a conditionally heteroskedastic error term that is linear in the regressors. Substituting
\[ Y_{1,t} = (a_0 + \delta_{w,0} w_t + \delta_{u,0} u_t) + (a_1 + \delta_{w,1} w_t + \delta_{u,1} u_t) X_{1,t} + (a_2 + \delta_{w,2} w_t + \delta_{u,2} u_t) Y_{2,t}, \quad (12) \]

where the parameters are now functions of both \( w_t \) and \( u_t \). Conditioning the \( \tau_1^{th} \) quantile of \( Y_{1,t} \) recursively on the \( \tau_2^{th} \) quantile of \( Y_{2,t} \), we arrive at a QQ relationship

\[
Q_{Y_{1,t}}(\tau_1|X_{1,t}, Q_{Y_{2,t}}(\tau_2|X_{2,t})) \\
= (a_0 + \delta_{w,0} F^{-1}_w(\tau_2) + \delta_{u,0} F^{-1}_u(\tau_1)) + (a_1 + \delta_{w,1} F^{-1}_w(\tau_2) + \delta_{u,1} F^{-1}_u(\tau_1)) X_{1,t} \\
+ (a_2 + \delta_{w,2} F^{-1}_w(\tau_2) + \delta_{u,2} F^{-1}_u(\tau_1)) Q_{Y_{2,t}}(\tau_2|X_{2,t}) \\
\equiv \alpha_0(\tau_2, \tau_1) + \alpha_1(\tau_2, \tau_1) X_{1,t} + \alpha_2(\tau_2, \tau_1) Q_{Y_{2,t}}(\tau_2|X_{2,t}),
\]

where the last line expresses the same relationship as (10).

### 3 Estimation and Inference

#### 3.1 Estimation

We propose a power series approach to estimate the linear semiparametric QQ framework. Without loss of generality, let the dimension of \( X_{1,t} \) be one. To motivate the power series method, first rewrite (7) by adding and subtracting some terms

\[
Y_{1,t} = \alpha_0(F^{-1}_w(\tau_2), u_t) + \alpha_1(F^{-1}_w(\tau_2), u_t) X_{1,t} + \alpha_2(F^{-1}_w(\tau_2), u_t) Y_{2,t} \\
+ [\alpha_0(w_t, u_t) - \alpha_0(F^{-1}_w(\tau_2), u_t)] \\
+ [\alpha_1(w_t, u_t) - \alpha_1(F^{-1}_w(\tau_2), u_t)] X_{1,t} \\
+ [\alpha_2(w_t, u_t) - \alpha_2(F^{-1}_w(\tau_2), u_t)] Y_{2,t} \\
= \alpha_0(F^{-1}_w(\tau_2), u_t) + \alpha_1(F^{-1}_w(\tau_2), u_t) X_{1,t} + \alpha_2(F^{-1}_w(\tau_2), u_t) Y_{2,t} + \Psi_t(w_t, u_t),
\]

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

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

where $\hat{\Psi}$ controls for $\Psi$. Observing that $\Psi_t(w_t, u_t)$ 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 $F_w^{-1}(\tau_2)$.

Since this is a univariate series expansion, a simple power series may be used. In a multivariate setting, one suggestion is to employ orthonormalized series expansion that eliminates the cross-products in the expansion so that the number of approximating terms may be reduced (Andrews, 1991). Defining $w_t(\tau_2) = w_t - F_w^{-1}(\tau_2)$, the expansion yields

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

where $\alpha_{i,k}(u_t)$ is the $k^{th}$ derivative of $\alpha_i$ around $F_w^{-1}(\tau_2)$ so that the derivatives of $\alpha_i$ may be treated as functions of $u_t$ only. More generally, we may expand each coefficient using a different number of polynomials. Defining $\alpha_{i,k}(F_w^{-1}(\tau_2), u_t) \equiv \varphi_{\tau_2,i,k}(u_t)$, we utilize the truncated regression model

$$
H_{Y_1}(w_t(\tau_2), u_t; \alpha, \varphi) = \alpha_0(\tau_2, u_t) + \alpha_1(\tau_2, u_t)X_{1,t} + \alpha_2(\tau_2, u_t)Y_{2,t}
$$

$$
+ \varphi_{\tau_2,0,K_0}(u_t)' P_{0,K_0}(w_t(\tau_2)) + \varphi_{\tau_2,1,K_1}(u_t)' P_{1,K_1}(w_t(\tau_2)) + \varphi_{\tau_2,2,K_2}(u_t)' P_{2,K_2}(w_t(\tau_2)),
$$

where $P_{i,K_i}(w(\tau_2))$ is the $K_i$ polynomial in $w_t(\tau_2)$ while $\varphi_{\tau_2,i,K_i}(u_t)$ is a vector associated with the derivatives of $\alpha_i$. For instance, $\varphi_{\tau_2,1,K_1} = [\varphi_{\tau_2,1,1} \quad \varphi_{\tau_2,1,2} \quad \ldots \quad \varphi_{\tau_2,1,K_1}]'$ is the vector of derivatives omitting the $u_t$-argument and $P_{1,K_1}(w_t(\tau_2)) = [w_t(\tau_2) \quad w_t(\tau_2)^2 / 2! \quad \ldots \quad w_t(\tau_2)^{K_1} / K_1!]'$ is the vector of the polynomial so that

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

---

7This is provided that the $\alpha$ coefficients are sufficiently smooth functions of $w_t$. 

11
Since we have expanded $w_t$ in $\alpha_i$ around $F_w^{-1}(\tau_2)$, the only innovation term remaining in $\alpha_i$ and $\varphi_{\tau_2,i,K_i}$ is $u_t$. In other words, the expansion separates $u_t$ from $w_t$ in $\alpha_i$ so that after controlling for $w_t(\tau_2)$ in the nuisance term, all the parameters will be functions of $u_t$ alone. Hence $\alpha_i$, with its $w_t$-argument now anchored at $F_w^{-1}(\tau_2)$, can be estimated using standard quantile regression treating $u_t$ as the only source of innovation.

With the truncation, $H Y_1(w_t(\tau_2), F^{-1}_w(\tau_1); \alpha, \varphi)$ may be used to approximate the conditional quantile function of $Y_{1,t}$. Let the difference between the true and the approximate conditional quantile functions be $\Gamma_{0,t} + \Gamma_{1,t} + \Gamma_{2,t}$, where $\Gamma_{i,t}$ defines a remainder term associated with the expansion of $\alpha_i$. To consistently estimate the conditional quantile function, it is imperative for $\Gamma_{i,t}$ to be asymptotically negligible as the number of approximating terms $K_i$ in the polynomial increases 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. Andrews (1991) and Newey (1997) examined the conditions for consistency and asymptotic normality for series estimation in the ordinary least squares regression by using a truncation method. Lee (2007) employed series expansion and regression splines in quantile regression with endogenous variables.

The estimation follows a two-step procedure that is summarized below:

1. **Obtain $\hat{w}_t(\tau_2)$ as the residual**, i.e. $\hat{w}_t(\tau_2) = Y_{2,t} - \hat{\beta}(\tau_2)'X_{2,t}$, after estimating a $\tau_2$th quantile regression of $Y_{2,t}$, obtaining

   $$\hat{\beta}(\tau_2) = \arg \min_{\beta} T^{-1} \sum_{t=1}^{T} \rho_{\tau_2}(Y_{2,t} - \beta'X_{2,t}).$$

2. **Using $\hat{w}_t(\tau_2)$, estimate a $\tau_1$th quantile regression of $Y_{1,t}$**, obtaining

   $$(\hat{\alpha}(\tau_2, \tau_1), \hat{\varphi}_{\tau_2,K}(\tau_1)) = \arg \min_{\alpha} T^{-1} \sum_{t=1}^{T} \rho_{\tau_1}(Y_{1,t} - H_{Y_1}'(\hat{w}_t(\tau_2), u_t; \alpha, \varphi)).$$

Note that $w_t(\tau_2)$ can be estimated from the residual because the model in the first step is assumed to exhibit only location shift. In the special case of heteroskedasticity of the linear form, the estimation
becomes straightforward as the expansion is not required. Referring to (12) again, consider

\[ Y_{1,t} = (\alpha_0 + \delta_{w,0}w_t + \delta_{u,0}u_t) + (\alpha_1 + \delta_{w,1}w_t + \delta_{u,1}u_t)X_{1,t} + (\alpha_2 + \delta_{w,2}w_t + \delta_{u,2}u_t)Y_{2,t} \]

\[ = (a_0 + \delta_{w,0}F_w^{-1}(\tau_2) + \delta_{u,0}u_t) + (a_1 + \delta_{w,1}F_w^{-1}(\tau_2) + \delta_{u,1}u_t)X_{1,t} + (a_2 + \delta_{w,2}F_w^{-1}(\tau_2) + \delta_{u,2}u_t)Y_{2,t} \]

\[ + \delta_{w,0}w_t(\tau_2) + \delta_{w,1}w_t(\tau_2)X_{1,t} + \delta_{w,2}w_t(\tau_2)Y_{2,t}, \]

where we have substituted \( w_t = w_t(\tau_2) + F_w^{-1}(\tau_2) \) in the last line. For the above, estimating \( \alpha_i(\tau_2, \tau_1) = a_i + \delta_{w,i}F_w^{-1}(\tau_2) + \delta_{u,i}F_u^{-1}(\tau_1) \) requires us to control \( \Psi^{-1} \) by replacing \( w_t(\tau_2) \) in there with \( \hat{w}_t(\tau_2) \), then follow up with a \( \tau_1 \) quantile regression.

### 3.2 Inference

Inference will be based on the asymptotic distribution presented below. The discussion here considers some general series that includes the power series as a special case.

First, we lay down some notation. Define \( \alpha_i(\tau_2, u_t) \equiv \alpha_{i,\tau_2}(u_t) \) for \( i = 0, 1, 2 \) so that \( \alpha_{\tau_2}(u) = [\alpha_{0,\tau_2}(u_t) \alpha_{1,\tau_2}(u_t) \alpha_{2,\tau_2}(u_t)]' \). In addition, denote \( \alpha_{i,\tau_2}(F_u^{-1}(\tau_1)) \equiv \alpha_{i,\tau_2}(\tau_1) \). Let the original information vector at time \( t \), not including the polynomials from the series expansion, be \( X_{1,t} = [1 \ X_{1,t}' \ Y_{2,t}'] \). The design matrix is thus a \( T \times p \) matrix \( X_1 \). Without loss of generality, we consider a one-dimensional \( X_1 \) so that \( p = 3 \). For the other variables that have a time subscript, a lack of time subscript is used to denote their vector (matrix) counterparts if the variable is a scalar (vector).

As a result of the series expansion, the additional regressors will form a \( T \times \bar{\lambda} \) matrix of polynomials \( P_\lambda(w(\tau_2)) = [P_{0,K_0}(w(\tau_2)) \ P_{1,K_1}(w(\tau_2)) \ P_{2,K_2}(w(\tau_2))] \), where the number of terms in the polynomials is \( \bar{\lambda}(T) = K_0(T) + K_1(T) + K_2(T) \). The notation makes it explicit that the polynomials are functions of \( w(\tau_2) \). The design matrix then combines the original regressors \( X_1 \) and the polynomials \( P_\lambda(w(\tau_2)) \) to form \( X_1(w(\tau_2)) = [X_1 \ P_\lambda(w(\tau_2))] \), which has \( \lambda = p + \bar{\lambda} \) columns. 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 \( \hat{P} \equiv P_\lambda(\hat{w}(\tau_2)) \) and thus the design matrix \( \hat{X}_1 \equiv X_1(\hat{w}(\tau_2)) \). With appropriate regularity conditions, we have \( \hat{w}(\tau_2) = w(\tau_2) + o_p(1) \) so that \( \hat{X}_1 = X_1 + o_p(1) \). This is true as long as the estimated parameters in the first-step regression are
consistent, i.e. \( \hat{\beta}(\tau_2) = \beta(\tau_2) + o_p(1) \).

Let the coefficients on the polynomials be \( \varphi_{\tau_2} = [\varphi'_{0,\tau_2} \varphi'_{1,\tau_2} \varphi'_{2,\tau_2}]' \), bearing in mind that they are functions of \( u_t \), the innovation in \( Y_{1,t} \). Hence, the combined parameter vector is a \( \lambda \)-vector \( \theta_{\tau_2} = [\alpha'_{\tau_2} \varphi'_{\tau_2}]' \). The truncation of the series introduces a remainder term associated with each of the \( \alpha \) coefficient that has been expanded. The remainder is a multiplication of a \( T \times 3 \) vector \( \Gamma = [\Gamma_0 \ \Gamma_1 \ \Gamma_2] \) and a \( 3 \times 1 \) vector of ones denoted by \( i_3 \), where \( \Gamma_i \) is a \( T \times 1 \) vector of the remainder term associated with estimating \( \alpha_i \). In period \( t \) notation, \( \Gamma_t \) is a \( 3 \times 1 \) vector which is the transpose of the \( t^{th} \) row of \( \Gamma \).

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

\[
Y_{1,t} = \theta_{\tau_2}(\tau_1)'X_{1,t} + \Gamma'_ti_3 + u_t(\tau_1)
\]

and

\[
Y_{2,t} = \beta(\tau_2)'X_{2,t} + w_t(\tau_2),
\]

where \( \Gamma'_ti_3 = Q_{Y_{1,t}}(\tau_1|\bar{X}_{1,t}) - \theta_{\tau_2}(\tau_1)'X_{1,t} \) reflects the fact that \( \theta_{\tau_2}(\tau_1)'X_{1,t} \) is an approximation of \( Q_{Y_{1,t}}(\tau_1|\bar{X}_{1,t}) \). Since \( \bar{X}_{1,t} \) is unknown, feasible estimation requires replacing \( X_{1,t} \) with \( \hat{X}_{1,t} \) after obtaining \( \hat{w}_t(\tau_2) \) from the second equation. Using \( \hat{X}_{1,t} \) introduces a generated regressor problem that will have implications for inference, owing to the fact that we are actually estimating

\[
Y_{1,t} = \theta_{\tau_2}(\tau_1)'\hat{X}_{1,t} + \Gamma'_ti_3 + \underbrace{\theta_{\tau_2}(\tau_1)'(X_{1,t} - \hat{X}_{1,t})}_{\Phi_t} + u_t(\tau_1),
\]

where \( \Phi_t \) is a term that is introduced by using \( \hat{X}_{1,t} \), which is an additional source of imprecision that will lead to increasing the standard error of \( \hat{\theta}_{\tau_2}(\tau_1) \). This can be seen in its asymptotic distribution derived later in the section. We now examine the large sample properties of \( \hat{\theta}_{\tau_2}(\tau_1) \) and derive its asymptotic distribution. The large sample theory utilizes the following assumptions:

**Assumption A1.** Let \( \{Y_{1,t}, t \geq 1\} \) and \( \{Y_{2,t}, t \geq 1\} \) be \( L_1 \)-integrable sequences of random variables defined on the probability space \((\Omega, \mathcal{F}, P_1)\) and \((\Omega, \mathcal{F}, P_2)\) having a nondecreasing sub \( \sigma \)-fields \( \mathcal{G}_{i,0} \subset \mathcal{G}_{i,1} \subset \ldots \subset \mathcal{G}_i \) for \( i = 1, 2 \), where \( \mathcal{G}_{i,0} \) is the trivial \( \sigma \)-field, \( \mathcal{G}_{1,t-1} = \sigma(\{X_{1,j}\}_{j=0}^t, \{Y_{1,k}\}_{k=0}^{t-1}) \)
and $\mathcal{F}_{2,t-1} = \sigma(\{X_{2,j}\}_{j=0}^t, \{Y_{2,k}\}_{k=0}^{t-1})$.

**Assumption A2.** The $\lambda$-dimensional parameter space $\Theta$ is compact.

**Assumption A3.** There exist positive constants $s$, $\zeta$ and $K$, and a sequence of numbers $K\lambda^{-s}$ such that $\max_t \max_i |\Gamma_{i,t}| < K\lambda^{-s}$ and $\lambda = \mathcal{O}(T^\zeta)$.

**Assumption A4.** The conditional distribution function of $u_t(\tau_1)$, denoted by $F_t$, is continuously differentiable with conditional density $f_t$ that is bounded above by a constant $C_f^{\max}$ and bounded below by a constant $C_f^{\min}$ at $u_t(\tau_1) = 0$.

For the next assumption, define $\hat{D}_{1,T} = \hat{X}_1'\hat{X}_1/T$ and $D_{1,T} = X_1'X_1/T$. For each $\hat{\theta}_{\tau_2} \in \Theta$, define $\hat{M}_{1,T}(\hat{\theta}_{\tau_2}) = \hat{X}_1'\hat{F}(\hat{\theta}_{\tau_2})\hat{X}_1/T$ and $M_{1,T}(\hat{\theta}_{\tau_2}) = X_1'F(\hat{\theta}_{\tau_2})X_1/T$. In addition, let $\hat{F}(\hat{\theta}_{\tau_2})$ be a diagonal matrix with $t$ element $f(\eta\hat{Y}_t(\hat{\theta}_{\tau_2}))$ and $\hat{F}(\hat{\theta}_{\tau_2})$ be a diagonal matrix with $t$ element $f(\eta\hat{Y}_t(\hat{\theta}_{\tau_2}))$, where $0 < \eta < 1$, $\hat{Y}_t(\hat{\theta}_{\tau_2}) = (\hat{\theta}_{\tau_2} - \theta_{\tau_2}(\tau_1))'\hat{X}_{1,t} + \theta_{\tau_2}(\tau_1)'(\hat{X}_{1,t} - X_{1,t}) - \Gamma_i't^3$ and $\hat{\gamma}_t(\hat{\theta}_{\tau_2}) = (\hat{\theta}_{\tau_2} - \theta_{\tau_2}(\tau_1))'\hat{X}_{1,t} - \Gamma_i't^3$. For $M_{1,T}(\hat{\theta}_{\tau_2}(\tau_1))$, we simply denote $M_{1,T}$.

**Assumption A5.** $D_{1,T}$ and $M_{1,T}(\hat{\theta}_{\tau_2})$ converge to positive definite matrices $D$ and $M_{1}(\hat{\theta}_{\tau_2})$ respectively. The minimum eigenvalues of $M_{1,T}(\hat{\theta}_{\tau_2})$ and $M_{1}(\hat{\theta}_{\tau_2})$, defined by $K_{\min}(M_{1,T}(\hat{\theta}_{\tau_2}))$ and $K_{\min}(M_{1}(\hat{\theta}_{\tau_2}))$, are bounded away from zero for all $T$ and uniformly in $\hat{\theta}_{\tau_2} \in \Theta$.

**Assumption A6.** Let the $j$ element of $X_{1,t}$ be $X_{1,t}^{(j)}$. There exists a constant $\Delta$ such that $E|X_{1,t}^{(j)}|^3 \leq \Delta < \infty$ for all $t$ and $j = 1, \ldots, p$.

**Assumption A7.** $\hat{\beta}(\tau_2)$ is a consistent estimator of $\beta(\tau_2)$.

By assuming that $X_{1,t}$ is $\mathcal{F}_{1,t-1}$-measurable, A1 implicitly captures the fact that conditioning on $\mathcal{F}_{1,t-1}$ implies conditioning on $w_t$ also. Assumption A2 requires the compactness of the parameter set while Assumption A3 is required to bound the remainder term, which is also required in Andrews
(1991) and Newey (1997). We consider the case where the dimension of the approximating functions may increase at a polynomial rate that is controlled by $\zeta$.

Assumption A4 requires the density function of $u_t(\tau_1)$ to be bounded above. At $u_t(\tau_1) = 0$, its density must be bounded away from zero. Note that $u_t(\tau_1)$ is the re-centering of $u_t$ and such operation will not alter the shape of the distribution of $u_t$. Therefore, A4 may be restated for the conditional distribution and density functions of $u_t$ instead. Likewise, we may also restate A4 for the conditional distribution and density functions of $Y_{1,t}$ as its distribution is derived from the conditional distribution of $u_t$.

Assumption A5 and A6 impose the existence of certain moment conditions. These conditions are the counterparts of Assumption F(iii) of Andrews (1991) that requires random variables and series functions to be bounded, which in turn can be achieved if the series functions themselves are already bounded, an instance being the trigonometric series, or if the possibly unbounded series functions are defined on a bounded support. Assumption A7 implies that $\hat{\theta}(\tau_2)$ is a consistent estimator of $\theta(\tau_2)$ as $\hat{\theta}(\tau_2) = -(\hat{\beta}(\tau_2) - \beta(\tau_2))'X_{2,t}$ and $\hat{\beta}(\tau_2)$ is consistent by A7. This in turn can be weakened by imposing conditions that ensure consistency of the first step estimators.

We first proceed by establishing consistency in Proposition 1, then the rate of convergence in 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 $\hat{\theta}(\tau_2)$, which may be used to obtain the asymptotic distribution. The technical details of the proofs are relegated to Appendix B.

**Proposition 1.** (Consistency) Under A1-A4, A6 and A7, $\hat{\theta}(\tau_2) - \theta(\tau_2) = o_p(1)$.

That $\hat{\theta}(\tau_2)$ 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) Let $\|A\| = tr(A'A)^{1/2}$. Under A1-A7, if $s > (1 - 2\zeta)/2\zeta$ and $\zeta \in (0, 1/2)$, then $\|\hat{\theta}(\tau_2) - \theta(\tau_2)\| = O_p(\lambda/\sqrt{T})$.

The convergence rate of $\hat{\theta}(\tau_2) - \theta(\tau_2)$ may be inferred from Proposition 2 as $O_p(T^{\zeta-1/2})$. Incidentally, it is straightforward to establish that this is also the rate of convergence in mean-squared. The parameter $\zeta$, defined on the interval $(0, 1/2)$, clearly demonstrates the tension between the speed
of convergence of $\hat{\theta}_{\tau_2}(\tau_1)$ and the rate of decay of the remainder term. For univariate power series expansion considered in this paper, the parameter $s$ is also a smoothness parameter, being the number of times the $\alpha(w_t, u_t)$ coefficients are differentiable in $w_t$. By rearranging the condition in Proposition 2, we obtain $\zeta > 1/2(1 + s)$, reflecting the lower bound in the rate of growth in the number of approximating terms. Clearly, we must have $s \geq 1$ and the smoother $\alpha(w_t, u_t)$ is in $w_t$ the smaller is the minimum possible rate of growth in the number of approximating terms. Lee (2007) considered $\zeta < 1/8$ for the power series and our conditions in this context require that $s \geq 4$, which is slightly less restrictive than $s \geq 5$ required by Lee (2007).\footnote{We also obtain the same range for $\zeta$ as Zernov et al. (2009), who 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 $\zeta \in (0, 1/2)$.}

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\footnote{In this expression, we have used the fact that $f(\eta \Upsilon_t(\theta_{\tau_2}(\tau_1)))$ converges to $f(0)$ in probability. Hence, by Dominated Convergence Theorem, $E[f_t(\eta \Upsilon_t(\theta_{\tau_2}(\tau_1)))X_{1,t}^{(j)}\theta_{\tau_2}(\tau_1)]$ converges to $E[f_t(0)X_{1,t}^{(j)}\theta_{\tau_2}(\tau_1)]$ for each $j$ and $k$ element of $X_{1,t}$ and $\theta_{\tau_2}(\tau_1)$ respectively.}

$$\sqrt{T}(\hat{\theta}_{\tau_2}(\tau_1) - \theta_{\tau_2}(\tau_1)) = -\bar{M}_{1,T}^{-1}T^{-1/2}\sum_{t=1}^{T}X_{1,t}\psi_{\tau_1}(Y_t - \theta_{\tau_2}(\tau_1)'X_{1,t})$$

$$-\bar{M}_{1,T}^{-1}T^{-1}\sum_{t=1}^{T}E[f_t(0)X_{1,t}\theta_{\tau_2}(\tau_1)']X_{1,t}'X_{2,t}\sqrt{T}(\hat{\beta}(\tau_2) - \beta(\tau_2)) + o_p(1),$$

(13)

where $\bar{M}_{1,T} = E[M_{1,T}]$. Let $M_{3,T} = T^{-1}\sum_{t=1}^{T}E[f_t(\eta \Upsilon_t(\theta_{\tau_2}(\tau_1)))X_{1,t}\theta_{\tau_2}(\tau_1)']X_{1,t}'X_{2,t}$ be a $\lambda \times p_2$ matrix and $\beta$ be a $p_2$ vector, where $X_{1,t}' = \partial X_{1,t}(w_t(\tau_2))/\partial w$. We add the following assumptions:

**Assumption A5’**. A5 plus $M_{3,T}$ converges to a $M_3$ matrix with full column rank.

**Assumption A8.** $\sqrt{T}(\hat{\beta}(\tau_2) - \beta(\tau_2)) = N(0, \Omega_{\beta(\tau_2)})$, where $\Omega_{\beta(\tau_2)}$ is a $p_2 \times p_2$ asymptotic covariance matrix of $\sqrt{T}(\hat{\beta}(\tau_2) - \beta(\tau_2))$. 

With the additional assumptions, we have:

**Proposition 3.** (Asymptotic Normality) Under A1-A5’ and A6-A8, \( \sqrt{T} \Omega_{\theta_{\tau_2}(\tau_1)}^{-1/2} \left( \hat{\theta}_{\tau_2}(\tau_1) - \theta_{\tau_2}(\tau_1) \right) \Rightarrow N(0, I) \).

Using (13), the asymptotic covariance matrix of \( \sqrt{T}(\hat{\theta}_{\tau_2}(\tau_1) - \theta_{\tau_2}(\tau_1)) \) may be expressed as

\[
\Omega_{\theta_{\tau_2}(\tau_1)} = \tau_1 (1 - \tau_1) M_1^{-1} D_1 M_1^{-1} + M_1^{-1} M_3 \Omega_{\beta(\tau_2)} M_3' M_1^{-1}.
\] (14)

We may further refine the expression for \( \Omega_{\beta(\tau_2)} \), starting from the linear representation

\[
\sqrt{T}(\hat{\beta}(\tau_2) - \beta(\tau_2)) = T^{-1/2} M_{2,T}^{-1} \sum_{t=1}^{T} X_{1,t} \psi_{\tau_1} (Y_{2,t} - \beta(\tau_2)' X_{2,t}) + o_p(1),
\]

where \( M_{2,T} = T^{-1} \sum_{t=1}^{T} g_t(0) X_{2,t} X_{2,t}' \) is a \( p_2 \times p_2 \) matrix with a limiting matrix \( M_2 \) and \( g_t \) is the conditional distribution of \( w_t(\tau_2) \). Based on the above, we have \( \Omega_{\beta(\tau_2)} = \tau_2 (1 - \tau_2) M_2^{-1} D_2 M_2^{-1} \), where \( D_2 \) is the \( p_2 \times p_2 \) limiting matrix of \( D_{2,T} = T^{-1} \sum_{t=1}^{T} X_{2,t} X_{2,t}' \).

The asymptotic covariance matrix of \( \sqrt{T}(\hat{\theta}_{\tau_2}(\tau_1) - \theta_{\tau_2}(\tau_1)) \) in (14) is a general one that includes the possibility of conditional heteroskedasticity. For the actual estimation, it is more computationally convenient to treat \( w_t \) and \( u_t \) as both conditionally homoskedastic instead. In this case, we achieve further simplification of the covariance matrix formula as \( M_1(\theta_{\tau_2}(\tau_1)) = f(F^{-1}(\tau_1)) D_1 \), where \( f(F^{-1}(\tau_1)) \) in turn is \( f(0) \) given that \( F^{-1}(\tau_1) = 0 \). Under A5 and A7, \( D_1 \) may be consistently estimated using \( \hat{D}_{1,T} \) while \( \hat{f}(0) \) may be obtained by inverting the quantile density, i.e. \( \hat{s}(\tau_1) \), estimated using the nonparametric method of Siddiqui (1961). In addition, \( \hat{f}(0) \) is used to estimate \( M_3 \) through \( \hat{M}_{3,T} = \hat{f}(0) T^{-1} \sum_{t=1}^{T} \hat{\theta}_{\tau_2}(\tau_1)' \hat{X}_{1,t} \hat{X}_{1,t}' \). To estimate \( \Omega_{\theta_{\tau_2}(\tau_1)} \), we estimate \( D_2 \) using \( D_{2,T} \) and \( M_2 \) using \( \hat{g}(0) D_{2,T} \), where \( \hat{g}(0) \) is the inverse of the quantile density estimator and absence of the circumflex over \( D_{2,T} \) expresses the fact that no generated regressors are used to form this matrix.

To robustly estimate \( \Omega_{\theta_{\tau_2}(\tau_1)} \) in the presence of conditional heteroskedasticity, we may first estimate the covariance matrix of \( \sqrt{T}(\hat{\beta}(\tau_2) - \beta(\tau_2)) \) using \( \hat{\Omega}_{\beta(\tau_2)} = \tau_2 (1 - \tau_2) \hat{M}_{2,T}^{-1} D_{2,T} \hat{M}_{2,T}^{-1} \), where \( \hat{M}_{2,T} = T^{-1} \sum_{t=1}^{T} \hat{g}_t(0) X_{2,t} X_{2,t}' \) and \( \hat{g}_t(0) \) is the Hendricks-Koenker density estimator (see Koenker
2005, p. 80). The latter is

\[
\hat{g}_t(0) = \max \left\{ 0, \frac{2b_k}{X'_{2,t} \hat{\beta}(\tau_2 + b_k) - X'_{2,t} \hat{\beta}(\tau_2 - b_k) - e} \right\},
\]

where \(e\) is a small number to prevent division by zero and \(b_k\) is the bandwidth with the Bofinger (1975) and Hall and Sheather (1988) bandwidths as possible candidates. Then to estimate \(\hat{\Omega}_{\hat{\theta}_2(\tau_1)}\) robustly, we use the robust estimator of \(\hat{\Omega}_{\hat{\beta}(\tau_2)}\) together with \(\hat{D}_{1,T}, \hat{M}_{1,T} = T^{-1} \sum_{t=1}^{T} \hat{f}_t(0) \hat{X}_{1,t} \hat{X}'_{1,t} \) and \(\hat{M}_{3,T} = T^{-1} \sum_{t=1}^{T} \hat{f}_t(0) \hat{X}_{1,t} \hat{X}'_{1,t} \hat{\theta}_{\tau_2}(\tau_1) X'_{2,t}\), where \(\hat{f}_t(0)\) is the Hendricks-Koenker density estimator.

### 3.3 Monte Carlo Simulation

In this section, we compare the performance of the model under various assumptions about the order of the polynomial. Consider the data generating process

\[ Y_{1,t} = a_0 + a_1 X_{1,t} + (a_2 + \delta(\lambda e^w_t + u_t)) Y_{2,t} \tag{15} \]

and

\[ Y_{2,t} = b_0 + b_1 X_{1,t} + b_2 X_{2,t} + w_t, \tag{16} \]

where \((a_0, a_1, a_2, \delta, \lambda) = (3, 4, 4, 5, 3), (b_0, b_1, b_2) = (1, 2, 3), X_{1,t} \sim t_3, X_{2,t} \sim N(15, 2), w_t \sim N(0, 0.5)\) and \(u_t \sim N(0, 1)\). This generating process is similar to the benchmark model of Ma and Koenker (2006), except that the slope coefficient on \(Y_{2,t}\) in (15) is specified as \(a_2 + \delta(\lambda e^w_t + u_t)\) while they adopted a specification of \(a_2 + \delta(\lambda w_t + u_t)\) instead. In modifying the data generating process of Ma and Koenker, the slope coefficient on \(Y_{2,t}\) is now a nonlinear function of \(w_t\) and this motivates the power series expansion. If their data generating process is adopted, the expansion is no longer needed as the primary regression can be motivated as having an error term that exhibits conditional heteroskedasticity of the linear form. We would like to estimate \(\alpha_2\) from

\[ Y_{1,t} = a_0 + a_1 X_{1,t} + \alpha_2(w_t, u_t) Y_{2,t}, \]

understanding that \(\alpha_2(w_t, u_t)\) is an unknown function of \(w_t\) and \(u_t\) from the researcher’s perspective.
The true value \( \alpha_2(\tau_2, \tau_1) \equiv \alpha_2(F_{\tau_2}^{-1}(\tau_2), F_{\tau_1}^{-1}(\tau_1)) \), which is of interest, is \( 4 + 5(3 \exp(F_{\tau_2}^{-1}(\tau_2)) + F_{\tau_1}^{-1}(\tau_1)) \). To estimate this, we consider expansions where polynomials of \( w_t(\tau) \) up to the tenth order are included. This model can be expressed as

\[
H_{Y_{1.t}}(\hat{w}_t(\tau_2)) = a_0 + a_1 X_{1.t} + \alpha_2 Y_{2.t} + \sum_{k=1}^{I} \phi_k \frac{\hat{w}_t(\tau_2)^k}{k!} Y_{2.t}
\]

(17)

where \( k \) indexes the power of \( \hat{w}_t \) with \( I \) ranging from 1 to 10. The Monte Carlo experiment is carried out by simulating data from (15) and (16) 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 points in \( \tau \), resulting in 81 regressions corresponding to each \( \tau_1 \) and \( \tau_2 \) located on the grid. The exercise employs 1000 simulations. Confining to \( \tau_1 = \tau_2 = \tau \), Tables 1 and 2 report the true parameter values, the average of the estimated values of the simulations, together with their bias and root mean-squared error (RMSE) corresponding to 500 and 1000 generated observations.

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 model (Panel B) to the cubic model (Panel C) where the estimates are obtained based on simulations with 500 observations. In these plots, the larger estimated values are more lightly shaded. From the panels, it can be seen 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. However, the estimated surface becomes very close to the true surface even in the extremes when a quadratic, cubic or quartic model is used.

Comparing Tables 1 and 2, it can be seen that the bias and RMSE decline as the sample size increase. Fixing the sample size, the average of the RMSE across \( \tau \) (of the nine reported for each \( \tau \)) declines to a minimum as the order of expansion increases from the first to the third. However, the average RMSE begins to rise as the order of expansion increases to the forth and beyond. While the cubic model has the smallest average RMSE, the quintic model has the smallest squared bias. This exercise therefore recommends using a third to fifth order power series expansion to estimate a model of similar sample sizes.
4 Empirical Results

For the actual empirical implementation, we consider two policy instruments: M1 and M2 money supply.\textsuperscript{10} For the monetary process equation, we regress the monetary instrument on twelve of its own lags as well as the first lag of the change in Treasury yield.\textsuperscript{11} For the output process equation, we specifically consider

\[ y_t = \alpha_I(w_t, u_t) + \sum_{i=1}^{K_y} \alpha_{y,i}(w_t, u_t)y_{t-i} + \alpha_r(w_t, u_t)\Delta r_{t-1}. \] (18)

A possible future extension of (18) is to allow the intercept and slope parameters respond directly to the quantiles of the lagged shocks as well. At this juncture, estimating an impulse response function of quantiles as such is beyond the scope of the paper’s theoretical work. We justify the inclusion of \( w_t \) but not its lags as we expect from Cover’s findings that the contemporaneous negative shock would have the largest impact on output growth. In a rather ad-hoc manner, the task of controlling for the lagged effects of money supply shocks rests on the ability of the lagged change in Treasury yield to capture the history of the shocks themselves.

We consider (18) with twelve lags of output growth and estimate \( \alpha_I(\tau_2, \tau_1), \alpha_{y,i}(\tau_2, \tau_1) \) and \( \alpha_r(\tau_2, \tau_1) \) using a cubic expansion given that the cubic model has produced the smallest average RMSE in our Monte Carlo exercise. For parsimony, the paper only considers expanding the coefficients on the first four lags of output growth, i.e. \( \alpha_{y,i}(\tau_2, \tau_1) \) for \( i = 1, \ldots, 4 \).

Monthly time series from Datastream is used. Output growth is defined as the growth rate of the industrial production index. 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. Of particular interest is the response of the intercept term, which we will examine next, given the original location shift

\textsuperscript{10}While the Federal funds rate is also a monetary policy instrument, the upper quantiles of the Federal funds innovation will represent restrictive policies while the lower quantiles will be expansive so that the indexation scheme is opposite to the one when the money supply is used.

\textsuperscript{11}In his model, Cover also included the lagged government budget surplus and the ratio of the unemployed over the employed. However, these variables were usually statistically insignificant.
specification implicit in Cover’s model as shown in (4).

**The Intercept Term**

Panels A and B plot the surface of the intercept term when the monetary instruments are M1 and M2 money supply growth respectively. Not surprisingly, the surface is downward sloping as the quantile of output growth declines, implying that the intercept term falls with lower quantiles of output growth. When money supply shock influences output growth, the surface would also vary along the quantiles of the shock, where an increasingly restrictive shock is located in the left tail of the shock’s distribution. Hence, the surface should be tilted towards the \((0, 0, z_{\text{min}})\) vertex should money supply shock be non-neutral.

Focusing on M1 money supply as the monetary instrument, the “tilt” in the intercept surface is particularly pronounced in the upper quantiles of output growth. This implies that when output growth is large, restrictive shocks to M1 money supply are particularly effective in slowing down output growth than it is so when applied to output growth arising from the middle to lower quantiles of the distribution. When M2 is the monetary instrument, the “tilt” in the surface appears to be more consistent throughout the output growth distribution. Thus for the middle to the lower quantiles of output growth, restrictive shocks to M2 money supply are more effective in reducing output growth than are restrictive shocks to M1. Only in the right tail of output growth does the response of output growth appear to be similar for restrictive policies based on either M1 or M2.

Now, do money supply shocks influence output growth asymmetrically? To address this, the subplots in Figure 3 demonstrate how the intercept term responds to the various quantiles of money supply shock when output growth is at the 10th, 50th and 90th percentiles. Panels A and B of Figure 3 display these subplots when M1 and M2 money supply are adopted as the monetary instrument respectively. The dotted horizontal line in each subplot reflects the value of the intercept term when the money supply shock is at its median. Relative to the median money supply shock, we refer the quantiles of money supply shock below the median as restrictive and those above as expansive.

When output growth is at the 10th percentile, Panel A shows that both restrictive and expansive M1 money supply shock are ineffective in shifting the intercept away from the horizontal line. However, the median output growth is influenced by M1 money supply shock asymmetrically. Here, ceteris
paribus, output growth stays around 0.23% per month for values of M1 money supply shocks that are at least the median, but declines to 0.15% per month when the shock declines to the 10th percentile, indicating that restrictive policy is more effective when applied to the median output growth. Such asymmetry is also observed at the 90th percentile of output growth, since changing the monetary stance from the median to the 90th percentile increases output growth by 13 basis points (bps) per month, but decreases it by 20 bps per month when the money supply shock is restricted from the median to the 10th percentile. Therefore, restrictive M1 money supply shocks are more influential than expansive shocks when applied to output growth that is typical or is bullish, not when it is bearish.

However, when M2 money supply is considered, Panel B suggests that all 10th, 50th and 90th percentiles of output growth respond asymmetrically to restrictive as opposed to expansive shocks. For instance, Panel B shows that the 10th percentile of output growth increases only by 10 bps per month when the money supply shock increases from the median to the 90th percentile, but declines by 23 bps per month when the shock tightens from the median to the 10th percentile. The same experiment adds -4 bps (9 bps) to the median (90th percentile) output growth in the expansive direction but contracts 16 bps (22 bps) in the restrictive direction.

To further observe the asymmetric effect of money supply shocks, Figure 4 plots the change in the intercept term brought about by restrictive versus expansive policy, where restrictive policy here specifically refers to lowering the money supply shock from the median to the 10th percentile while expansive policy refers to raising it from the median to the 90th percentile. The solid (dotted) line in Figure 4 plots the change in the intercept term, for each quantile of output growth, following a restrictive (expansive) policy. This change is specifically defined as $\alpha_I(0.5, \tau_1) - \alpha_I(0.1, \tau_1)$ when restrictive policy is applied to some $\tau_1^{th}$ quantile of output growth and $\alpha_I(0.9, \tau_1) - \alpha_I(0.5, \tau_1)$ when otherwise. Therefore, the two lines will look identical if both restrictive and expansive policies affect output growth symmetrically. Clearly, whether M1 or M2 money supply is considered, the lines deviate substantially at each quantile of output growth, implying the money supply shock has asymmetric effects regardless the state of output growth.

Figure 4 also addresses the question: when applied to different levels of output growth, would the effects of the same restrictive or expansive monetary policy stance differ? Unambiguously, the
answer is yes. On the contrary, if the same restrictive or expansive stance affects all quantiles of output growth similarly, then the lines in Figure 4 will be horizontal. In Panel A, both restrictive and expansive M1 money supply shocks are clearly more effective when output growth is large while Panel B shows that both restrictive and expansive M2 money supply shocks are more effective in both tails of output growth.

As a final note, we would like to suggest that the asymmetry in the money-output relationship is not necessarily the same as “pushing on the string”. Strictly interpreted, the latter describes an asymmetric money-output phenomenon where restrictive policies are effective but expansive policies are not. Focusing on M2, “pushing on the string” therefore appears to be relevant when output growth is at center of its distribution as the influence of restrictive M2 is effective but that of expansive M2 is very weak. However, elsewhere in the output growth distribution, even though restrictive shocks are more effective than expansive shocks, the latter are not ineffective. For instance, the M2 money supply shock may influence the tails of output growth asymmetrically, but expansive shocks are effective as well even though they are not as influential as restrictive shocks.

**Slope on $y_{t-1}^{12}$**

Panels C and D of Figure 2 plot the estimated slope surface for $y_{t-1}$ corresponding to M1 and M2 money supply. Contrasting the intercept surfaces, both figures show that the slope tends to be flat across most quantiles of money supply shock and output growth, although it is slightly more elevated in the left tail of output growth. The latter implies that the left tail of contemporaneous output growth tends to react slightly more strongly to lagged output growth than elsewhere in its distribution.

Consider the effects of varying the monetary stance shown in the section plot of Figure 5. Except at the 10th percentile of output growth, Panel A shows that the slope on $y_{t-1}$ responds very weakly to changes in the M1 money supply shock. This weak response is also generally true when M2 money supply is used. Hence, the relationship between the quantile of contemporaneous output growth and the first lag of output growth is generally robust to variations in the money supply shock.

---

12To keep the discussion concise, estimation results related to the higher lags of output growth are omitted.
Slope on $\Delta r_{t-1}$

Panels E and F of Figure 2 plot the estimated slope surface for $\Delta r_{t-1}$ corresponding to M1 and M2 money supply. Both figures show that the surface is usually elevated in the lower quantiles of output growth and money supply shock. When M2 money supply is used, the surface appears to be like a saddle, displaying larger elevations in both analogous tails of output growth and money supply shock and depression around the center.

Figure 6 presents the section plot. In Panel A, the slopes on $\Delta r_{t-1}$ in the 10th and 50th percentile output growth regressions are not statistically different from zero along most quantiles of the money supply shock. At the 90th percentile of output growth, the null of zero can generally be rejected for the estimated slope. This observation is also similar when M2 money supply is used, where at the 90th percentile of output growth, the slope is statistically significant when the money supply shock is less than or equal to its median. This implies that a large increase in $\Delta r_{t-1}$ may adversely affect output growth in the next period especially when output growth is in the right tail.

5 Conclusion

Using a newly developed quantile-based framework, this paper further investigates into the nonlinearities that may be present between how money supply shock and output growth are related. First, it examines whether the same quantile of output growth responds differently to various quantiles of money supply shock, and finds that each quantile of output growth is typically more responsive towards restrictive than expansive shocks. Second, it investigates whether the same restrictive or expansive money supply shock may affect output growth differently when output growth is large or small. When M1 money supply is used, the right tail of output growth is disproportionately more sensitive to both restrictive and expansive money supply shocks. When M2 money supply is used, both tails of output growth become more sensitive than the center is to these shocks.

As the paper uses the money supply as the instrument, one natural extension is to consider a model with interest rate as the instrument instead. In addition, the paper estimates a reduced-form relationship between the money supply shock and output growth. A relevant extension, which is also in line with using the interest rate, is to estimate a structural relationship possibly motivated from a
New Keynesian DSGE model. This research agenda is currently pursued by the author.

References


Ma, Lingjie, and Roger Koenker, “Quantile regression methods for recursive structural equation mod-


Appendix A

Linear Quantile Regression: A Brief Review

Linear econometric models may exhibit location, or scale shift, or both under specific assumptions about the structure of the error term. The location shift model arises when the conditional quantiles are differentiated only by the intercept while the slope coefficients remain the same. The model exhibits scale shifts if variations in the slope coefficients contribute to differentiating the conditional quantiles. A simple case of a model exhibiting both location and scale shift is

\[ Y_t = a_0 + a_1 X_t + (\delta_0 + \delta_1 X_t) u_t, \]  

so that (19) is a model with a conditionally heteroskedastic error term \((\delta_0 + \delta_1 X_t) u_t\), where \(u_t\) is an i.i.d. process. By rewriting (19) as

\[ Y_t = (a_0 + \delta_0 u_t) + (a_1 + \delta_1 u_t) X_t, \]  

we can see that the innovation term \(u_t\) is a shifter of both the intercept and slope parameter. With the monotonicity of \(Y_t\) in \(u_t\), it can be argued from (20) that the \(\tau\)th quantile of \(Y_t\) conditioning on \(X_t\) directly corresponds to the \(\tau\)th quantile of \(u_t\).

Let \(F_u(\cdot)\) be the distribution function of \(u_t\) and \(F_u^{-1}(\tau)\) be its \(\tau\)th quantile. The \(\tau\)th conditional quantile of \(Y_t\) may be expressed as

\[ Q_{Y_t}(\tau|X_t) = (a_0 + \delta_0 F_u^{-1}(\tau)) + (a_1 + \delta_1 F_u^{-1}(\tau)) X_t \]
\[ = \alpha_0(\tau) + \alpha_1(\tau) X_t. \]

When \(\delta_1 = 0\), the error term in (20) becomes conditionally homoskedastic. The conditional quantile
of $Y_t$ then becomes $Q_{Y_t}(\tau|X_t) = \alpha_0(\tau) + a_1 X_t$, exhibiting only location shift. Likewise, if $\delta_0 = 0$, the conditional quantile of $Y_t$ will be purely differentiated by changes in the scale parameter.

Another way to look at (19) is to first define $u_t(\tau) = u_t - F_u^{-1}(\tau)$ so that the $\tau^{th}$ quantile of $u_t(\tau)$ is re-centered to zero. Then, substituting $u_t = u_t(\tau) + F_u^{-1}(\tau)$ into (20), we may express $Y_t$ as

$$Y_t = (a_0 + \delta_0 F_u^{-1}(\tau)) + (a_1 + \delta_1 F_u^{-1}(\tau))X_t + (\delta_0 + \delta_1 X_t)u_t(\tau)$$

$$= Q_{Y_t}(\tau|X_t) + u_t(\tau|X_t),$$

(21)

where we have defined $u_t(\tau|X_t) = (\delta_0 + \delta_1 X_t)u_t(\tau)$, which also has a $\tau^{th}$ conditional quantile of zero. The last expression in (21) is called a quantile representation and is convenient for elucidating what estimation in quantile regression entails as estimating $Q_{Y_t}(\tau|X_t)$ involves searching for both the intercept and slope coefficients that renders the $\tau^{th}$ quantile of $\hat{u}_t(\tau|X_t)$, the sample analog of $u_t(\tau|X_t)$, to zero. In the population context, the population parameters $\alpha(\tau)$ are those that minimize

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

where $\rho_\tau(u) = (\tau - \mathbb{I}(u < 0))u$ is an asymmetric loss function. These parameters also set the population score function $E[X_t \psi_\tau(Y_t - \alpha' X_t)]$ to zero, where $\psi_\tau(u) = \tau - \mathbb{I}(u < 0)$. In the actual estimation, the population quantile objective function is replaced with a sample analog which $\hat{\alpha}(\tau)$ minimizes:

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

---

13 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 $\tau^{th}$ quantile of $\hat{u}_t(\tau|X_t)$.

14 Notice that $E[\mathbb{I}(u_t(\tau|X_t) < 0)|X_t] = \tau$ holds because the $\tau^{th}$ conditional quantile of $u_t(\tau|X_t)$ is zero. Therefore, since $u_t(\tau|X_t) = Y_t - \alpha(\tau)'X_t$ by construction, it can be argued using the law of iterated expectations that $\alpha(\tau)$ will set the population score function to zero.

15 Recognizing that $E[\mathbb{I}(u_t(\tau|X_t) < 0)|X_t] - \tau = 0$ is a population moment provides the basis for method of moments estimation for quantiles as discussed by Chernozhukov and Hansen (2005).
The above objective function is differentiable except at $Y_t = \alpha'X_t$, yielding the sample score function as

$$W(\alpha) = T^{-1} \sum_{t=1}^{T} X_t \psi(\tau_t) (Y_t - \alpha'X_t).$$

The sample score function evaluated at $\hat{\alpha}(\tau)$ will be zero except on a measure zero set, where the quantile objective function is not differentiable, and this set corresponds to the set of observations satisfying $Y_t = \hat{\alpha}(\tau)'X_t$.

**Appendix B**

**Derivations and Proofs**

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

$$L_T(\tilde{\theta}) = T^{-1} \sum_{t=1}^{T} \left[ \rho_{\tau_1} (u_t(\tau_1) - ((\tilde{\theta} - \theta)'\hat{X}_{1,t} + (\hat{X}_{1,t} - \bar{X}_{1,t})'\theta - \Gamma_{i3}^t)) - \rho_{\tau_1} (u_t(\tau_1)) \right]$$

(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). $\hat{\theta}$ is the minimizer of (22) and the first order condition is

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

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

**Lemma 1.** *(Uniform Law of Large Numbers)* Under A1-A4, A6 and A7, the objective function, $L_T(\tilde{\theta})$, defined in (22), satisfies

$$\sup_{\tilde{\theta} \in \Theta} \left| L_T(\tilde{\theta}) - E[L_T(\tilde{\theta})] \right| \xrightarrow{p} 0$$

as $T \to 0$. 

31
Proof: Using Knight’s identity, i.e. \( \rho_r(u - v) - \rho_r(u) = -v \psi_r(u) + \int^v_0 I(0 < u \leq e) de \), and letting 
\[ \tilde{\gamma}_t(\theta) = (\bar{\theta} - \theta)'\tilde{X}_{1,t} + \theta' (\tilde{X}_{1,t} - X_{1,t}) - \Gamma'_i i_3, \]
we may express

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

\[
+ \sup_{\theta \in \Theta} |T^{-1} \sum_{t=1}^T \left( \int_0^{\tilde{\gamma}_t(\theta)} I(0 < u_t(\tau_1) \leq e) - E[\int_0^{\tilde{\gamma}_t(\tilde{\theta})} I(0 < u_t(\tau_1) \leq e) de] \right)|. \tag{24}
\]

We now show that (23) is \( o_p(1) \). 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) is \( o_p(1) \) pointwise. Hence, consider some \( \tilde{\theta}_t \in \Theta \). Applying Chebyshev inequality and the law of total variance, we have

\[
P \left( |T^{-1} \sum_{t=1}^T (\tilde{\gamma}_t(\tilde{\theta}_t) \psi_{\tau_1}(u_t(\tau_1)) - E[\tilde{\gamma}_t(\tilde{\theta}_t) \psi_{\tau_1}(u_t(\tau_1))]| \geq \delta/2 \right) \\
\leq \frac{4}{\delta^2} T^{-1} E[Var[[((\tilde{\theta}_t - \theta)^' \tilde{X}_{1,t} + \theta' (\tilde{X}_{1,t} - X_{1,t}) - \Gamma'_i i_3) \psi_{\tau_1}(u_t(\tau_1))]|X_{1,t}]] \\
\leq \frac{4}{\delta^2} T^{-1} E[((\tilde{\theta}_t - \theta)^' \tilde{X}_{1,t} + \theta' (\tilde{X}_{1,t} - X_{1,t}) - \Gamma'_i i_3)^2 Var[\psi_{\tau_1}(u_t(\tau_1))]|X_{1,t}]] \\
\leq \frac{4\gamma_1(1 - \tau_1)}{\delta^2} T^{-1} E[(9K^2 \lambda^{-2s} + 6K \lambda^{-s})|\tilde{\theta}_t - \theta)^' \tilde{X}_{1,t}| + 6K \lambda^{-s}|\theta' (\tilde{X}_{1,t} - X_{1,t})|] \\
+ 2 |(\tilde{\theta}_t - \theta)^' \tilde{X}_{1,t}|(\tilde{X}_{1,t} - X_{1,t})'\theta| + |(\tilde{\theta}_t - \theta)^' \tilde{X}_{1,t}|^2 + |(\tilde{X}_{1,t} - X_{1,t})'\theta|^2 \\
= \Theta(T^{-1}),
\]

which implies that A2 of Newey (1991) is satisfied, where the last line follows from the application of A3, A6, A7 and the Monotone Convergence Theorem. To verify A3a of Newey (1991), consider 
\[ |T^{-1} \sum_{t=1}^T \psi_{\tau_1}(u_t(\tau_1))(\tilde{\gamma}_t(\tilde{\theta}) - \tilde{\gamma}_t(\theta))| \leq T^{-1} \sum_{t=1}^T |\tilde{\gamma}_t(\tilde{\theta}) - \tilde{\gamma}_t(\theta)| \leq T^{-1} \sum_{t=1}^T \max_j |\tilde{\gamma}_t^{(j)}(\tilde{\theta}) - \tilde{\gamma}_t^{(j)}(\theta)|, \]
where \( \tilde{\gamma}_t^{(j)} \) denotes the \( j \) element of \( \tilde{\theta} \). Now, \( \sum_{j=1}^\lambda |\tilde{\theta}_t^{(j)} - \theta_t^{(j)}| = \|\tilde{\theta} - \theta\|_1 \) is a Manhattan norm while \( T^{-1} \sum_{t=1}^T \max_j |\tilde{X}_{1,t}^{(j)}| = \Theta(1) \). These two conditions are sufficient for A3a of Newey (1991) to hold and thus (23) is \( o_p(1) \) following Corollary 2.2 of Newey (1991).

Next, we show that (24) is \( o_p(1) \). To do so, we verify the assumptions of Andrews (1987). Assumption A1 of Andrews (1987) requires that the parameter space be compact, which is A2 in this
paper. It is also straightforward to verify the counterpart of Assumption A2a of Andrews (1987), which imposes that \( \int_0^{T_t} (I(0 < u_t(\tau_1) \leq e)) \, de \) is a random variable and for \( \hat{\theta} \in \| \hat{\theta} - \theta \| \) where \( \| \hat{\theta} - \theta \| \) is sufficiently small, \( \sup_{\hat{\theta}} \int_0^{T_t} (I(0 < u_t(\tau_1) \leq e)) \, de \) and \( \inf_{\hat{\theta}} \int_0^{T_t} (I(0 < u_t(\tau_1) \leq e)) \, de \) are random variables for all \( \hat{\theta} \in \Theta \).

Now, we verify Assumptions A2b and A3 of Andrews (1987). Assumption A2b requires that \( T^{-1} \sum_{t=1}^{T} \int_0^{T_t} (I(0 < u_t(\tau_1) \leq e)) \, de \) satisfies pointwise law of large numbers. To do so, we verify that \( E[\int_0^{T_t} (I(0 < u_t(\tau_1) \leq e)) \, de] < \infty \). Without loss of generality, assume that \( \hat{T}_t(\hat{\theta}) = (\hat{\theta} - \theta)^\prime X_{1,t} + \theta'(\tilde{X}_{1,t} - \tilde{X}_{1,t}) - \Gamma_t' i_3 > 0 \). By A7, \( \tilde{X}_{1,t} \overset{p}{\rightarrow} X_{1,t} \). Since \( \| \hat{\theta} - \theta \|^\prime \tilde{X}_{1,t} \) is bounded above by a multiple of \( |Y_{1,t}| \) which is integrable by A1, Dominated Convergence Theorem implies that \( (\hat{\theta} - \theta)^\prime \tilde{X}_{1,t} \) in \( \hat{T}_t(\hat{\theta}) \) may be replaced with \( (\hat{\theta} - \theta)^\prime X_{1,t} \) with an \( o(1) \) error. Likewise, we may drop \( \theta'(\tilde{X}_{1,t} - X_{1,t}) \) in \( \hat{T}_t(\hat{\theta}) \) as the expectation of this term is \( o(1) \) by the Monotone Convergence Theorem. Therefore, considering \( |(\hat{\theta} - \theta)^\prime X_{1,t} + |\Gamma_t' i_3| > (\hat{\theta} - \theta)^\prime \tilde{X}_{1,t} - \Gamma_t' i_3 > 0 \), we check that \( E[\int_0^{T_t} (I(0 < u_t(\tau_1) \leq e)) \, de] \leq E[\int_0^{\max} (3K \lambda^{-s} + |(\hat{\theta} - \theta)^\prime X_{1,t}|) \, de] \leq C_f^\max E[(3K \lambda^{-s} + \| \hat{\theta} - \theta \|^\prime \max_j |X^{(j)}_{1,t}|)^2] = O(1) \) where we have used A4 to bound the density function and \( O(1) \) follows from A3 and A6. With a bounded first moment, the pointwise law of large numbers follows.

To verify A3 of Andrews (1987), we need to show that for all \( \theta \in \Theta \), \( \sup_{\hat{\theta} \in \Theta} \| \hat{\theta} - \theta \| \rightarrow 0 \), we have

\[
\sup_{t \geq 1} \left| T^{-1} \sum_{t=1}^{T} \left( E[\int_0^{T_t} (I(0 < u_t(\tau_1) \leq e)) \, de] - E[\int_0^{T_t} (I(0 < u_t(\tau_1) \leq e)) \, de)] \right) \right| \rightarrow 0. \quad (25)
\]

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 \( (\hat{\theta} - \theta)^\prime X_{1,t} - \Gamma_t' i_3 > 0 \), we
have

\[
E\left[\sup_{\theta \in \Theta} (\bar{\theta} - \theta) \chi_{1,t} - \Gamma_t^{i_3}\right]
- \left\{I(0 < u_t(\tau_1) \leq e)de\right]\right] - E\left[\int_0^{\bar{\theta} - \theta) \chi_{1,t} - \Gamma_t^{i_3}\right]
\leq E\left[\int_0^{\sup_{\theta \in \Theta} (\bar{\theta} - \theta) \chi_{1,t} + 3K\lambda^{-s}} - ((\bar{\theta} - \theta) \chi_{1,t} - 3K\lambda^{-s})^2\right]
\leq \frac{C_f}{f} \left[2\sup_{\theta \in \Theta} ((\bar{\theta} - \theta) \chi_{1,t} + 6K\lambda^{-s} + \sup_{\theta \in \Theta} (\bar{\theta} - \theta) \chi_{1,t})\right]
\leq \frac{C_f}{f} \left[2\max_j |\chi_{1,t}^{(j)}| \sup_{\theta \in \Theta} ((\bar{\theta} - \theta) ||(6K\lambda^{-s} + \max_j |\chi_{1,t}^{(j)}| \sup_{\theta \in \Theta} (\bar{\theta} - \theta))\right]
\leq 0,

where the last inequality follows from A7, \(\sup_{\theta \in \Theta} ||\bar{\theta} - \theta|| \leq 2\sup_{\theta \in \Theta} ||\bar{\theta} - \theta|| \to 0\) 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. \(\square\)

**Proof of Proposition 1:** Define \(L(\theta) = E[L_T(\theta)]\). Clearly, \(L\) is minimized at \(\theta\). Following the argument in Theorem 2.1 of Newey and McFadden (1994), we have \(L(\theta) < L_T(\theta) + \delta/3 < L_T(\theta) + 2\delta/3 < L(\theta) + \delta\), 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 \(\hat{\theta}\) being a minimizer of \(L_T\). Focusing on the last term, express the summand in \(L(\theta)\) using Knight’s identity as \(\rho_{\tau_1}(u_t(\tau_1) - \Gamma_t^{i_3}) - \rho_{\tau_1}(u_t(\tau_1)) = -\Gamma_t^{i_3}\psi_{\tau_1}(u_t(\tau_1)) + \int_0^{\Gamma_t^{i_3}} \mathbb{I}(0 < u_t(\tau_1) \leq e)de\). Taking conditional expectations and using A3, we can show that \(L(\theta) = E[f(\eta\Gamma_t^{i_3})\psi_{\tau_1}(u_t(\tau_1)) + \int_0^{\Gamma_t^{i_3}} \mathbb{I}(0 < u_t(\tau_1) \leq e)de\). Since \(\lambda \to 0\) as \(T \to 0\), so that in turn \(L(\theta) \to 0\) by the continuity of \(L(\cdot)\), we have \(L(\hat{\theta}) < \delta\) asymptotically for any arbitrary \(\delta\). \(\square\)

**Lemma 2. (Stochastic Equicontinuity)** Under A1-A7,

\[
\sup_{||\theta - \hat{\theta}|| \leq \epsilon} \left\|\mathbb{W}_{T,\tau_1}(\hat{\Theta}(\hat{\theta})) - \mathbb{W}_{T,\tau_1}(0) - E[\mathbb{W}_{T,\tau_1}(\hat{\Theta}(\hat{\theta})) - \mathbb{W}_{T,\tau_1}(0)]\right\| = o_p(\lambda/\sqrt{T})
\]
where $\varepsilon_T = (\lambda/\sqrt{T}) \log T^\kappa$.

**Proof:** Let $\varepsilon_T = (\lambda/\sqrt{T}) \log T^\kappa$ for some $\kappa > 0$. Partition the interval $\|\hat{\theta} - \theta\| \leq ([1/\delta_T] + 1) \delta_T \varepsilon_T$, where $\delta_T$ is a decreasing sequence, as a union of a class $E$ of closed cubes $E_k$ with center $\theta_k$ such that for every $\hat{\theta} \in E_k$, $\|\hat{\theta} - \theta_k\| < \delta_T \varepsilon_T$. Therefore, for each dimension of the parameter space, we partition it with intervals of length $\delta_T \varepsilon_T$ so that there are $[1/\delta_T] + 1$ partitions. Given that the dimension of the parameter space is $\lambda$, such partitioning generates $([1/\delta_T] + 1)^\lambda$ number of cubes.

Define $s_t(\hat{\theta}) = -\mathbb{X}_{1,t}[\psi_T(u_t(\tau_1) - \hat{\Upsilon}_t(\hat{\theta})) - \psi_T(u_t(\tau_1))]$. The Lemma holds if

$$\sup_{\|\hat{\theta} - \theta\| \leq \varepsilon_T} \left\| \frac{1}{T} \sum_{t=1}^{T} (s_t(\hat{\theta}) - E[s_t(\hat{\theta})]) \right\| = o_p(\lambda/\sqrt{T}).$$

Consider

$$\sup_{\|\hat{\theta} - \theta\| \leq \varepsilon_T} \left\| \frac{1}{T} \sum_{t=1}^{T} (s_t(\hat{\theta}) - E[s_t(\hat{\theta})]) \right\| \leq \max_{E_k \in E} \left\| \frac{1}{T} \sum_{t=1}^{T} (s_t(\theta_k) - E[s_t(\theta_k)]) \right\|$$

$$= \max_{E_k \in E} \left( \frac{1}{T} \sum_{t=1}^{T} (s_t(\hat{\theta}) - s_t(\theta_k)) - E[s_t(\hat{\theta}) - s_t(\theta_k)] \right) \right\|.$$  \hspace{1cm} (26)

Focusing on the first term in (26), for some sequence $\gamma_T$, consider

$$P(\max_{E_k \in E} \left\| \frac{1}{T} \sum_{t=1}^{T} (s_t(\theta_k) - E[s_t(\theta_k)]) \right\| > \gamma_T) \leq \sum_{j=1}^{\lambda} \sum_{E_k \in E} P(\left\| \frac{1}{T} \sum_{t=1}^{T} (s_t^{(j)}(\theta_k) - E[s_t^{(j)}(\theta_k)]) \right\| > \gamma_T),$$

where we have used the fact that $s_t$ is a $\lambda$-vector. Now, by definition, $\{s_t^{(j)}, \mathcal{F}_{1,t}\}$ is an adapted stochastic sequence. Since $E[E[s_t^{(j)}(\theta_k)|\mathcal{F}_{1,t-1}]] = E[s_t^{(j)}(\theta_k)|\mathcal{F}_{1,t-1}]$ by smoothing, $s_t^{(j)}(\theta_k) - E[s_t^{(j)}(\theta_k)]$
In order for the probability expression to converge to zero, we require that
\[ \kappa > \zeta \]

may choose some \( \psi \) to \( \zeta < \) to \( P(\delta \text{ is a martingale difference sequence.}) \)

The first inequality follows from the Cauchy-Schwarz inequality, the second inequality follows from A6 and the third inequality follows from the fact that \( \psi^A \leq \psi_\tau \). We apply the Hoeffding inequality for martingales in Theorem 1 of Lee and Su (2002), that is,

\[
P\left( \left| \sum_{t=1}^{T} \left( s_t^{(j)}(\theta_k) - E[s_t^{(j)}(\theta_k)] \right) \right| > CT \right) \leq \exp \left( \frac{-C^2T^2}{2 \sum_{t=1}^{T} p_{1,t} + \frac{2CT^3}{3}} \right).
\]

Let \( \delta_T = \mathcal{O}(T^{-\delta}) \), where \( \delta \) is some positive constant, and this also implies that \( (1/\delta_T + 1)^\lambda = \mathcal{O}(T^{\lambda\delta}) \). Applying law of large numbers, we have \( T(\sum_{t=1}^{T} p_{1,t}) = T\mathcal{O}_p(\lambda^{-s} + (\lambda/T^{1/2}) \log T^\kappa) \leq \mathcal{O}_p(T^{\zeta/2 + 1/2} \log T^\kappa) \) where the inequality follows from setting \( s = (1 - 2\zeta)/2\zeta \) and \( \lambda = \mathcal{O}(T^\zeta) \) as stated in A3. Let \( C = \sqrt{2}T^{-b} \log T^\kappa \) be the constant and choose \( b = (3 - 2\zeta)/4 \). Consequently, \( P(T^{-1}| \sum_{t=1}^{T} (s_t^{(j)}(\theta_k) - E[s_t^{(j)}(\theta_k)]) > \sqrt{2}T^{-b} \log T^\kappa) \leq \exp (- \log T^\kappa) = T^{-\kappa} \). We have

\[
P\left( \max_{E_k \in E} \left\| \frac{1}{T} \sum_{t=1}^{T} \left( s_t(\theta_k) - E[s_t(\theta_k)] \right) \right\| > \sqrt{2}T^{-b} \log T^\kappa \right)
\]

\[
\leq \lambda \mathcal{O}(T^{\lambda\delta}) P \left( \left| \frac{1}{T} \sum_{t=1}^{T} \left( s_t^{(j)}(\theta_k) - E[s_t^{(j)}(\theta_k)] \right) \right| > \sqrt{2}T^{-b} \log T^\kappa \right)
\]

\[
\leq \mathcal{O}(T^{\zeta + \lambda\delta - \kappa}).
\]

In order for the probability expression to converge to zero, we require that \( \kappa > \zeta + \lambda\delta \). Hence, we may choose some \( \kappa \) that satisfies \( \kappa = \mathcal{O}(\lambda) \) so that \( T^{-b} \log T^\kappa = T^{-b} \kappa \log T = \mathcal{O}(T^{-b+\zeta}) \). For \( \max_{E_k \in E} \left\| \frac{1}{T} \sum_{t=1}^{T} (s_t(\theta_k) - E[s_t(\theta_k)]) \right\| = o_p(1) \), we require that \( b > \zeta \) holds, which corresponds to \( \zeta < 1/2 \). Next, note that \( \lambda/\sqrt{T} = \mathcal{O}(T^{\zeta - 1/2}) \). In order for \( T^{-b} \log T^\kappa = o(\lambda/\sqrt{T}) \) so that \( \max_{E_k \in E} \left\| \frac{1}{T} \sum_{t=1}^{T} (s_t(\theta_k) - E[s_t(\theta_k)]) \right\| = o_p(\lambda/\sqrt{T}) \), we check if \( T^{-b} \log T^\kappa = o(T^{\zeta - 1/2}) \), which
Proof of Proposition 2

Let \( r_{k,t} = (\theta_k - \theta)^T \hat{X}_{1,t} + \theta' (\hat{X}_{1,t} - X_{1,t}) - \Gamma' i_3 \). Focusing on the second term in (26), express

\[
\max_{E_k \in \mathcal{E}} \sup_{\|\theta - \theta_k\| \leq \delta_T \epsilon_T, \tilde{\theta} \in E_k} \left\| \frac{1}{T} \sum_{t=1}^{T} (s_t(\tilde{\theta}) - s_t(\theta_k) - E[s_t(\tilde{\theta})] - s_t(\theta_k)) \right\|
\leq \max_{E_k \in \mathcal{E}} \left\| \frac{1}{T} \sum_{t=1}^{T} X_{1,t} \left( \psi_t(u_t(\tau_1) - r_{k,t} - \delta_T \epsilon_T \|\hat{X}_{1,t}\|) - \psi_t(u_t(\tau_1) - r_{k,t} + \delta_T \epsilon_T \|\hat{X}_{1,t}\|) \right) \right\|
\leq \left\| u_t(\tau_1) \leq r_{k,t} + \delta_T \epsilon_T \|\hat{X}_{1,t}\| \right\|
\]

Observe that \( \psi_t(u_t(\tau_1) - r_{k,t} - \delta_T \epsilon_T \|\hat{X}_{1,t}\|) - \psi_t(u_t(\tau_1) - r_{k,t} + \delta_T \epsilon_T \|\hat{X}_{1,t}\|) = \mathbb{I}(r_{k,t} - \delta_T \epsilon_T \|\hat{X}_{1,t}\| \leq u_t(\tau_1) \leq r_{k,t} + \delta_T \epsilon_T \|\hat{X}_{1,t}\|) \). As before, considering the \( j \) element in (27), we check that

\[
E[(X_{1,t}^{(j)} \mathbb{I}(r_{k,t} - \delta_T \epsilon_T \|\hat{X}_{1,t}\| \leq u_t(\tau_1) \leq r_{k,t} + \delta_T \epsilon_T \|\hat{X}_{1,t}\|))] \leq \mathcal{O}(|X_{1,t}^{(j)}|^2 C_f \max \delta_T \epsilon_T \|\hat{X}_{1,t}\|(1 + o_p(1))
\leq 2\lambda^{1/2} |X_{1,t}^{(j)}|^2 C_f \max \delta_T \epsilon_T \max_k |X_{1,t}^{(k)}| |(1 + o_p(1))
\leq 2\lambda^{1/2} C_f \max \delta_T \epsilon_T \max_k |X_{1,t}^{(k)}|^3 (1 + o_p(1)) := p_{2,t},
\]

where we can show that \( \sum_{t=1}^{T} p_{2,t} = \mathcal{O}(T^{2\zeta - \delta + 1/2} \log T^\kappa) \). Applying the probability inequality again, we can show that \( P(|T^{-1} \sum_{t=1}^{T} (s_t(\tilde{\theta}) - s_t(\theta_k) - E[s_t(\tilde{\theta})] - s_t(\theta_k))| > \sqrt{2} T^{-d} \log T^\kappa) < T^{-\kappa}, \) for \( d = (3 - 4\zeta + 2\delta)/4 \). For \( d > b \), we choose \( \delta > \zeta \) so that the second term in (26) is also \( o_p(\lambda/\sqrt{T}) \).

\[\square\]

Proof of Proposition 2: Define \( \hat{Y}_t(\tilde{\theta}) = (\tilde{\theta} - \theta)^T \hat{X}_{1,t} + \theta(\hat{X}_{1,t} - X_{1,t}) - \Gamma' i_3 \). Without loss of generality, assume that \( (\tilde{\theta} - \theta)^T \hat{X}_{1,t} + \theta(\hat{X}_{1,t} - X_{1,t}) \) Also, recall that

\[
Y_{1,t} = \theta^T X_{1,t} + \Gamma' i_3 + u_t(\tau_1)
\]

\[
= \theta^T \hat{X}_{1,t} + \Gamma' i_3 + \theta(\hat{X}_{1,t} - \hat{X}_{1,t}) + u_t(\tau_1),
\]

From now on, let \( \theta \) be the population parameter to simplify the notation. Denote the first order
condition as
\[ \hat{\mathbb{W}}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\theta})) = T^{-1} \sum_{t=1}^{T} \hat{X}_{1,t} \psi_{\tau_1} \left( u_t(\tau_1) - \hat{\Upsilon}_t(\hat{\theta}) \right), \]
where \( \hat{\mathbb{W}}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\theta})) = 0 \) except for a finite number of points since \( \hat{\theta} \) is the minimizer. In addition, denote
\[ \mathbb{W}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\theta})) = T^{-1} \sum_{t=1}^{T} X_{1,t} \psi_{\tau_1} \left( u_t(\tau_1) - \hat{\Upsilon}_t(\hat{\theta}) \right). \]
That is, the difference between \( \hat{\mathbb{W}}_{T,\tau_1} \) and \( \mathbb{W}_{T,\tau_1} \) is that the former multiplies \( \psi_{\tau_1} \) with \( \hat{X}_{1,t} \) while the latter with \( X_{1,t} \). Expand \( E[\hat{\mathbb{W}}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\theta}))] \) around \( \hat{\Upsilon}_t(\hat{\theta}) = 0 \):

\[
E[\hat{\mathbb{W}}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\theta}))] = T^{-1} \sum_{t=1}^{T} E[\hat{X}_{1,t} E[\psi_{\tau_1} (u_t(\tau_1) - \hat{\Upsilon}_t(\hat{\theta})) | \hat{X}_{1,t}]] \]
\[
= T^{-1} \sum_{t=1}^{T} E[\hat{X}_{1,t} (F(\hat{\Upsilon}_t(\hat{\theta}))) - F(0))] \]
\[
= T^{-1} \sum_{t=1}^{T} E[X_{1,t} (F(\hat{\Upsilon}_t(\hat{\theta}))) - F(0))] + T^{-1} \sum_{t=1}^{T} E[(\hat{X}_{1,t} - X_{1,t})(F(\hat{\Upsilon}_t(\hat{\theta}))) - F(0))]. \tag{28}
\]
Now, \( T^{-1} \sum_{t=1}^{T} E[(\hat{X}_{1,t} - X_{1,t})(F(\hat{\Upsilon}_t(\hat{\theta}))) - F(0))] = o(T^{-1/2}). \) For the first term in (28), we apply the Mean Value Theorem, thus obtaining

\[
E[\hat{\mathbb{W}}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\theta}))] = T^{-1} \sum_{t=1}^{T} \left( E[X_{1,t} (F(-\Gamma t) - F(0))] + E[X_{1,t} f(\eta \hat{\Upsilon}_t(\hat{\theta})) X_{1,t} | \hat{\Upsilon}_t(\hat{\theta}) - \theta] + E[X_{1,t} f(\eta \hat{\Upsilon}_t(\hat{\theta})) \theta | \hat{X}_{1,t} - X_{1,t}] \right) \]
\[
+ o(T^{-1/2}) \tag{29}
\]
where \( 0 < \eta < 1. \) By A7, since \( \hat{X}_{1,t} \) is a smooth function of \( \hat{w}_t \) which converges almost surely to \( w_t, \) \( \hat{X}_{1,t} \) also converges almost surely to \( X_{1,t} \). In addition, recall that the \( i \)th diagonal element of \( F(\theta) \) is \( f_i(\eta \hat{\Upsilon}_t(\hat{\theta})) \) while \( \hat{\theta} \) is consistent by Proposition 1. Consider \( \hat{F}(\hat{\theta}) = \hat{F}(\theta) - F(\theta) + F(\theta). \) Then, applying the Slutsky Theorem, we have \( \hat{F}(\hat{\theta}) - F(\theta) \overset{p}{\to} 0 \) by A7 and \( F(\hat{\theta}) \overset{p}{\to} F(\theta) \) by Proposition 1. Hence, \( \hat{M}_{1,T}(\theta) = E[X_{1,T} F(\theta) X_{1}/T]. \) To simplify the notation further, let \( \tilde{M}_{1,T} \) and \( F \) correspond to the values where the population parameter \( \theta \) is in the argument. We may
then rewrite the vector counterpart of (29) as

\[
E[\hat{W}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\theta}))]
= \hat{M}_{1,T}(\hat{\theta} - \theta) - T^{-1} \sum_{t=1}^{T} E[f_t(\eta \Upsilon_t(\theta))X_{1,t}]\Gamma_t'\iota_3 + T^{-1} \sum_{t=1}^{T} E[f_t(\eta \Upsilon_t(\theta))X_{1,t}\theta'](\hat{X}_{1,t} - X_{1,t})
+ o(T^{-1/2}).
\quad (30)
\]

By rearranging (30), we obtain

\[
\hat{\theta} - \theta
= \hat{M}_{1,T}^{-1}E[\hat{W}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\theta}))] + \hat{M}_{1,T}^{-1}T^{-1} \sum_{t=1}^{T} E[f_t(\eta \Upsilon_t(\theta))X_{1,t}]\Gamma_t'\iota_3
- \hat{M}_{1,T}^{-1}T^{-1} \sum_{t=1}^{T} E[f_t(\eta \Upsilon_t(\theta))X_{1,t}\theta'](\hat{X}_{1,t} - X_{1,t}) + o(T^{-1/2})

= \hat{M}_{1,T}^{-1}W_{1,T}(0) + \hat{M}_{1,T}^{-1}T^{-1} \sum_{t=1}^{T} E[f_t(\eta \Upsilon_t(\theta))X_{1,t}]\Gamma_t'\iota_3 - \hat{M}_{1,T}^{-1}T^{-1} \sum_{t=1}^{T} E[f_t(\eta \Upsilon_t(\theta))X_{1,t}\theta'](\hat{X}_{1,t} - X_{1,t})

+ \hat{M}_{1,T}^{-1}(\hat{W}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\theta})) - \hat{W}_{T,\tau_1}(\hat{\Upsilon}_t(\theta))) + \hat{M}_{1,T}^{-1}(\hat{W}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\theta})) - \hat{W}_{T,\tau_1}(\hat{\Upsilon}_t(\theta)))

+ o(T^{-1/2}).
\quad (31)
\]

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

\[
\sup_{\|\hat{\theta} - \theta\| \leq \varepsilon_T} \|W_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\theta})) - W_{T,\tau_1}(\hat{\Upsilon}_t(\theta)) - E[\hat{W}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\theta})) - \hat{W}_{T,\tau_1}(\hat{\Upsilon}_t(\theta))]| = o_p(\lambda/\sqrt{T}),
\quad (32)
\]

where we have used the fact that \(E[\hat{W}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\theta}))] = 0\). To establish the rate for (B), observe that

\[
\hat{W}_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\theta})) - W_{T,\tau_1}(\hat{\Upsilon}_t(\hat{\theta})) \leq \frac{1}{T} \sum_{t=1}^{T} (\hat{X}_{1,t} - X_{1,t})\psi_{\tau_1}(u_t(\tau_1) - \hat{\Upsilon}_t(\hat{\theta}))
\leq \max_t \|\hat{X}_{1,t} - X_{1,t}\| \frac{1}{T} \sum_{t=1}^{T} \psi_{\tau_1}(u_t(\tau_1) - \hat{\Upsilon}_t(\hat{\theta})),
\]

39
where \( \max_t \| \hat{X}_{1,t} - X_{1,t} \| = o_p(1) \) by A7. To apply Chebyshev inequality, check that

\[
\Var \left( \frac{1}{T} \sum_{t=1}^{T} \psi_{\tau_1}(u_t(\tau_1) - \hat{Y}_t(\hat{\theta})) \right) \leq T^{-2} \sum_{t=1}^{T} \left[ f(\eta \hat{Y}_t(\hat{\theta}))((\hat{\theta} - \theta)'X_{1,t} + \theta'(\hat{X}_{1,t} - X_{1,t}) - \Gamma_t i_3) \right] \\
\leq T^{-2} \sum_{t=1}^{T} C_f^{\max} (\|\hat{\theta} - \theta\|\|X_{1,t}\| + \|\theta\|\|\hat{X}_{1,t} - X_{1,t}\| + K\lambda^{-s}) \\
= o_p(T^{-1}),
\]

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 \( \hat{W}_{T,\tau_1}(\hat{Y}_t(\hat{\theta})) - \hat{W}_{T,\tau_1}(\hat{Y}_t(\hat{\theta})) = o_p(T^{-1}) \).

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

\[
\hat{\theta} - \theta = - \bar{M}_{1,T}^{-1} \sum_{t=1}^{T} X_{1,t} \psi_{\tau_1}(u_t(\tau_1)) + \bar{M}_{1,T}^{-1} \sum_{t=1}^{T} E[f_t(\eta \hat{Y}_t(\hat{\theta}))X_{1,t}] \Gamma_t i_3 \\
- \bar{M}_{1,T}^{-1} \sum_{t=1}^{T} E[f_t(\eta \hat{Y}_t(\hat{\theta}))X_{1,t}'](\hat{X}_{1,t} - X_{1,t}) + o_p(\lambda/\sqrt{T}). \tag{33}
\]

Recall that \( \Gamma = [\Gamma_0 \; \Gamma_1 \; \Gamma_2] \), we have \( \bar{M}_{1,T}^{-1} \sum_{t=1}^{T} E[f_t(\eta \hat{Y}_t(\hat{\theta}))X_{1,t}] \Gamma_0, t = \bar{M}_{1,T}^{-1} E[X_{1}'F] \Gamma_0 \) using A3. Using the fact that \( \|\Gamma_0\| = (\Gamma_0^T \Gamma_0)^{1/2} \leq K\lambda^{1/2-s} \), consider \( \|\bar{M}_{1,T}^{-1} E[X_{1}'F] \Gamma_0\| \leq \|\bar{M}_{1,T}^{-1} E[X_{1}'F]\| \|\Gamma_0\| \). In addition, since Jensen’s inequality implies \( \|\bar{M}_{1,T}^{-1} E[X_{1}'F]\|^2 \leq E[\|\bar{M}_{1,T}^{-1} X_{1}'F\|^2] \), consider

\[
\|\bar{M}_{1,T}^{-1} E[X_{1}'F]\|^2 = T^{-2} tr(FX_{1} M_{1,T}^{-1} \bar{M}_{1,T}^{-1} X_{1}'F) \\
\leq C_f^{\max} K_{\min}(M_{1,T})^{-2} T^{-1} tr(X_{1}'F X_{1}/T) \\
\leq C_f^{\max} K_{\min}(M_{1,T})^{-2} T^{-1} K_{\max}(M_{1,T}) tr(I_\lambda) \\
= o_p(\lambda/T),
\]

which follows from \( K_{\max}(M_{1,T}) = O_p(1) \). Therefore, \( \|\bar{M}_{1,T}^{-1} \sum_{t=1}^{T} E[f_t(\eta \hat{Y}_t(\hat{\theta}))X_{1,t}] \Gamma_t, t\| = O(\sqrt{\lambda/T})O(\lambda^{1/2-2}) \). Repeating with \( \Gamma_1 \) and \( \Gamma_2 \), we have \( \|\bar{M}_{1,T}^{-1} \sum_{t=1}^{T} E[f_t(\eta \hat{Y}_t(\hat{\theta}))X_{1,t}] \Gamma_t i_3\| = O(\sqrt{\lambda/T})O(\lambda^{1/2-2}) \). Observe that \( O(\lambda^{1/2-s}) = O(T^{(1/2-s)}) \) and \( (1/2 - s)\zeta < \zeta \) since \( \zeta < 1/2 \). Therefore, \( O(T^{(1/2-s)}) = o(T^\zeta) \), which in turn implies that \( O(\lambda^{1/2-s}) = o(\lambda) \). Hence,
\( O(\sqrt{T}/T)O(\lambda^{s-1/2}) = o(\lambda/\sqrt{T}) \).

Next, consider the fact that \( M_{1,T}^{-1}T^{-1} \sum_{t=1}^{T} X_{1,t} \psi_{\tau_1}(u_t(\tau_1)) \leq M_{1,T}^{-1}T^{-1} \sum_{t=1}^{T} X_{1,t} = M_{1,T}^{-1}T^{-1} \bar{X}_{1,i\lambda} \).

Since
\[
\| M_{1,T}^{-1}T^{-1} \bar{X}_{1,i\lambda} \| = T^{-2} \lambda^{\min} (M_{1,T})^{-2} \| \bar{X}_{1,i\lambda} \| \leq T^{-2} \lambda^{\min} (M_{1,T})^{-2} \| i_\lambda \| \| X_{1,i\lambda} \| \leq T^{-2} \lambda^{\min} (M_{1,T})^{-2} \lambda K_{\max} (D_{1,T}) \| r(I_\lambda) \| = \| o_p(\lambda^2/T) \|,
\]

this implies that \(\| M_{1,T}^{-1}T^{-1} \sum_{t=1}^{T} X_{1,t} \psi_{\tau_1}(u_t(\tau_1)) \| = O_p(\lambda/\sqrt{T}) \).

Finally, consider \(\| M_{1,T}^{-1}T^{-1} \sum_{t=1}^{T} E[f_i(q_t(\theta))X_{1,t}](\bar{X}_{1,i\lambda} - X_{1,i\lambda}) \| \). Now, \(\| M_{1,T}^{-1}T^{-1}E[X_{1,i\lambda}'] \| = O(\sqrt{T}/T) \) and \(\| \bar{X}_{1,i\lambda} - X_{1,i\lambda} \| = O_p(\sqrt{T}/T) \). Together imply that \(\| M_{1,T}^{-1}T^{-1} \sum_{t=1}^{T} E[f_i(q_t(\theta))X_{1,t}](\bar{X}_{1,i\lambda} - X_{1,i\lambda}) \| = O_p(\lambda^2/T) \). This result, combining with the above, establishes the proposition after applying the triangular inequality. \( \square \)

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

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

\[
Var(c'X_{1,t} \psi_{\tau_1}(Y_t - \theta'X_{1,t})) = E[c'X_{1,t} \psi_{\tau_1}(Y_t - \theta'X_{1,t})] - (E[c'X_{1,t} \psi_{\tau_1}(Y_t - \theta'X_{1,t})])^2 \geq E[c'X_{1,t}^2]E[\psi_{\tau_1}(Y_t - \theta'X_{1,t})^2|X_{1,t}] - E[c'X_{1,t}^2|F_t(\Gamma_{i3}) - \tau_1^2] = E[c'X_{1,t}^2 F_t(\Gamma_{i3})(1 - F_t(\Gamma_{i3}))],
\]

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 \( E[c'X_{1,t}^2] = O(\lambda^2) \). In addition,
since $E|c'X_{1,t}|^2 > 0$, there is a positive constant $L$ such that $E|c'X_{1,t}|^2 > L\lambda^2$. Hence, $\Psi_T^2 \geq \min_t TE[c'X_{1,t}|2F_t(\Gamma_t'\iota_3)(1 - F_t(\Gamma_t'\iota_3))] \geq LT\lambda^2$, so that

$$\sum_{t=1}^{T} E \left[ \frac{|c'X_{1,t}\psi_1(Y_t - \theta'X_{1,t})|^2}{\Psi_T^2} \mathbb{I} \left( \left| \frac{c'X_{1,t}\psi_1(Y_t - \theta'X_{1,t})}{\Psi_T} \right| > \epsilon \right) \right] \leq \sum_{t=1}^{T} E \left[ \frac{|c'X_{1,t}|^{2+\delta}}{L^{1+\delta}T^{1+\delta}\lambda^{2+\delta}\epsilon^{1+\delta}} \right] \leq T \frac{\lambda^{2+\delta}\max_j c_j^{2+\delta}\Delta}{L^{1+\delta}T^{1+\delta}\lambda^{2+\delta}\epsilon^{1+\delta}} \to 0$$

since $\delta > 0$. Therefore, $T^{-1/2} \sum_{t=1}^{T} X_{1,t}\psi_1(Y_t - \theta'X_{1,t})$ is asymptotically normal by the Lindeberg-Feller Central Limit Theorem and the Crámer-Wold device. □

**Proof of Proposition 3:** This follows from Lemma 3, A5′ and A8. □
Figure 1: Estimated $\hat{\alpha}_2(\tau_2, \tau_1)$ from Monte Carlo simulation based on 500 observations.

This figure plots the estimates of $\hat{\alpha}_2$ using the regression function in (17). Panel A shows the true parameter value $\alpha_2 + \delta(\lambda e^{F_x^{-1}(\tau_2)} + F^{-1}_u(\tau_1))$ while Panels B, C, and D plot $\hat{\alpha}_2$ estimated from the linear, quadratic and cubic expansion models.
Figure 2: Estimated parameters of the output process equation.

This figure shows the estimated parameters in the output process equation using either M1 or M2 money supply as the monetary instrument. \textit{Tau of monetary shock} and \textit{Tau of output} index the quantile of the money supply shock and output growth respectively.

A. Intercept, M1 Money Supply
B. Intercept, M2 Money Supply
C. $y_{t-1}$, M1 Money Supply
D. $y_{t-1}$, M2 Money Supply
E. $\Delta r_{t-1}$, M1 Money Supply
F. $\Delta r_{t-1}$, M2 Money Supply
Figure 3: **Section plots of the intercept term.**

This figure plots the section of the intercept fixing output growth at the 10th, 50th or 90th percentile. Panels A and B plot the intercept sections based on M1 and M2 money supply as the monetary instrument. The horizontal line plots the median of the money supply shock. The dash lines represent the two standard deviation bands.

A. M1 Money Supply

B. M2 Money Supply
Figure 4: Changes in the intercept term for restrictive versus expansive monetary policy.

This figure plots the change in the intercept term of the output process equation when the monetary policy is restrictive or expansive relative to the median position. Panels A and B correspond to using M1 and M2 money supply as the monetary instrument. The solid line plots the change in the intercept when the money supply shock increases from the median to 90th percentile, i.e. $\alpha_I(0.9, \tau) - \alpha_I(0.5, \tau)$ for some $\tau$ quantile of output growth. The dotted line plots the change when the money supply shock declines from the median to the 10th percentile, i.e. $\alpha_I(0.5, \tau) - \alpha_I(0.1, \tau)$.

A. M1 Money Supply

B. M2 Money Supply
Figure 5: **Section plots of the slope coefficient on y_{t-1}**.

This figure plots the section of the slope coefficient on $y_{t-1}$ fixing output growth at the 10th, 50th or 90th percentile. Panels A and B plot the slope sections based on M1 and M2 money supply as the monetary instrument. The horizontal line plots the median of the money supply shock. The dash lines represent the two standard deviation bands.

A. M1 Money Supply

B. M2 Money Supply
Figure 6: **Section plots of the slope coefficient on $\Delta r^t_{t-1}$**.

This figure plots the section of the slope coefficient on $\Delta r^t_{t-1}$ fixing output growth at the 10th, 50th or 90th percentile. Panels A and B plot the slope sections based on M1 and M2 money supply as the monetary instrument. The horizontal line plots the median of the money supply shock. The dash lines represent the two standard deviation bands.

A. M1 Money Supply

B. M2 Money Supply
Table 1: Monte Carlo simulation based on 500 Observations

This table estimates $\hat{\alpha}_2(\tau_2, \tau_1)$ for $\tau_1 = \tau_2 = \tau$ based on (17). $I$ denotes the degree of polynomial in the power series expansion. The estimation is repeated 1000 times using 500 generated observations.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I = 1$</td>
<td>2.3502</td>
<td>8.3538</td>
<td>12.8233</td>
<td>16.9900</td>
<td>20.7534</td>
<td>24.9784</td>
<td>28.7811</td>
<td>34.1802</td>
<td>41.6116</td>
</tr>
<tr>
<td>bias</td>
<td>-3.1453</td>
<td>-1.2858</td>
<td>-0.0950</td>
<td>1.0414</td>
<td>1.7534</td>
<td>2.6860</td>
<td>2.6623</td>
<td>3.1242</td>
<td>2.7345</td>
</tr>
<tr>
<td>average rmse</td>
<td>5.7477</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.2511</td>
<td>-0.3519</td>
<td>-0.5866</td>
<td>-0.5530</td>
<td>-0.6100</td>
<td>-0.4649</td>
<td>0.0468</td>
<td>0.3622</td>
<td>0.6780</td>
</tr>
<tr>
<td>average rmse</td>
<td>5.2772</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.7019</td>
<td>0.4760</td>
<td>0.0458</td>
<td>-0.0353</td>
<td>-0.1472</td>
<td>-0.3213</td>
<td>-0.3746</td>
<td>0.1351</td>
<td>-0.7686</td>
</tr>
<tr>
<td>average rmse</td>
<td>5.4664</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.2879</td>
<td>0.2939</td>
<td>0.1875</td>
<td>0.0603</td>
<td>0.1967</td>
<td>0.4091</td>
<td>0.0598</td>
<td>-0.1545</td>
<td>-0.6969</td>
</tr>
<tr>
<td>average rmse</td>
<td>5.7119</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.4458</td>
<td>0.152</td>
<td>0.0436</td>
<td>0.1967</td>
<td>0.1063</td>
<td>0.0491</td>
<td>-0.1654</td>
<td>-0.2845</td>
<td>-0.9699</td>
</tr>
<tr>
<td>average rmse</td>
<td>5.4664</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.4112</td>
<td>0.2538</td>
<td>0.1534</td>
<td>0.2116</td>
<td>-0.1387</td>
<td>-0.1904</td>
<td>-0.1484</td>
<td>-0.4641</td>
<td>-0.4641</td>
</tr>
<tr>
<td>average rmse</td>
<td>5.8065</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.8014</td>
<td>0.1611</td>
<td>0.4281</td>
<td>0.1216</td>
<td>0.0348</td>
<td>0.0158</td>
<td>-0.0158</td>
<td>-0.3453</td>
<td>-0.4737</td>
</tr>
<tr>
<td>average rmse</td>
<td>5.9791</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.4459</td>
<td>0.1329</td>
<td>0.0974</td>
<td>0.1782</td>
<td>-0.1804</td>
<td>-0.0499</td>
<td>-0.2751</td>
<td>0.1945</td>
<td>-1.1292</td>
</tr>
<tr>
<td>rmse</td>
<td>7.9280</td>
<td>5.5988</td>
<td>4.7076</td>
<td>4.6292</td>
<td>4.7691</td>
<td>5.1226</td>
<td>5.7284</td>
<td>7.0144</td>
<td>10.8679</td>
</tr>
<tr>
<td>average rmse</td>
<td>6.2629</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.5269</td>
<td>0.6924</td>
<td>0.0595</td>
<td>0.2799</td>
<td>0.0634</td>
<td>-0.3766</td>
<td>0.0367</td>
<td>-0.2283</td>
<td>-0.3865</td>
</tr>
<tr>
<td>rmse</td>
<td>7.9815</td>
<td>5.7101</td>
<td>4.8820</td>
<td>4.6013</td>
<td>4.9457</td>
<td>5.0767</td>
<td>5.9700</td>
<td>7.1807</td>
<td>10.9551</td>
</tr>
<tr>
<td>average rmse</td>
<td>6.3670</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.8439</td>
<td>0.3643</td>
<td>-0.1016</td>
<td>-0.0950</td>
<td>0.1090</td>
<td>-0.0522</td>
<td>-0.0713</td>
<td>-0.2813</td>
<td>-1.1286</td>
</tr>
<tr>
<td>average rmse</td>
<td>6.5668</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Monte Carlo simulation based on 1000 Observations

This table estimates $\hat{\alpha}_2(\tau_2, \tau_1)$ for $\tau_1 = \tau_2 = \tau$ based on (17). $I$ denotes the degree of polynomial in the power series expansion. The estimation is repeated 1000 times using 1000 generated observations.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>bias</td>
<td>-3.2159</td>
<td>-1.5256</td>
<td>0.0251</td>
<td>0.9144</td>
<td>1.8524</td>
<td>2.3939</td>
<td>2.8442</td>
<td>2.7857</td>
<td>2.7776</td>
</tr>
<tr>
<td>rmse</td>
<td>5.5689</td>
<td>3.5463</td>
<td>2.7144</td>
<td>2.9665</td>
<td>3.4330</td>
<td>3.9166</td>
<td>4.6164</td>
<td>5.2503</td>
<td>7.0282</td>
</tr>
<tr>
<td>average rmse</td>
<td>4.3379</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>-0.0982</td>
<td>-0.4784</td>
<td>-0.5960</td>
<td>-0.8152</td>
<td>-0.5471</td>
<td>-0.2732</td>
<td>-0.0525</td>
<td>0.2264</td>
<td>1.2196</td>
</tr>
<tr>
<td>average rmse</td>
<td>3.7631</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.6248</td>
<td>0.4637</td>
<td>0.0286</td>
<td>0.0713</td>
<td>-0.0605</td>
<td>-0.2027</td>
<td>-0.0035</td>
<td>-0.0414</td>
<td>-1.402</td>
</tr>
<tr>
<td>average rmse</td>
<td>3.7287</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.4001</td>
<td>0.0708</td>
<td>-0.0170</td>
<td>0.1122</td>
<td>0.0066</td>
<td>-0.0301</td>
<td>0.2264</td>
<td>1.2196</td>
<td></td>
</tr>
<tr>
<td>average rmse</td>
<td>3.7631</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.1987</td>
<td>0.0434</td>
<td>0.1574</td>
<td>0.0929</td>
<td>0.0066</td>
<td>-0.0301</td>
<td>-0.0397</td>
<td>-0.1937</td>
<td>0.0800</td>
</tr>
<tr>
<td>average rmse</td>
<td>3.8661</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.4075</td>
<td>-0.0695</td>
<td>-0.0398</td>
<td>-0.1220</td>
<td>0.0346</td>
<td>0.0604</td>
<td>-0.0347</td>
<td>-0.1091</td>
<td>0.0554</td>
</tr>
<tr>
<td>average rmse</td>
<td>3.9316</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.4678</td>
<td>0.1035</td>
<td>0.1845</td>
<td>-0.1983</td>
<td>-0.1224</td>
<td>0.1456</td>
<td>-0.1226</td>
<td>-0.2840</td>
<td>-0.8161</td>
</tr>
<tr>
<td>average rmse</td>
<td>4.0495</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.1989</td>
<td>0.1693</td>
<td>-0.0830</td>
<td>0.0792</td>
<td>0.0783</td>
<td>-0.1472</td>
<td>0.1816</td>
<td>-0.1912</td>
<td>-0.2082</td>
</tr>
<tr>
<td>average rmse</td>
<td>4.2235</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.2490</td>
<td>0.2468</td>
<td>-0.1583</td>
<td>0.1689</td>
<td>0.0818</td>
<td>-0.1665</td>
<td>0.1503</td>
<td>-0.3159</td>
<td>-0.1161</td>
</tr>
<tr>
<td>average rmse</td>
<td>4.4266</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bias</td>
<td>0.3071</td>
<td>0.1024</td>
<td>0.0497</td>
<td>-0.0548</td>
<td>-0.0382</td>
<td>0.1213</td>
<td>-0.1814</td>
<td>-0.3339</td>
<td>-0.4264</td>
</tr>
<tr>
<td>average rmse</td>
<td>4.5190</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>