OPTIMAL MEDIAN UNBIASED ESTIMATION OF COEFFICIENTS ON HIGHLY PERSISTENT REGRESSORS

PIOTR ELIASZ

ABSTRACT. This paper derives an optimal estimator for the slope coefficient on highly persistent and predetermined regressors in an otherwise standard linear regression. Optimality pertains to the class of procedures that are median unbiased irrespectively of the degree of persistence. It holds for a wide class of monotone loss functions. The optimality statement generalizes to confidence sets. The estimator, which is based on inversion of the Jansson-Moreira (2004) statistic, dominates currently available alternatives in terms of expected square losses across the domain of near nonstationarity. In the empirical application we document encouraging performance of the proposed estimator for forecasting asset returns.

1. INTRODUCTION

This paper addresses the problem of optimal estimation of coefficients on highly persistent and predetermined regressors in an otherwise standard linear regression. This issue draws economists' attention for at least two reasons. First, it is well known that currently available procedures are likely to produce biased estimates in this setting. Second, this problem is of high empirical relevance because high persistence and potential endogeneity are widespread across economic time series. Numerous financial indexes as well as macroeconomic indicators, among them inflation, unemployment and gross domestic product, fall into this category.

Date: November 10, 2004. JOB MARKET PAPER. I thank Ben Bernanke, Andrea Gager, Jianqing Fan, P. Jeganathan, Ulrich Müller, Christopher Sims and particularly my advisor, Mark Watson, for helpful comments on earlier drafts of this paper. Financial support from the Fellowship of Woodrow Wilson Scholars is gratefully acknowledged.

Theoretical studies provide a variety of estimators that exhibit some form of efficiency, either in the standard setting of stationary regressors or when regressors are exactly integrated (for efficient estimators in cointegrating regressions see Johansen (1988), Phillips and Loretan (1991), Saikkonen (1991), Stock and Watson (1993)). What this paper concerns, however, is a framework that falls in between – estimation when regressors are not necessarily integrated but nonetheless too persistent to be usefully described as stationary. In the econometric literature such variables are often referred to as nearly integrated. In this setting we are aware of only one procedure with certain efficiency properties, the asymptotically centering estimator of Jeganathan (1997) (see also Cox and Llatas (1991)).[1] It is not entirely satisfactory, however, since it is subject to a similar lack of robustness as noted by Elliott (1998) in the context of cointegration methods, as we will discuss further below.

The objective of this paper is to develop an alternative approach, with analytically demonstrable optimality, that would be robust to the degree of persistence in a regressor. To this end we exploit a recent development of optimal testing procedures for the slope coefficient on nearly integrated regressors by Jansson and Moreira (2004) in conjunction with the theory of optimal median unbiased estimation in the presence of nuisance parameters of Pfanzagl (1979). The resulting estimator, which is based on inversion of the Jansson-Moreira statistic, is optimal in the class of procedures that are conditionally median unbiased, a restriction made precise later in the text. This is established under a very general specification of a loss function, spanning asymmetric, concave or bounded shapes, a particularly attractive feature given recent evidence on asymmetric losses in macroeconomic forecasting by Elliott, Komunjer, and Timmermann (2004). The optimality statement is obtained as a corollary of a stronger result that provides optimal confidence sets in regressions

[1]We may add that this framework presents no inherent difficulties from the perspective of Bayesian analysis. The AC estimator is in fact a limit of a sequence of Bayes estimators under proper priors.

with highly serially correlated regressors, an issue investigated by Stock and Watson (1996) and more recently Campbell and Yogo (2003).

Monte Carlo evaluation of the proposed estimator demonstrates that it dominates currently available procedures across the domain of near nonstationarity, with relative efficiency gains increasing with the degree of endogeneity. We also evaluate the robustness of the median unbiased estimator to the choice of a loss function and find that it retains desirable characteristics under asymmetric losses, in contrast to some of the alternative procedures.

Turning to empirical applicability we note that the property of median unbiasedness characterizes estimators that are as likely to underestimate as to overestimate the true value, a highly desirable characteristic when any systematic biases affect the estimation. It is particularly relevant in forecasting, a setting in which median unbiased estimation results in same frequency of positive and negative forecast errors.[2]

We illustrate an application of our methodology with an exercise in forecasting asset returns. Although the question of predictability of asset returns has received significant attention in the literature, little is known about relative merits of different estimation methods in forecasting. We demonstrate that the optimal median unbiased estimator performs well relative to the available statistical alternatives and document robustness of this result with respect to the choice of a loss function.

We begin with the description of the model in the next section. For clarity it takes a simple form of a predictive regression model. Optimality of the proposed median unbiased estimator is developed in Section 3 for finite samples and asymptotically equivalent results are presented in Section 4. Monte Carlo evidence is discussed in Section 5. Extensions to more general settings as well as feasible inference are discussed in Section 6 and the empirical application is described in Section 7. Proofs are collected in an appendix.

[2]In the class of linear predictions this holds under symmetrically distributed innovations and no uncertainty about the sign of the most recent realization of the regressor.

2. Preliminaries

Throughout the paper we employ a setting and notation used earlier by Jansson and Moreira (2004). We therefore consider a bivariate predictive regression model where the observed data $\{(y_t, x_t) : 1 \leq t \leq T\}$ is generated by

$$y_t = \alpha + \beta x_{t-1} + \varepsilon_t^y \tag{1}$$

$$x_t = \gamma x_{t-1} + \varepsilon_t^x \tag{2}$$

where

$A1 :\ x_0 = 0,$

$A2 :\ \varepsilon_t = \left(\varepsilon_t^y, \varepsilon_t^x\right)' \sim i.i.d.\mathcal{N}\left(0, \Sigma\right)$ and

$$\Sigma = \begin{bmatrix} \sigma_{yy} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{xx} \end{bmatrix}$$

is a known, positive definite matrix.

These assumptions are relaxed later in the text. In particular, we modify (2) to accommodate $x_0 \neq 0$ and replace the Gaussian assumption on the errors with moment restrictions.

Our choice of the predictive regression model (1), rather than a standard regression of the form

$$y_t = \alpha + \beta x_t + \varepsilon_t^y \tag{3}$$

is driven in part by expositional simplicity of the former, given results of Jansson and Moreira (2004) and earlier Jeganathan (1997) who have demonstrated their asymptotic equivalence in the nearly integrated regressors setting considered in this paper.[3] It is therefore important to stress that methods developed in this paper are

[3] The two models are asymptotically equivalent in a sense that their respective sufficient statistics have the same asymptotic representation.

asymptotically valid and applicable in the standard regression setting (3). In addition, this framework, which plays a prominent role in empirical finance, suits well a possible application of our procedure to forecasting.

We are concerned with estimators of β treating γ as a nuisance parameter. Although finite sample theory of the median unbiased estimation to be presented in the next section does not hinge on any further assumptions on the parameter γ, the motivation for the paper and the asymptotic theory all come from a local-to-unity parametrization of γ, that is $\gamma = 1 + c/T$, where c is a negative constant. This stipulates that the root of the process $\{x_t\}$ approaches unity, a setting in which even asymptotically it remains difficult to differentiate between a stationary and an integrated series. The local-to-unity parametrization has often been employed in the closely related problem of inference on autoregressive coefficient near one; see, among other papers, Cavanagh, Elliott, and Stock (1995), Elliott, Rothenberg, and Stock (1996) and Elliott and Stock (2001).

Before we proceed with the theory of optimal median unbiased estimation of β we first note that the only alternative estimator with demonstrable optimality that holds uniformly over the domain of c is the Gaussian maximum likelihood estimator (GMLE) of β. Following Jeganathan (1995) and earlier statistical texts we call it an asymptotically centering (AC) estimator, since it is defined as a centering of a local quadratic approximation to the likelihood function. It takes a simple form of the OLS estimator with the OLS estimate of γ plugged-in, as detailed in the following Lemma:

Lemma 1. *In the model* (1)-(2) *the AC estimator takes the form of the OLS estimator from regressing* $y_t - \sigma_{xx}^{-1}\sigma_{xy}\left(x_t - \hat{\gamma}x_{t-1}\right)$ *on a constant and* x_{t-1}*, where* $\hat{\gamma}$ *is the OLS estimator from regressing* x_t *on* x_{t-1}*.*

Cox and Llatas (1991) and Jeganathan (1997) applied this methodology to the problem closely related to ours and demonstrated that the AC estimator minimizes asymptotic variance in the class of $M-estimators$, defined in terms of the asymptotic score representation as specified in Jeganathan (1995, page 851). As evident from Lemma 1, however, the AC procedure is sensitive to the first-step error in the estimation of γ. Under the local-to-unity parametrization this translates to a similar lack of robustness as noted by Elliott (1998) in the context of efficient cointegrating regression estimators. Under the infeasible scenario in which we know the value of c, the GMLE(c) has been considered by Phillips (1991), among others, and is known to deliver the minimax risk.

In this paper we choose to search for an estimator in the class of median unbiased procedures, that is such that are equally likely to fall above or below the true parameter value. Median unbiasedness is a very desirable property. Possibly its most noteworthy characteristic is the robustness to the choice of a loss function. The optimal estimator we will derive in the next section will minimize risk for any loss function that is non-decreasing as we move away from the true value, a class that spans asymmetric or bounded shapes. We make the following definition:

Definition 1. *Let $\mathcal{L}_\beta$ denote the class of monotone (also referred to as quasiconvex) loss functions $L(\cdot,\beta):\mathbb{R}\to[0,\infty)$, where $\beta\in\mathbb{R}$, such that: (i) $L(\beta,\beta)=0$ and (ii) $L(\cdot,\beta)\leq\bar{L}$, where $\bar{L}\in\mathbb{R}^+$, is a convex set.*

On the theoretical front Pfanzagl (1985) demonstrated that if we rank estimators according to risk under monotone loss then only estimators with the same median bias are comparable under this criterion.[4] It naturally leads to considering median unbiased estimators, which he further shows are always admissible in this setting. Our attention, therefore, turns to procedures $m_T(y,x)$ that will

$$\min_{m_T} E^{Y,X}_{\beta,\gamma}\left(L\left(m_T\left(y,x\right),\beta\right)\right) \tag{4}$$

[4]To see this consider two loss functions of the form $L_1(b,\beta)=1-1[b\leq\beta]$ and $L_2(b,\beta)=1-1[b\geq\beta]$, where $1[\cdot]$ is the indicator function. Note that these are members of $\mathcal{L}_\beta$. Corresponding risks equal probabilities of the estimator falling on each side of the true parameter value, necessitating equal median bias.

subject to:

$$P_{\beta,\gamma}^{Y,X}\left\{m_T\left(y,x\right)\leq\beta\right\}=P_{\beta,\gamma}^{Y,X}\left\{m_T\left(y,x\right)\geq\beta\right\}=\frac{1}{2}\tag{5}$$

where $L\in\mathcal{L}_\beta$ and (Y,X) are the random variables described by the system (1)-(2). $P_{\beta,\gamma}^{Y,X}$ specifies the probability measure generated by (Y,X) over the class of Borel sets in the Euclidean space, endowed with a Lebesgue measure μ. Expectation is with respect to $P_{\beta,\gamma}^{Y,X}$. The side condition (5) specifies that the estimator must be median unbiased in the family $P_{\beta,\gamma}^{Y,X}$.

It turns out, unfortunately, that there is no optimal member in the class of estimators restricted only by (5). This observation (which is implicit in the proof to Theorem 1, see the next section) is related to the work of Shaffer (1991), who shows in the setting of standard linear regression with random regressors that the Gauss-Markov theorem does not necessarily apply in the class of unconditionally mean unbiased estimators, whereas it always holds in the restricted class of conditionally mean unbiased procedures. In our quest for an optimal procedure we proceed along similar lines and restrict the class of estimators to those which are median unbiased conditionally on statistics that summarize variability of a regressor. Since distribution of such statistics is, for any fixed value of γ, independent of the parameter of interest β, they are the specific ancillaries.

There is an ongoing debate in the statistical literature about the validity and appropriateness of inference conditional on ancillaries. Many advocated conditioning on ancillaries, which appears to had been widely accepted until sound conditional procedures were shown to be unconditionally inadmissible by Brown (1990). (This become known as Brown ancillarity paradox, see Brown's paper and discussions in a special issue of the Annals of Statistics, 1990). As many discussants of Brown argue (see particularly Berger), this understanding should not divert our attention from conditional inference since, for the analysis of the sample at hand, we would often not like good performance of our estimator to come from the distribution of the ancillary at the expense of possibly terrible conditional characteristics. Given that conditioning provides means to eliminate nuisance parameters and enough simplicity to derive an optimal estimator we proceed in the conditional framework. In the next section we work out the finite sample solution to this problem.

3. Finite Sample Theory

We start with the description of the likelihood and sufficient statistics in the model (1)-(2). Let $\sigma_{yy.x} = \sigma_{yy} - \sigma_{xy}^2 \sigma_{xx}^{-1}$. Under assumptions $A1 - A2$ $Y \times X$ induce the family of distributions $dP_{\beta,\gamma}^{Y,X}(y,x)$, indexed by $(\beta,\gamma) \in \mathbb{R}^2$, that takes the form

$$(2\pi)^{-T}\left(\sigma_{yy.x}\sigma_{xx}\right)^{-T/2}\exp\left\{-\frac{1}{2}\sigma_{yy.x}^{-1}\sum_{t=1}^{T}\left[y_t - \alpha - \beta x_{t-1} - \sigma_{xx}^{-1}\sigma_{xy}\left(x_t - \gamma x_{t-1}\right)\right]^2 - \frac{1}{2}\sigma_{xx}^{-1}\sum_{t=1}^{T}\left(x_t - \gamma x_{t-1}\right)^2\right\} d\mu\left(y,x\right), \tag{6}$$

or, concentrating with respect to α (this is equivalent to the density of the maximal invariant with respect to transformations $(y_t, x_t) \rightarrow (y_t + a, x_t)$, $a \in \mathbb{R}$),

$$C_\Sigma h\left(y,x\right)\exp\left[\beta S_\beta + \gamma S_\gamma - \frac{1}{2}\left(\beta - \sigma_{xx}^{-1}\sigma_{xy}\gamma\right)^2 S_{\beta\beta} - \frac{1}{2}\gamma^2 S_{\gamma\gamma}\right] d\mu,$$

which we recognize as a curved exponential family with sufficient statistics:

$$\begin{aligned}
S_\beta &= \sigma_{yy.x}^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}\left(y_t - \sigma_{xx}^{-1}\sigma_{xy}x_t\right),\\
S_\gamma &= \sigma_{xx}^{-1}\sum_{t=1}^{T}x_{t-1}x_t - \sigma_{xx}^{-1}\sigma_{xy}S_\beta,\\
S_{\beta\beta} &= \sigma_{yy.x}^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}^2,\\
S_{\gamma\gamma} &= \sigma_{xx}^{-1}\sum_{t=1}^{T}x_{t-1}^2.
\end{aligned}$$

Here $\tilde{x}_{t-1} = x_{t-1} - T^{-1}\sum_{s=1}^{T} x_{s-1}$, $C_\Sigma = (2\pi)^{-1}\left(\sigma_{yy}\sigma_{xx} - \sigma_{xy}^2\right)^{-1/2}$ is a normalizing constant and

$$h(y,x) = \exp\left\{-\frac{1}{2}\left[\sigma_{yy.x}^{-1}\sum_{t=1}^{T}\left(\tilde{y}_t - \sigma_{xx}^{-1}\sigma_{xy}\tilde{x}_t\right)^2 + \sigma_{xx}^{-1}\sum_{t=1}^{T}x_t^2\right]\right\}$$

is independent of β, γ ($\tilde{y}_t$ denotes deviation of y_t from the average). Let $S = \left(S_\beta, S_\gamma, S_{\beta\beta}, S_{\gamma\gamma}\right)$ and the subset we will use for conditioning $S_C = \left(S_\gamma, S_{\beta\beta}, S_{\gamma\gamma}\right)$. Note that S_C is sufficient for γ. We denote by $P_{\beta,\gamma}^{S}$ and $P_{\beta,\gamma}^{S_C}$ the probability measures induced by S and S_C, respectively. Absorbing the $C_\Sigma h(y,x)$ factor into the measure μ we can write the distribution of the family $P_{\beta,\gamma}^{S}$, where notation makes clear that we operate now on the family of sufficient statistics (which we justify in the proof of Theorem 1) as

$$dP_{\beta,\gamma}^{S}(s) = \exp\left(\beta s_\beta + \gamma s_\gamma - \frac{1}{2}\left(\beta - \sigma_{xx}^{-1}\sigma_{xy}\gamma\right)^2 s_{\beta\beta} - \frac{1}{2}\gamma^2 s_{\gamma\gamma}\right) d\mu(s),$$

where $s = \left(s_\beta, s_\gamma, s_{\beta\beta}, s_{\gamma\gamma}\right) \in \mathbb{R}^4$. From Lemma 2.7.8 in Lehmann (1997) there exist measures ξ and ν_{s_C}, such that the marginal and conditional probability measures are described by

$$dP_{\beta,\gamma}^{S_C}(s_C) = \exp\left(\gamma s_\gamma - \frac{1}{2}\left(\beta - \sigma_{xx}^{-1}\sigma_{xy}\gamma\right)^2 s_{\beta\beta} - \frac{1}{2}\gamma^2 s_{\gamma\gamma}\right) d\xi(s_C), \tag{7}$$

$$dP_{\beta}^{S_\beta|s_C}(s_\beta) = \exp\left(\beta s_\beta\right) d\nu_{s_C}(s_\beta). \tag{8}$$

Note that the conditional distribution is independent of γ. The proposed estimator is based on an inverted median of the $dP_{\beta}^{S_\beta|s_C}(s_\beta)$ distribution, derived earlier by Jansson and Moreira (2004) for testing point hypothesis on β. Specifically, let

$$F_{\beta}^{S_\beta|s_C}(u) = P_{\beta}^{S_\beta|s_C}\left\{s_\beta \le u\right\} = \int_{-\infty}^{u} \exp\left(\beta s_\beta\right) d\nu_{s_C}(s_\beta)$$

denote the corresponding cumulative distribution function. Its continuity in β and s_β, necessary for existence and uniqueness of the inverse function, is verified in the appendix. Since $F_{\beta}^{S_\beta|s_C}(u)$ is increasing and continuous in u for every $\beta \in \mathbb{R}$ and $S_C = s_C$, there exists a median function $med_T(\beta, s_C)$, such that $F_{\beta}^{S_\beta|s_C}(med_T(\beta, s_C)) =$

0.5. Since the family admits a monotone likelihood ratio, $med_T(\beta, s_C)$ is increasing in β (see Lehmann (1997, Thm. 3.3.2 (ii))). Therefore the inverse function $m_T^*(\cdot, s_C)$ exists $(m_T^*(med_T(\beta, s_C), s_C) = \beta)$ and $s_\beta \to m_T^*(s_\beta, s_C)$ defines the estimator on the partition $S_C = s_C$. Integrating over the support of S_C we obtain the unconditional representation of the estimator

$$s \to m_T^*(s). \tag{9}$$

Since Gaussian densities are absolutely continuous, we may restrict attention to non-randomized estimators.

Optimality of the proposed median unbiased estimator is summarized in the following theorem.

Theorem 1. *Estimator $s \to m_T^*(s)$ is optimal in the following sense: $m_T^*(s)$ minimizes risk*

$$E_{\beta,\gamma}^{Y,X}[L(\cdot, \beta)]$$

among all estimators of β that are median unbiased conditionally on $(S_{\beta\beta}, S_{\gamma\gamma}) = (s_{\beta\beta}, s_{\gamma\gamma})$, for any loss function $L \in \mathcal{L}_\beta$ and for every $(\beta, \gamma) \in \mathbb{R}^2$.

Note that $(S_{\beta\beta}, S_{\gamma\gamma})$, which are specific ancillaries for β, summarize variability of x and in that sense the estimator $m_T^*(s)$ is optimal in the class of estimators that are median unbiased irrespectively of the degree of variability in the regressor. A related result, strengthening the Rao-Blackwell Theorem and constraining the set of optimal estimators (under convex loss) to functions of complete sufficient statistics has been obtained by Lehmann and Scheffé (1950). Extension to a broader class of quasiconvex loss functions, employed in this paper, is discussed in Brown, Cohen, and Strawderman (1976).

This theorem can actually be stated as a corollary of a stronger result, providing optimal confidence bounds for parameter β. In this form it is stated and proved in the appendix. Here, to highlight the critical steps, we discuss the proof briefly, relegating mathematical detail to the appendix. The theorem generalizes the result in

Lehmann (1997, p.94-95), derived unconditionally under no nuisance parameters, by specifying the results in Pfanzagl (1979) to a curved exponential family considered in this paper. The proof consists of the following steps:

(1) Existence and uniqueness of $m_T^*(s)$ is verified by demonstrating continuity and monotonicity of the relevant conditional probability measure.

(2) Optimality on the partition $S_C = s_C$ is demonstrated along the lines of Lehmann (1997). Specifically, because the likelihood conditional on $S_C = s_C$ is monotone in β, we apply the theory of uniformly most accurate confidence bounds for β. What this says is that there exist lower $\left(m_T^l\right)$ and upper $\left(m_T^u\right)$ bounds for β at, respectively, $1-\alpha_l$ and $1-\alpha_u$ confidence levels (we take $\alpha_l + \alpha_u \leq 1$), such that

$$P_\beta^{S_\beta|s_C}\left\{m_T^l \leq \beta'\right\} \leq P_\beta^{S_\beta|s_C}\left\{\tilde{m}_T^l \leq \beta'\right\} \qquad \forall \beta' \leq \beta,$$

where $\tilde{m}_T^l$ is any other lower confidence bound, and similarly

$$P_\beta^{S_\beta|s_C}\left\{m_T^u \geq \beta'\right\} \leq P_\beta^{S_\beta|s_C}\left\{\tilde{m}_T^u \geq \beta'\right\} \qquad \forall \beta' \geq \beta$$

for the upper bound, where $\tilde{m}_T^u$ denotes any other upper bound. Lehmann (1997, Problem 3.21) shows that m_T^l will minimize $E_\beta^{S_\beta|s_C}\left[L_l\left(\cdot,\beta\right)\right]$ at its level α_l for any function L_l that is nonincreasing in m_T^l for $m_T^l < \beta$ and 0 for $m_T^l \geq \beta$. A symmetric argument shows that m_T^u will minimize $E_\beta^{S_\beta|s_C}\left[L_u\left(\cdot,\beta\right)\right]$ for any L_u that is nondecreasing in m_T^u for $m_T^u > \beta$ and 0 for $m_T^u \leq \beta$. It is taking this argument to the limit of $\alpha_l = \alpha_u = 0.5$, which results in a "point" confidence set that by construction coincides with the median unbiased estimator $m_T^*(s)$ derived above, and minimizes

$$E_\beta^{S_\beta|s_C}\left[L\left(\cdot,\beta\right)\right]$$

for any function $L \in \mathcal{L}_\beta$ in the class of procedures that are median unbiased conditionally on $S_C = s_C$.

(3) Next we demonstrate that $m_T^*(s)$ is median unbiased unconditionally, for all $(\beta,\gamma)\in\mathbb{R}^2$. This follows from measurability of $m_T^*(s)$, since then we can use conditional unbiasedness to find:

$$P^S_{\beta,\gamma}\{m_T^*(s)\leq\beta\}=\int P_\beta^{S_\beta|s_C}\{m_T^*(s_\beta)\leq\beta\}\,dP^{S_C}_{\beta,\gamma}(s_C)=\frac{1}{2}.$$

(4) Finally, to demonstrate optimality in the class of procedures that are median unbiased with respect to the family $P_{\beta,\gamma}^{S|s_{\beta\beta},s_{\gamma\gamma}}$, we take any median unbiased estimator within this class and show that, conditionally on $(S_{\beta\beta},S_{\gamma\gamma})=(s_{\beta\beta},s_{\gamma\gamma})$, it is median unbiased on the partition $S_C=s_C$ (which follows from completeness of the family $P_{\beta,\gamma}^{S_C|s_{\beta\beta},s_{\gamma\gamma}}$). That means it is inferior on the partition, and hence, by integration, inferior unconditionally. This argument goes along the lines of Pfanzagl (1979). More specifically, since conditioning on all the sufficient statistics leaves any function independent of parameters, we can rewrite any median unbiased estimator $m_T(y,x)$ as a function of sufficient statistics, say $m_T(s)$ (see the appendix for a rigorous argument). Next, we use conditional median unbiasedness together with completeness of the family $dP_{\beta,\gamma}^{S|s_{\beta\beta},s_{\gamma\gamma}}$ (Lehmann, 1997, Thm. 4.3.1.), to arrive at

$$P_\beta^{S_\beta|s_C}\{m_T(s)\leq\beta\}=\frac{1}{2}\quad a.e.\ s_\gamma\in S_\gamma \tag{10}$$

with $(s_{\beta\beta},s_{\gamma\gamma})$ fixed, which means that $m_T(s)$ is median unbiased on the partition $S_C=s_C$ for all $\beta\in\mathbb{R}$. But we already know that it is $m_T^*(s)$ that is optimal on this partition and thus $m_T(s)$ is conditionally inferior. It is detailed in the appendix how this statement, through integration with respect to $P_{\beta,\gamma}^{S_\gamma|s_{\beta\beta},s_{\gamma\gamma}}$, generalizes to optimality of $m_T^*(s)$ within the specified class.

The property of median unbiasedness, however similar in spirit, nevertheless is not equivalent to equivariance. There are instances (see Jeganathan (1995)) in which these can be used interchangeably as side conditions, although in general they are not equivalent constraints. It turns out, however, that in the setting of this paper

$m_T^*(s)$ is equivariant with respect to shifts in y of the form specified in the following Lemma:

Lemma 2. *The optimal median unbiased estimator $m_T^*(s)$ is equivariant with respect to transformations of sufficient statistics induced by*

$$y_t \to y_t + \bar{\beta}x_{t-1}$$

for any $\bar{\beta} \in \mathbb{R}$.

4. Asymptotic Theory

In this section we are concerned with asymptotic localization of our procedure around particular values for parameters (β, γ), specifically $(\beta_0, 1)$. Localizing γ around unity is pertinent to the motivation of the paper to study estimation in the presence of highly persistent regressors, that may, nevertheless, not be exactly integrated. Zooming at some prespecified β, on the other hand, which we denote by β_0 and assume to be in the local neighborhood of the true value of β, exemplifies local approach to global estimation, well documented in the statistical literature (see particularly Chapter 4 of Shiryaev and Spokoiny (2000) and Chapter 6 of Le Cam and Yang (2000)). It is characterized by inference performed on a local quadratic approximation to the likelihood around some preliminary, "pilot" estimate. Specifically, we assume, following in part notation of Jansson and Moreira (2004), that

$$\begin{aligned}\beta &= \beta_T(b) = \beta_0 + \delta_T b,\\ \gamma &= \gamma_T(c) = 1 + T^{-1}c,\end{aligned}$$

where $(b, c) \in \mathbb{R}^2$ (c need not be negative) and $\delta_T = T^{-1}\sigma_{xx}^{-1/2}\sigma_{yy.x}^{1/2}$. The $O\left(T^{-1}\right)$ neighborhood is standard in the literature and makes the local experiment not easily differentiable from that at $(\beta_0, 1)$, a property referred to as contiguity in the statistical texts cited above. β_0, our initial approximation to the true value of β, needs to fall in the $O_p\left(T^{-1}\right)$ neighborhood of the true value, a condition clearly satisfied by a super-consistent OLS or GMLE estimators. When viewed from this perspective estimation of this section can be thought of as providing a local correction to the OLS or GMLE estimators, that as argued earlier may be severely (locally) biased.

We keep errors *i.i.d.* and Gaussian for simplicity of exposition since it has been shown by Jansson and Moreira (2004) to be a least favorable distributional assumption. In addition, Jeganathan (1997) showed that if extra nuisance parameters describe the distribution of the errors their likelihood asymptotically separates from that of (β, γ) and hence does not affect inference on parameters of interest.

With the specified parametrization we find, building on (6), that the probability distribution of (Y, X) is

$$dP^{Y,X}_{\beta_T(b),\gamma_T(c)} = \exp\left(bR_\beta + cR_\gamma - \frac{1}{2}(b-\lambda c)^2 R_{\beta\beta} - \frac{1}{2}c^2 R_{\gamma\gamma}\right) dP^{Y,X}_{\beta_0,1}, \tag{11}$$

where $dP^{Y,X}_{\beta_0,1}$ corresponds here to (6) evaluated at $(\beta_0, 1)$, $\lambda = \sigma_{xy}\left(\sigma_{xx}\sigma_{yy.x}\right)^{-1/2}$ and $R = \left(R_\beta, R_\gamma, R_{\beta\beta}, R_{\gamma\gamma}\right)$ is a set of asymptotically sufficient statistics

$$R_\beta = \sigma_{xx}^{-1/2}\sigma_{yy.x}^{-1/2}T^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}\left(y_t - \beta_0 x_{t-1} - \sigma_{xx}^{-1}\sigma_{xy}\Delta x_t\right),$$

$$R_\gamma = \sigma_{xx}^{-1}T^{-1}\sum_{t=1}^{T}x_{t-1}\Delta x_t - \lambda R_\beta,$$

$$R_{\beta\beta} = \sigma_{xx}^{-1}T^{-2}\sum_{t=1}^{T}\tilde{x}_{t-1}^2,$$

$$R_{\gamma\gamma} = \sigma_{xx}^{-1}T^{-2}\sum_{t=1}^{T}x_{t-1}^2.$$

Lemmas 3 and 4 in Jansson and Moreira (2004) provide weak limits as well as a joint distribution of R under the assumption of fixed (b, c). Limiting processes $\mathcal{R}(b,c) \equiv \left(\mathcal{R}_\beta(b,c), \mathcal{R}_\gamma(b,c), \mathcal{R}_{\beta\beta}(c), \mathcal{R}_{\gamma\gamma}(c)\right)$ are functionals of independent Wiener and Ornstein-Uhlenbeck processes. Their limiting distribution is a member of a curved exponential family and in this sense replicates the properties of S, its finite sample equivalent.

Form of the asymptotic distribution (11) puts it in the locally asymptotically quadratic (LAQ) family, described thoroughly in Le Cam and Yang (2000). Applications to time series are discussed in Jeganathan (1995, 1997) and more recently Ploberger (2004). In short, this concept applies to situations in which log-likelihood ratios become asymptotically quadratic and can be seen to be closely related to the Bernstein-von Mises phenomenon of Bayesian analysis of posterior distributions becoming approximately normal (that is, quadratic) around the true value under quite general conditions.

The particular model we are concerned with here actually falls into a subclass of the LAQ family defined by Jeganathan (1995) as locally asymptotically Brownian functional (LABF). It is this observation that gives the asymptotically centering estimator of Jeganathan (1997) its asymptotic efficiency in the class of $M-estimators$. Specifying a fixed value for c, on the other hand, puts the model in yet another subclass of the LAQ family known as a locally asymptotically mixed normal (LAMN) family (this gives the infeasible AC estimator its minimax optimality).

To facilitate analysis of weak conditional convergence we follow the approach of Jansson and Moreira (2004) and define asymptotic conditional median unbiasedness of a sequence of arbitrary estimators $\{m_T(\cdot)\}$ as

$$\lim_{T\to\infty} E^{Y,X}_{\beta_T(b),\gamma_T(c)}\left[\left(1\left\{m_T(R)\leq\beta\right\}-0.5\right)f\left(R_{\beta\beta},R_{\gamma\gamma}\right)\right]=0 \qquad \forall f\in C_b\left(\mathbb{R}^2\right) \quad (12)$$

where $C_b\left(\mathbb{R}^2\right)$ is a class of continuous, bounded and real-valued functions on $\mathbb{R}^2$ and the expectation refers to $P^{Y,X}_{\beta_T(b),\gamma_T(c)}$.

The analysis of this section may suggest that, given β_0 in the neighborhood of the true β, we may adapt the techniques of the previous section and arrive at the conditionally asymptotically optimal median unbiased estimator of b, the local component of β. It will, in analogy with the finite sample theory, be based on an inverted median of the distribution of $\mathcal{R}_\beta(b,c)\,|\,(\mathcal{R}_C(b,c)=r_C)$, where $r_C=\left(r_\gamma,r_{\beta\beta},r_{\gamma\gamma}\right)$ is a realization of $\mathcal{R}_C(b,c)=\left(\mathcal{R}_\gamma(b,c),\mathcal{R}_{\beta\beta}(c),\mathcal{R}_{\gamma\gamma}(c)\right)$. Note that both $\mathcal{R}$ and its realization r depend implicitly on β_0. We denote the corresponding asymptotic median function by $med_A(b,r_C)$. Its continuity and monotonicity is verified in the appendix. The asymptotic counterpart of the estimator of last section, $m^*_L(r)$ ("L"

stands for *Local*) then solves

$$r_\beta = med_A\left(m_L^*\left(r\right), r_C\right) \tag{13}$$

on the partition $\mathcal{R}_C\left(b,c\right) = r_C$. Integrating with respect to $\mathcal{R}_C\left(b,c\right)$ yields an unconditional representation $r \to m_L^*\left(r,\beta_0\right)$ where we add β_0 to the list of arguments to acknowledge that m_L^* estimates only a local component of β around β_0. It seems natural, then, to use

$$m^*\left(r,\beta_0\right) = \beta_0 + \delta_T m_L^*\left(r,\beta_0\right) \tag{14}$$

as a global estimator of β. It turns out that substituting a discretized preliminary $T^{-1} - consistent$ estimator for β_0 is asymptotically innocuous and the optimality of the median unbiased estimator generalizes to the asymptotic setting as made precise in the following Theorem:

Theorem 2. *Assume there exists a preliminary estimator m_T^0 such that:*

(a) $T\left(m_T^0 - \beta\right) = O_p\left(1\right)$, *where β is the true value;*
(b) m_T^0 *takes values on a prespecified discretized subset of the parameter space.*

Then the estimator $m^*\left(r, m_T^0\right)$ as defined in (13)-(14) *is optimal in the following sense:*

$$\begin{aligned}\underline{\lim}_{T\to\infty} E^{Y,X}_{\beta_T(b),\gamma_T(c)}\left[L\left(m_T\left(R\right),\beta\right)\right] &\geq \lim_{T\to\infty} E^{Y,X}_{\beta_T(b),\gamma_T(c)}\left[L\left(m^*\left(R, m_T^0\right),\beta\right)\right] \\ &= E^{\mathcal{R}}_{b,c}\left[L\left(m^*\left(\mathcal{R}\left(b,c\right),\beta_0\right),\beta\right)\right]\end{aligned}$$

among all sequences of asymptotically median unbiased estimators $\{m_T\left(\cdot\right)\}$ of β in the sense of (12) *and for every $\left(b,c\right) \in \mathbb{R}^2$ and $L \in \mathcal{L}_\beta$ with at most countably many discontinuities.*

Discretization of the preliminary estimator m_T^0 guarantees uniform convergence to experiments with β_0 fixed in the vicinity of β. It also guards the preliminary estimator from seeking peculiarities in the likelihood, an aspect that can make construction of global estimates invalid, see Le Cam and Yang (2000) for a detailed discussion.

Conditions (a) and (b) are satisfied by a discretized OLS estimator. One possibility is to compute the OLS estimator up to an approximation of order T^{-1}, that is replacing $m_T^{OLS}(y, x)$ by $\left[T m_T^{OLS}(y, x)\right] / T$, where $[\cdot]$ denotes the greatest lesser integer function.

5. Monte Carlo Results

In this section we subject the optimal median unbiased estimator (OMUB) to a Monte Carlo evaluation. We measure its performance across the relevant local-to-unity domain of $\gamma = \{1 + c/T : -20 \leq c \leq 0\}$ in terms of mean square losses, $L(m, \beta) = (m - \beta)^2$. We generate artificial data from the model (1)-(2) under assumptions $A1 - A2$ and the degree of correlation between innovations, $\rho = \sigma_{xy} \times \left(\sigma_{yy}\sigma_{xx}\right)^{-1/2}$, set at 0.5 or 0.9 (the latter is calibrated to our empirical exercise, in which the long-run correlations fall between 0.9 and 1).

Performance of the median unbiased estimator derived in previous sections is contrasted with that of some of the alternatives: the OLS, the GMLE($c = 0$) and the AC estimator, discussed in Section 2. Such comparison is interesting since these are the available alternatives that do not belong to the class of median unbiased estimators and therefore are not covered by the optimality statements of Theorems 1 and 3. It thus remains an open question how these estimators perform in relation to each other. Theoretical considerations would advocate use of ordinary (or generalized) least squares on stationary regressors (setting approximated by large negative values of c) and the GMLE($c = 0$) for exactly integrated variables. The OMUB is, on the other hand, designed to work well irrespectively of the value of c, as is the AC estimator. The last alternative we are looking at is based on pretesting for the unit root (using the point optimal unit root test of Elliott, Rothenberg, and Stock (1996) with size 0.05) and employing a procedure designed for either the stationary or the integrated specification, whichever is favored by the data. As discussed in Stock and Watson (1996), whether the size is fixed or tending to zero, this method selects between two incorrect models, introducing biases inherent in the respective procedures.

We compute the median unbiased estimator by simulating 20000 draws of the sufficient statistics R in the neighborhood of the true parameter value and use, following Polk, Thompson, and Vuolteenaho (2004), the nonparametric nearest neighbor technique to estimate the conditional median. Specifically, we search the parameter space for β_0 (that enters R) such that the realized value of R_β coincides with the conditional median of R_β at β_0, given $\left(R_{\beta\beta}, R\gamma\gamma\right)$. We estimate the latter with the median of 5 percent of draws for which the conditioning statistics fall closest (in the sup norm) to their respective conditioning values.[5] Results of the simulation are summarized in Figure 1. We notice that the median unbiased estimator dominates available alternatives across the domain of near nonstationary, with little exception of the Gaussian MLE that, by construction, performs marginally better at $c = 0$. It seems to be the most versatile of all the procedures, performing very well on integrated regressors but at the same time guarding against heavy losses as we move

[5]See Polk, Thompson, and Vuolteenaho (2004) for further details on the nearest neighbor technique and an alternative neural network procedure.

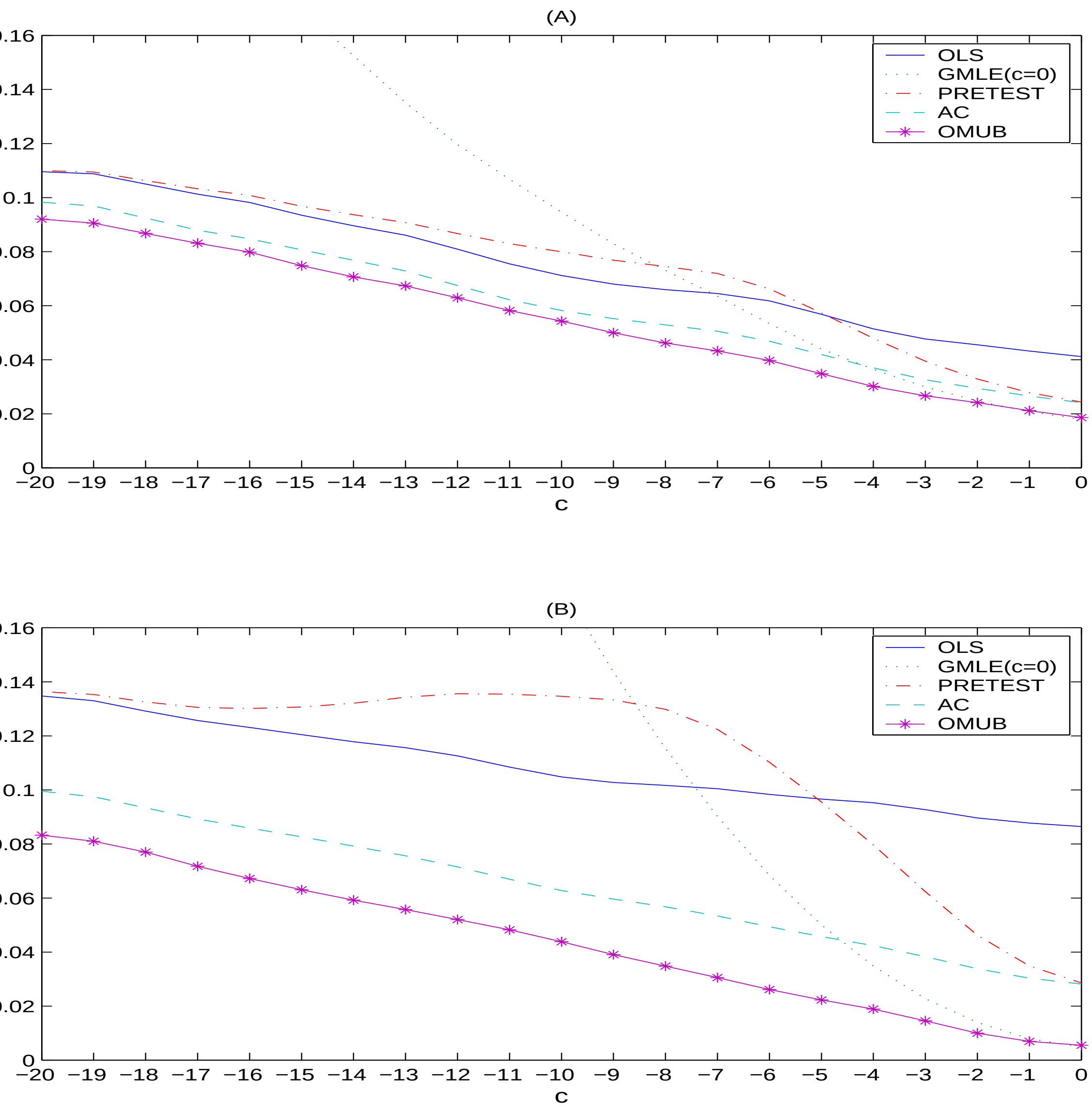

FIGURE 1. Mean Square Losses for $\rho = 0.5$ (A) and $\rho = 0.9$ (B). Scaled by T.

away from the unit root.[6] By comparing two panels of the figure we notice that relative efficiency gains attained by the median unbiased estimator increase with the degree of endogeneity.

[6]Bayesian procedures are represented in this comparison by the AC estimator, which, as noted earlier, is the limit of a sequence of Bayes estimators under proper priors that flatten out. An alternative approach, advocated in the literature, would involve imposing Jeffrey prior on the parameter γ and searching for the mean of the marginal posterior of β. This, however, is sensitive to the sample

To verify the theoretical robustness of the median unbiased estimator to the choice of a loss function we present a second set of results for two asymmetric loss functions that put different weight on positive and negative absolute deviations. Specifically, we choose:

$$L(m, \beta; a) = \{a - 3 \times 1[m < \beta]\}(m - \beta) \tag{15}$$

with $a = \{1, 2\}$. We expect the median unbiased estimator to perform similarly well irrespectively of which of the specifications is used, in contrast to the other estimators which are likely to be sensitive to any asymmetries in the penalty imposed on positive and negative errors. This is in fact the case, as documented in Figure 2.

6. Feasible Inference

In this section we extend the basic framework employed earlier for demonstrational purposes and indicate how the proposed estimator can be applied empirically. We follow Jansson and Moreira (2004) who show, in the context of the model of this paper, that the Gaussian assumption on the errors is least favorable in a sense that the model retains the same asymptotic structure under more general, non-parametric specification of the errors, and hence the same risk can be achieved. They also demonstrate that substituting consistent estimators for quantities specifying long-run behavior of the $\{\varepsilon_t\}$ process, assumed known in the preceding analysis, does not alter asymptotic reasoning. We reproduce some of their results for easy reference. Specifically, we modify the description of the $\{x_t\}$ process as follows:

$$\begin{aligned} x_t &= \mu_x + \chi_t, \\ \chi_t &= \gamma\chi_{t-1} + \varphi(l)\,\varepsilon_t^x, \end{aligned}$$

where

size as the prior on γ remains fixed and does not concentrate around 1, which results in large losses when T is large.

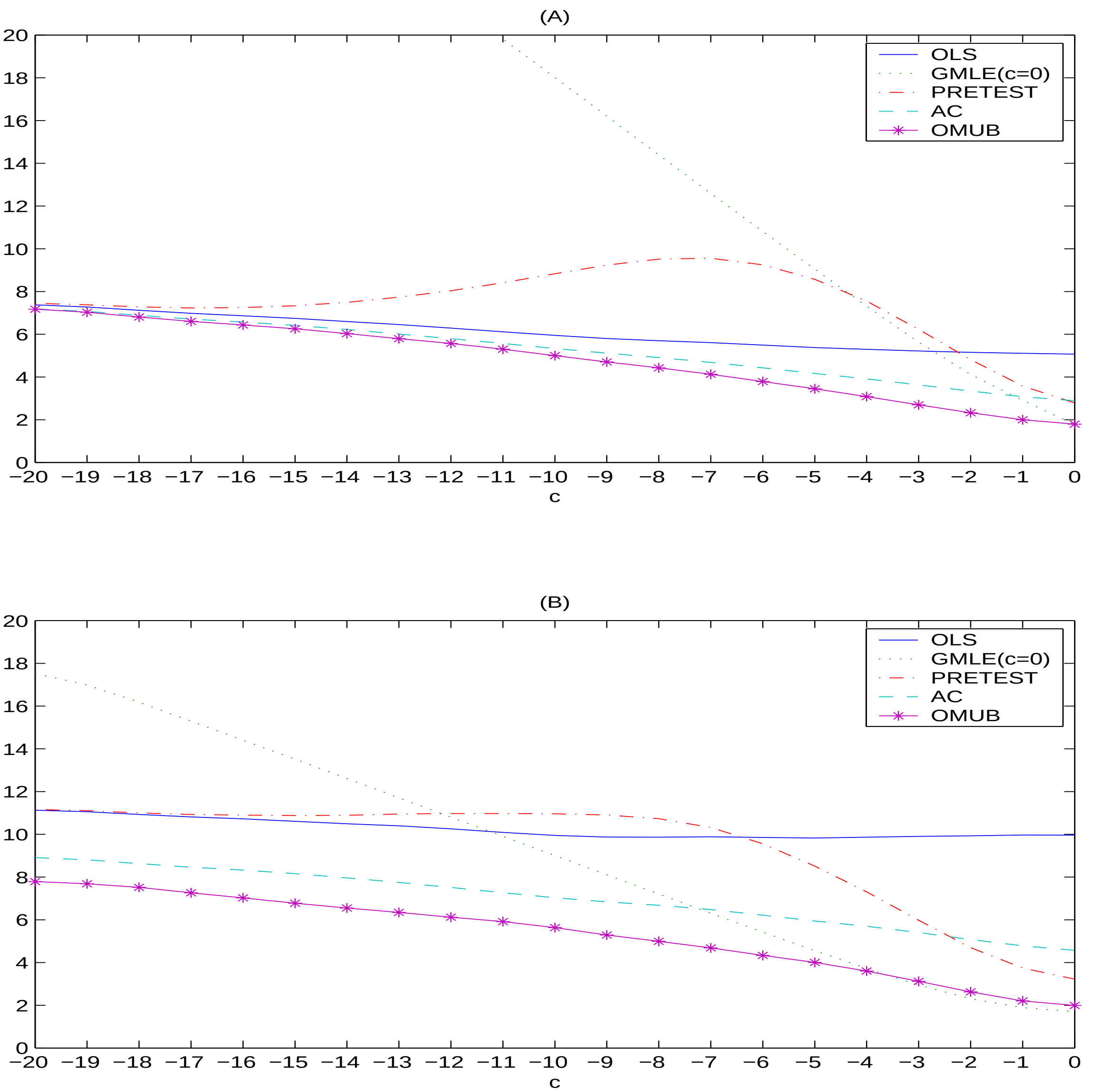

FIGURE 2. Mean Asymmetric Losses (specification (15)) with $a = 1$ (A) and $a = 2$ (B). $\rho = 0.9$, scaled by T.

$A1^* : \chi_0 = 0$,

$A2^* : \varphi(l) = 1 + \sum_{i=1}^{\infty} \varphi_i l^i$ admits $\varphi(1) \neq 0$ and $\sum_{i=1}^{\infty} i \left|\varphi_i\right| < \infty$,

$A3^* : E\left(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, ...\right) = 0$, $E\left(\varepsilon_t \varepsilon_t' | \varepsilon_{t-1}, \varepsilon_{t-2}, ...\right) = \Sigma$, where Σ is some positive definite matrix and $\sup_t E\left[\left\|\varepsilon_t\right\|^{2+\rho}\right] < \infty$ for some $\rho > 0$, where $\varepsilon_t = \left(\varepsilon_t^y, \varepsilon_t^x\right)'$.

Jansson and Moreira (2004) show that under this assumption appropriately modified sufficient statistics, with no unknown nuisance parameters, converge to the same limit as obtained in the Gaussian setting of Section 4. This, in part, has also been noted by Jeganathan (1997), who shows that the likelihood representation of (β, γ) will asymptotically separate from that of the errors. Let the long-run variance of $\left(\varepsilon_t^y, \varphi(l)\, \varepsilon_t^x\right)$ be specified as

$$\Omega = \begin{pmatrix} \omega_{yy} & \omega_{yx} \\ \omega_{xy} & \omega_{xx} \end{pmatrix} = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[\begin{pmatrix} \varepsilon_t^y \\ \varphi(l)\, \varepsilon_t^x \end{pmatrix} \begin{pmatrix} \varepsilon_s^y \\ \varphi(l)\, \varepsilon_s^x \end{pmatrix}' \right]$$

with $\hat{\Omega}$ its consistent estimator, partitioned accordingly. Let $\hat{\omega}_{yy.x} = \hat{\omega}_{yy} - \hat{\omega}_{xx}^{-1}\hat{\omega}_{xy}^2$, $\hat{\lambda} = \hat{\omega}_{xy}\left(\hat{\omega}_{xx}\hat{\omega}_{yy.x}\right)^{-1/2}$, $x_0 = x_1$ and define $\hat{\chi}_t = x_t - x_1$. Then Jansson and Moreira (2004) prove (in Theorem 6) that the statistics

$$\begin{aligned}
\hat{R}_\beta &= \hat{\omega}_{xx}^{-1/2}\hat{\omega}_{yy.x}^{-1/2} T^{-1} \sum_{t=1}^{T} \tilde{x}_{t-1}\left(y_t - \beta_0 x_{t-1}\right) \\
&\quad - \hat{\lambda}\left[\frac{1}{2}\left(\hat{\omega}_{xx}^{-1} T^{-1} \hat{\chi}_T^2 - 1\right) - \hat{\omega}_{xx}^{-1} T^{-2} \hat{\chi}_T \sum_{t=1}^{T} \hat{\chi}_{t-1}\right], \\
\hat{R}_\gamma &= \frac{1}{2}\left(\hat{\omega}_{xx}^{-1} T^{-1} \hat{\chi}_T^2 - 1\right) - \hat{\lambda}\hat{R}_\beta, \\
\hat{R}_{\beta\beta} &= \hat{\omega}_{xx}^{-1} T^{-2} \sum_{t=1}^{T} \tilde{x}_{t-1}^2, \\
\hat{R}_{\gamma\gamma} &= \hat{\omega}_{xx}^{-1} T^{-2} \sum_{t=1}^{T} \hat{\chi}_{t-1}^2,
\end{aligned}$$

jointly converge to $\mathcal{R}$, the limit of R specified in Section 4. Consequently, we have the following theorem:

Theorem 3. *Let $\hat{R} = \left(\hat{R}_\beta, \hat{R}_\gamma, \hat{R}_{\beta\beta}, \hat{R}_{\gamma\gamma}\right)$. Then under assumptions $A1^* - A3^*$, with any $\hat{\Omega} \to_p \Omega$, we have*

$$\lim_{T\to\infty} E_{\beta_T(b),\gamma_T(c)}^{Y,X}\left[L\left(m^*\left(\hat{R}, \beta_0\right), \beta\right)\right] = E_{b,c}^{\mathcal{R}}\left[L\left(m^*\left(\mathcal{R}(b,c), \beta_0\right), \beta\right)\right]$$

for every $(b, c) \in \mathbb{R}^2$, where $L \in \mathcal{L}_\beta$ with at most countably many discontinuities.

What this says is that, asymptotically, the median unbiased estimator based on the feasible counterpart of R achieves optimality as specified in Theorem 2.

Finally we note that Jansson and Moreira (2004) provide integral representation of the joint distribution of $\mathcal{R}(b, c)$ in their Lemma 4 and Theorem 7. In principle it can be used to obtain precise estimates of the median function, however numerical integration procedures are not yet stable enough to produce robust estimates away from β_0.[7] In the empirical application, to be discussed next, we will therefore use the nearest neighbor technique for conditioning, as described in the last section.[8]

7. Empirical Application

As noted in the introduction, the optimal median unbiased estimator proposed in this paper is particularly attractive in forecasting macroeconomic or financial indexes. Assumptions of our model well describe the observed characteristics of the common and successful explanatory variables, such as dividend-price ratio or earnings-price ratio in financial applications or various macroeconomic fundamentals (measures of output or inflation) employed more broadly. Importantly, any changes in these variables are very persistent and are likely not to be exogenous with respect to the series of interest.

In this section we consider one possible application, in which we employ the OMUB estimator to forecast excess asset returns with the earnings-price ratio, shown to carry forecasting power by Campbell and Yogo (2003). Although this problem has been extensively studied in the past (see Campbell and Yogo (2003), Polk, Thompson, and Vuolteenaho (2004), Torous, Valkanov, and Yan (2001), Lanne (2002) and references therein) all theoretical and empirical investigations we are familiar with almost exclusively concern testing predictability of asset returns, rather than an arguably equally relevant task of point (or interval) forecasting. The following application is not meant as a comprehensive study of this very interesting subject but as

[7] I thank Michael Jansson for sharing the numerical routines that perform the integration.

[8] Matlab routine that computes the OMUB estimator is available from the author upon request.

an illustration of the possible and appealing use of the methods developed in this paper.

Data we use come from the recent study by Campbell and Yogo (2003) and span 1952-2002.[9] Returns are value-weighted monthly indexes from NYSE/AMEX markets provided by the Center for Research in Security Prices (CRSP). Excess returns are computed in logs over the riskfree return, taken to be the 1-month T-bill rate. The predictive variable is the log earnings-price ratio, computed by Campbell and Yogo (2003) as a moving average of earnings of the S&P 500 over the past ten years divided by the current price. As evident from Figure 3, excess returns demonstrate

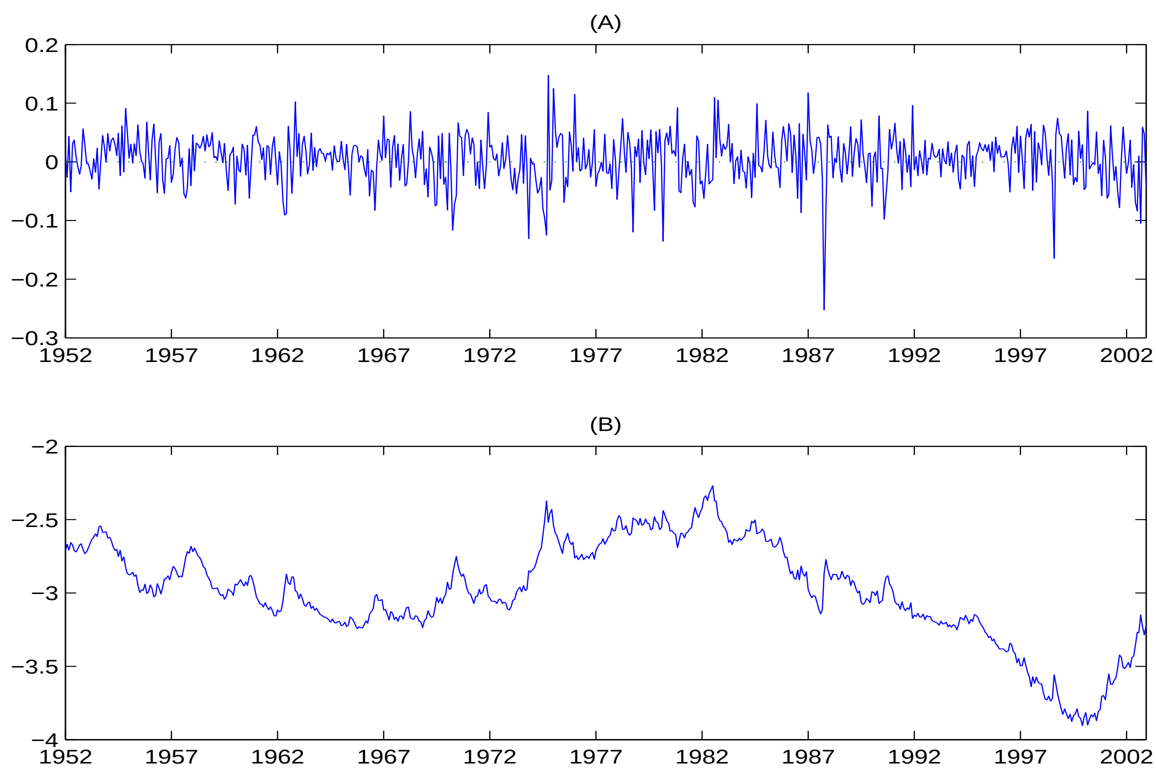

FIGURE 3. Excess asset returns (A) and Earnings-price ratio (B)

high volatility whereas our chosen predictor, the earnings-price ratio appears to be

[9]I thank John Campbell and Motohiro Yogo for providing access to the data set.

highly persistent. In fact, Campbell and Yogo (2003) estimate its confidence interval for the autoregressive coefficient c to be $[-6.95,\ 3.86]$, containing the exact unit root. Also, by construction, shocks to earnings-price ratio are highly negatively correlated with the return shocks; over the entire sample the long-run correlation is being estimated at -0.96.

We compute the median unbiased estimator as described in the preceding sections, employing discretized OLS (as a pilot estimate) and a long-run covariance matrix estimated using a Quadratic Spectral kernel with automatic bandwidth selection, as specified in Andrews (1991). The alternative procedures that we consider are OLS, GMLE(c=0) and AC estimators and the Bayes estimator with the Jeffrey prior on γ.[10]

We present our results in the form of rolling estimates of the slope coefficient over 1980-2002 in Figure 4. At every date estimation results are obtained using observa-

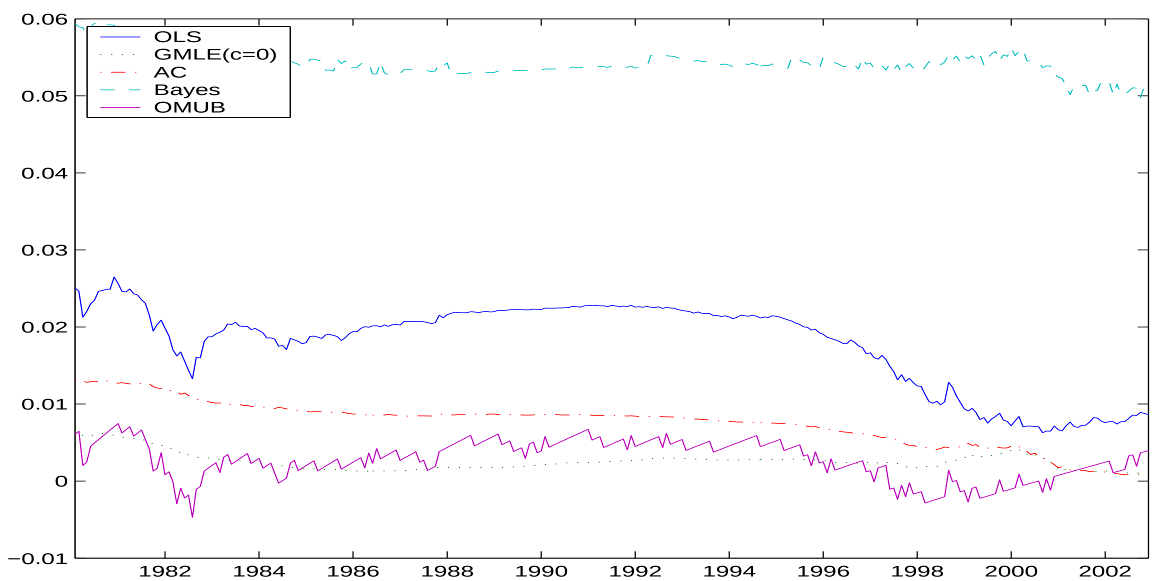

FIGURE 4. Estimates of β coefficient

tions from the beginning of sample up until the specified date. We notice that the

[10]Jeffrey prior is proportional to $1/\sqrt{1-\gamma^2}$ and hence favors values of γ close to 1.

median unbiased estimates are in general small in magnitude, of similar order as the GMLE(c=0) and AC estimates, but in contrast appear to be more versatile, accommodating to the changing economic environment. In particular, they are the only estimates that oscillate around zero in the regime-switching years at the beginning of the 80's and at the end of 90's.

Next, we compute the average losses from rolling pseudo out-of-sample forecasts over this period and detail the results in Table 1. Columns of this table summarize efficiency of estimators when measured with squared deviations (SD), absolute deviations (AD) or linear-asymmetric loss specification (15) with $a = 1$ or $a = 2$, respectively.

	SD	AD	$a = 1$	$a = 2$
OLS	2.74	39.79	47.52	71.84
GMLE(c=0)	1.92	32.83	49.71	48.77
AC	1.93	32.94	49.63	49.18
Bayes	2.40	38.36	63.02	52.06
OMUB	1.93	32.87	49.33	49.28

TABLE 1. Average loss in rolling forecasts over 1980-2002 measured with: square deviations (SD), absolute deviations (AD) and linear-asymmetric formulation (15) with $a = 1$ or $a = 2$. Scaled by 10^3.

Our results indicate that the OMUB estimator performs consistently well, either producing the lowest average loss or falling within a very small margin of the respective most efficient procedure. Results in the last two columns, corresponding to asymmetric losses (different costs of positive and negative errors), are very interesting in view of our earlier theoretical finding of robustness of the median unbiased procedure to the choice of a loss function. Although estimation errors only partly contribute to the observed forecast errors we would nevertheless expect to see some evidence of robustness in the empirical results. This is in fact the case; average losses under two asymmetric specifications are almost equal (similar observation holds for the GMLE(c=0) and AC procedures), in contrast to the OLS and Bayes estimators.

Qualitatively similar results are obtained when dividend-price ratio is used as an explanatory variable or if the sample period is extended back to the end of 1926, when available data starts.

8. Summary and Concluding Remarks

In this paper we have investigated optimal estimation procedures for the slope coefficient in regressions with a highly persistent and predetermined regressor. In this empirically relevant framework we have designed an optimal estimator in the class of conditionally median unbiased procedures. Importantly, the optimality statement extends to interval estimation and holds under a very general specification of the loss function.

Although simple, the theoretical framework employed in this paper yields itself to some interesting empirical applications. The preliminary results on forecasting asset returns call for further work in the area. It would be particularly interesting to see if any improvements in forecasting individual stocks at higher frequencies can be attained. In macroeconomics, on the other hand, one issue that has recently received a lot of attention and that can easily be investigated using the techniques developed in this paper concerns the effect of the level of inflation on the frequency of price adjustments.

There are important extensions of this work that merit further inquiry. First, the method should be modified to apply to a multivariate regression. This has already been resolved by Polk, Thompson, and Vuolteenaho (2004) for the testing problem considered by Jansson and Moreira (2004); their analysis can be adapted to the setting of this paper. On another front, we may expect the median unbiased estimator to perform well under heavy-tailed distributions (see Thompson (2003) for a recent study), but this remains to be established.

Last but not least, it seems that the general strategy of constructing optimal median unbiased estimators presented in this paper can be applied to other econometric models, whenever curvature of the likelihood can be circumvented by conditioning on sufficient statistics for the nuisance parameters.

9. APPENDIX: PROOFS

Proof of Lemma 1. Note that the local representation of the probability distribution (11), $dP^{Y,X}_{\beta_T(b),\gamma_T(c)}$, (which by construction is invariant to shifts of the form $(y_t, x_t) \rightarrow (y_t + a, x_t)$, $a \in \mathbb{R}$) can alternatively be written as

$$\exp\left[\theta' W_T(\theta_0) - \frac{1}{2}\theta' K_T \theta\right] dP^{Y,X}_{\beta_0,1}$$

with $\theta = (b, c)'$, $\theta_0 = (\beta_0, 1)'$, $W_T = \left(R_\beta, R_\gamma\right)'$ and

$$K_T = \left[\begin{array}{cc} R_{\beta\beta} & -\lambda R_{\beta\beta} \\ -\lambda R_{\beta\beta} & \lambda^2 R_{\beta\beta} + R_{\gamma\gamma} \end{array}\right].$$

Since K_T is necessarily positive definite and the sequences of probability measures $\left\{P^{Y,X}_{\beta_T(b),\gamma_T(c)}\right\}$ and $\left\{P^{Y,X}_{\beta_0,1}\right\}$ are contiguous (see Jeganathan (1997), Theorem 1), the structure falls into the LAQ family (see Le Cam and Yang (2000, Definition 6.1) for a precise statement and conditions). The AC estimator (see Le Cam and Yang (2000, Section 6.3) or Jeganathan (1995)), defined as maximizing local approximation to the likelihood, is specified as

$$m^{AC}_T(y, x) = \hat{\theta}_T + \delta_T K_T^{-1} W_T\left(\hat{\theta}_T\right),$$

where $\hat{\theta}_T$ denotes any preliminary estimator satisfying conditions (a)-(b) of Theorem 2. It is a straightforward algebra to show that $m^{AC}_T(y, x)$ equals the OLS estimator from regressing $y_t - \sigma_{xx}^{-1}\sigma_{xy}\left(x_t - \hat{\gamma}x_{t-1}\right)$ on a constant and x_{t-1}, where $\hat{\gamma}$ is the OLS estimator from regressing x_t on x_{t-1}. □

Proof of Theorem 1. Theorem 1 is a corollary of the following, more general result:

Theorem 4. *Let m_T^l and m_T^u denote lower and upper confidence bounds for β, such that*

$$F_\beta^{S|s_C}(m_T^l) = \alpha_l, \tag{16}$$

$$F_\beta^{S|s_C}(m_T^u) = 1-\alpha_u, \tag{17}$$

where $\alpha_l+\alpha_u \leq 1$. Then m_T^l and m_T^u are uniformly most accurate unbiased confidence bounds for β in a sense that the interval $C_T^* = \left(m_T^l, m_T^u\right)$ minimizes

$$E_{\beta,\gamma}^{Y,X}\left[L_{CI}\left(\cdot,\beta\right)\right]$$

among all intervals that are conditionally unbiased at level $1-\alpha_l-\alpha_u$, where conditioning refers to $\left(S_{\beta\beta}, S_{\gamma\gamma}\right) = \left(s_{\beta\beta}, s_{\gamma\gamma}\right)$. The loss function is specified as

$$L_{CI}\left(C_T^*,\beta\right) = L_l(m_T^l,\beta) + L_u\left(m_T^u,\beta\right),$$

where L_l is nonincreasing in m_T^l for $m_T^l < \beta$ and 0 otherwise and L_u is nondecreasing in m_T^u for $m_T^u > \beta$ and 0 otherwise.

Theorem 1 follows on setting $\alpha_l = \alpha_u = 0.5$. This results in $m_T^* = m_T^l = m_T^u$ by continuity of $F_\beta^{S|s_C}(\cdot)$ (see proof to Theorem 4 below). Also note that any $L_{CI} \in \mathcal{L}_\beta$. □

Proof of Lemma 2. In the space of sufficient statistics the specified transformation reads

$$\begin{aligned} s_\beta &\to s_\beta + \bar{\beta}s_{\beta\beta}, \\ s_\gamma &\to s_\gamma - \sigma_{xx}^{-1}\sigma_{xy}\bar{\beta}s_{\beta\beta}, \\ s_{\beta\beta} &\to s_{\beta\beta}, \\ s_{\gamma\gamma} &\to s_{\gamma\gamma}. \end{aligned}$$

We may recall that the estimator $m_T^*(s)$ is defined as $s_\beta = med_T\left(m_T^*, s_C\right)$. To prove equivariance we need to show that $\bar{m}_T^*$, which solves this equation under specified transformations, that is

$$s_\beta + \bar{\beta}s_{\beta\beta} = med_T\left[\bar{m}_T^*, \left(s_\gamma - \sigma_{xx}^{-1}\sigma_{xy}\bar{\beta}s_{\beta\beta}, s_{\beta\beta}, s_{\gamma\gamma}\right)\right], \tag{18}$$

equals $\bar{m}_T^* = m_T^* + \bar{\beta}$ for any $\bar{\beta} \in \mathbb{R}$. After substituting for s_β condition (18) becomes:

$$med_T\left[\bar{m}_T^*, \left(s_\gamma - \sigma_{xx}^{-1}\sigma_{xy}\bar{\beta}s_{\beta\beta}, s_{\beta\beta}, s_{\gamma\gamma}\right)\right] - med_T\left[m_T^*, \left(s_\gamma, s_{\beta\beta}, s_{\gamma\gamma}\right)\right] = \bar{\beta}s_{\beta\beta}. \quad (19)$$

Rewriting the conditional median function and the sufficient statistics explicitly in terms of the underlying random variables yields:

$$\begin{aligned}
& med_T\left(m_T^*, s_C\right) \\
= \; & med\left\{\left(m_T^* - \sigma_{xx}^{-1}\sigma_{xy}\gamma\right)\sigma_{yy.x}^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}^2 + \sigma_{yy.x}^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}\varepsilon_t \right| \\
& \left. \begin{array}{l} \sigma_{xx}^{-1}\sum_{t=1}^{T}x_{t-1}x_t - \sigma_{xx}^{-1}\sigma_{xy}\sigma_{yy.x}^{-1}\left[\left(m_T^* - \sigma_{xx}^{-1}\sigma_{xy}\gamma\right)\sum_{t=1}^{T}\tilde{x}_{t-1}^2 + \sum_{t=1}^{T}\tilde{x}_{t-1}\varepsilon_t\right] = s_\gamma, \\ \sigma_{yy.x}^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}^2 = s_{\beta\beta},\; \sigma_{xx}^{-1}\sum_{t=1}^{T}x_{t-1}^2 = s_{\gamma\gamma} \end{array}\right\} \\
= \; & \left(m_T^* - \sigma_{xx}^{-1}\sigma_{xy}\gamma\right)s_{\beta\beta} + med\left[\sigma_{yy.x}^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}\varepsilon_t \right| \\
& \left. \begin{array}{l} \sigma_{yy.x}^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}\varepsilon_t = -\sigma_{xx}\sigma_{xy}^{-1}\left(s_\gamma - \sigma_{xx}^{-1}\sum_{t=1}^{T}x_{t-1}x_t\right) - \left(m_T^* - \sigma_{xx}^{-1}\sigma_{xy}\gamma\right)s_{\beta\beta}, \\ \sigma_{yy.x}^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}^2 = s_{\beta\beta},\; \sigma_{xx}^{-1}\sum_{t=1}^{T}x_{t-1}^2 = s_{\gamma\gamma} \end{array}\right]
\end{aligned}$$

and similar expression obtains for $med_T\left[\bar{m}_T^*, \left(s_\gamma - \sigma_{xx}^{-1}\sigma_{xy}\bar{\beta}s_{\beta\beta}, s_{\beta\beta}, s_{\gamma\gamma}\right)\right]$. Clearly, then

$$med_T\left[\bar{m}_T^*, \left(s_\gamma - \sigma_{xx}^{-1}\sigma_{xy}\bar{\beta}s_{\beta\beta}, s_{\beta\beta}, s_{\gamma\gamma}\right)\right] - med_T\left(m_T^*, s_C\right) = \left(\bar{m}_T^* - m_T^*\right)s_{\beta\beta} + q\left(m_T^*, \bar{m}_T^*, s_C\right) \quad (20)$$

where

$$\begin{aligned}
q\left(m_T^*, \bar{m}_T^*, s_C\right) = \; & med\left[\sigma_{yy.x}^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}\varepsilon_t \,\middle|\, \begin{array}{l} \sigma_{yy.x}^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}\varepsilon_t = \left(\bar{\beta} - \bar{m}_T^*\right)s_{\beta\beta} - Q, \\ \sigma_{yy.x}^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}^2 = s_{\beta\beta},\; \sigma_{xx}^{-1}\sum_{t=1}^{T}x_{t-1}^2 = s_{\gamma\gamma} \end{array}\right] \\
& - med\left[\sigma_{yy.x}^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}\varepsilon_t \,\middle|\, \begin{array}{l} \sigma_{yy.x}^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}\varepsilon_t = -m_T^*s_{\beta\beta} - Q, \\ \sigma_{yy.x}^{-1}\sum_{t=1}^{T}\tilde{x}_{t-1}^2 = s_{\beta\beta},\; \sigma_{xx}^{-1}\sum_{t=1}^{T}x_{t-1}^2 = s_{\gamma\gamma} \end{array}\right]
\end{aligned}$$

and $Q = \sigma_{xx}\sigma_{xy}^{-1}\left(s_\gamma - \sigma_{xx}^{-1}\sum_{t=1}^{T}x_{t-1}x_t\right) + \sigma_{xx}^{-1}\sigma_{xy}\gamma s_{\beta\beta}$. The two conditioning sets in the last display will coincide if and only if $\bar{m}_T^* = m_T^* + \bar{\beta}$, that is,

$$q\left(m_T^*, m_T^* + \bar{\beta}, s_C\right) = 0.$$

We further note that

$$q\left(m_T^*,\bar{m}_T^*,s_C\right) \;>\; 0 \; for \; \bar{m}_T^* < m_T^* + \bar{\beta}, \tag{21}$$

$$q\left(m_T^*,\bar{m}_T^*,s_C\right) \;<\; 0 \; for \; \bar{m}_T^* > m_T^* + \bar{\beta}. \tag{22}$$

From (19) and (20) we find that the sufficient condition for equivariance is

$$\left(\bar{m}_T^* - m_T^* - \bar{\beta}\right) s_{\beta\beta} + q\left(m_T^*,\bar{m}_T^*,s_C\right) = 0$$

for any $\bar{\beta} \in \mathbb{R},\ \gamma \in \mathbb{R}$ and $S_C = s_C$. This, by (21)-(22), is uniquely solved by

$$\bar{m}_T^* = m_T^* + \bar{\beta}.$$

□

Proof of Theorem 2. Throughout this proof let $(b,c) \in \mathbb{R}^2$ be fixed. Denote by B an arbitrarily large bounded interval around $\beta_T(b)$. Let $P^R_{\beta_T(b),\gamma_T(c)}$ and $\tilde{P}^R_{\beta_T(b),\gamma_T(c)}$ denote probability measures generated by R with fixed and estimated β_0, respectively, and satisfying conditions (a) and (b) of Theorem 2. Then, from Proposition 6.3.3 and Remark 6.3.5 in Le Cam and Yang (2000), we learn that $L_1 - norms$

$$\sup_B ||P^R_{\beta_T(b),\gamma_T(c)} - \tilde{P}^R_{\beta_T(b),\gamma_T(c)}||$$

will tend to zero uniformly over B. That means that experiments with an estimated β_0 are asymptotically equivalent to those with fixed β_0 in the vicinity of the true β, and hence we take β_0 as fixed in what follows (see Shiryaev and Spokoiny (2000, Chapter 4) for a detailed treatment).

Existence, uniqueness and continuity in b of the median function $med_A(b,r_C)$ of the distribution of $\mathcal{R}_\beta(b,c)\,|\,(\mathcal{R}_C(b,c)=r_C)$ follows from Lemma 4(b) in Jansson and Moreira (2004), who establish exponential representation of this distribution, and arguments of Theorem 4 of this paper (that use Theorem 2.7.9 (i) in Lehmann (1997)). Its continuity in conditioning arguments r_C follows from Lemma 4(a) and Lemma 11(a) of Jansson and Moreira (2004). These authors prove continuity of critical functions of $\mathcal{R}_\beta(0,0)\,|\,(\mathcal{R}_C(0,0)=r_C)$ that extends to the domain of (b,c)

through exponential representation of the family $\mathcal{R}\left(b,c\right)$ established in their Lemma 4(a). This result, Lemma 3 in Jansson and Moreira (2004) and continuous mapping theorem (CMT) yield

$$med_A\left(b,R_C\right)\to_d med_A\left(b,\mathcal{R}_C\right)$$

and similarly for its inverse function

$$m_L^*\left(R\right)\to_d m_L^*\left(\mathcal{R}\right) \tag{23}$$

Next we go along the lines of proof of Theorem 5 in Jansson and Moreira (2004). Define $M_L\left(\beta_0\right)$ to be the class of locally asymptotically median unbiased estimators $m_L\left(\cdot,\beta_0\right)$ satisfying

$$E_{\beta_0,1}^{Y,X}\left[\left(1\left\{m_L\left(\mathcal{R},\beta_0\right)\leq\beta\right\}-0.5\right)f\left(\mathcal{R}_{\beta\beta},\mathcal{R}_{\gamma\gamma}\right)e^{\Lambda(b,c)}\right]=0\qquad\forall f\in C_b\left(\mathbb{R}^2\right),$$

where $\Lambda\left(b,c\right)$ denotes the weak limit of the likelihood ratio $dP^{Y,X}_{\beta_T(b),\gamma_T(c)}/dP^{Y,X}_{\beta_0,1}$. By construction $m_L^*\left(\cdot,\beta_0\right)\in M_L\left(\beta_0\right)$. Applying arguments of Theorem 4 in the present context we have

$$E_{\beta_0,1}^{Y,X}\left[L\left(m_L^*\left(\mathcal{R},\beta_0\right),\beta\right)\Lambda\left(b,c\right)\right]\leq E_{\beta_0,1}^{Y,X}\left[L\left(m_L\left(\mathcal{R},\beta_0\right),\beta\right)\Lambda\left(b,c\right)\right]\qquad\forall m_L\in M_L.$$

From (23), Le Cam's third lemma (for convergence of the likelihood ratio) and Billingsley (1999, Theorem 3.5) we find that $m_L^*\left(\cdot,\beta_0\right)$ satisfies (12) and

$$\lim_{T\to\infty}E^{Y,X}_{\beta_T(b),\gamma_T(c)}\left[L\left(m_L^*\left(R,\beta_0\right),\beta\right)\right]=E_{\beta_0,1}^{Y,X}\left[L\left(m_L^*\left(\mathcal{R},\beta_0\right),\beta\right)\Lambda\left(b,c\right)\right].$$

Also, from Proof to Theorem 5 in Jansson and Moreira (2004) (their arguments apply here since $m_T\left(\cdot\right)=O_p\left(1\right)$ and $L\left(\cdot,\beta\right)$ has at most countably many discontinuities) we know that for any $\left\{m_T\left(\cdot\right)\right\}$ satisfying (12) there exists an $m_L\in M_L$ such that

$$\underline{\lim}_{T\to\infty}E^{Y,X}_{\beta_T(b),\gamma_T(c)}\left[L\left(m_T\left(R\right),\beta\right)\right]=E_{\beta_0,1}^{Y,X}\left[L\left(m_L\left(\mathcal{R}\right),\beta\right)\Lambda\left(b,c\right)\right].$$

Since β_0 is fixed these local arguments generalize to the global estimator $m^*\left(\cdot,\beta_0\right)$.

$\square$

Proof of Theorem 3. Convergence result $\hat{R} \rightarrow_d \mathcal{R}$ follows from Theorem 6 in Jansson and Moreira (2004). Then Theorem 3 follows from continuity of L (except for sets of measure zero) and $m_L^* (\cdot, \beta_0)$ (which holds by continuity of its inverse, $med_A (b, \cdot)$) and CMT. □

Proof of Theorem 4. We first note that, in addition to probability measures ξ and ν_{s_C} specified in the text (in (7) and (8)), we learn from Lemma 2.7.8 in Lehmann (1997) that there exist measures ζ and $v_{s_{\beta\beta}, s_{\gamma\gamma}}$, such that

$$\begin{aligned} dP_{\beta,\gamma}^{S_{\beta\beta}, S_{\gamma\gamma}} \left(s_{\beta\beta}, s_{\gamma\gamma}\right) &= \exp\left[-\frac{1}{2}\left(\beta - \sigma_{xx}^{-1}\sigma_{xy}\gamma\right)^2 s_{\beta\beta} - \frac{1}{2}\gamma^2 s_{\gamma\gamma}\right] d\zeta\left(s_{\beta\beta}, s_{\gamma\gamma}\right), \\ dP_{\beta,\gamma}^{S_C|s_{\beta\beta}, s_{\gamma\gamma}} \left(s_\gamma\right) &= \exp\left(\gamma s_\gamma\right) dv_{s_{\beta\beta}, s_{\gamma\gamma}}\left(s_\gamma\right). \end{aligned} \tag{24}$$

For later use we note that $d\nu_{s_C}(t)$,that enters (8), denotes $\exp\left(-\beta_0 s_\beta\right) dP_{\beta_0,\gamma_0}^{S|s_C}\left(s_\beta\right)$, where $(\beta_0, \gamma_0) \in \mathbb{R}^2$. Also, for any measurable functions f_1 and f_2, we have:

$$P_{\beta,\gamma}^{S_C}\left\{f_1\left(s_C\right) \in A\right\} = \int P_{\beta,\gamma}^{S_C|s_{\beta\beta}, s_{\gamma\gamma}}\left\{f_1\left(s_C\right) \in A\right\} dP_{\beta,\gamma}^{S_{\beta\beta}, S_{\gamma\gamma}}\left(s_{\gamma\gamma}\right), \tag{25}$$

$$P_{\beta,\gamma}^{S}\left\{f_2\left(s\right) \in A\right\} = \int P_{\beta}^{S|s_C}\left\{f_2\left(s\right) \in A\right\} dP_{\beta,\gamma}^{S_C}\left(s_C\right), \tag{26}$$

where A is any Borel set on the real line. We first verify continuity of the cumulative distribution function

$$F_\beta^{S|s_C}(u) = \int_{-\infty}^{u} \exp\left(\beta s_\beta\right) d\nu_{s_C}\left(s_\beta\right).$$

Continuity in β, for any fixed u and γ, follows from Theorem 2.7.9 (i) in Lehmann (1997). With β fixed, on the other hand, we have:

$$\lim_{u_2 \downarrow u_1} \int_{u_1}^{u_2} \exp\left(\beta s_\beta\right) d\nu_{s_C}\left(s_\beta\right) = 0$$

since density ν_{s_C}, which is a conditional density in the joint exponential family as specified above, is uniformly bounded. Continuity in u follows.

From Corollary 3.5.3 (ii) in Lehmann (1997) we learn that (because of continuity and monotonicity) there exist lower $\left(m_T^l\right)$ and upper $\left(m_T^u\right)$ uniformly most accurate unbiased confidence bounds for β at, respectively, $1-\alpha_l$ and $1-\alpha_u$ confidence levels (we take $\alpha_l+\alpha_u \leq 1$). That is:

$$P_\beta^{S|s_C}\left\{m_T^l \leq \beta'\right\} \leq P_\beta^{S|s_C}\left\{\tilde{m}_T^l \leq \beta'\right\} \qquad \forall \beta' \leq \beta,$$

where $\tilde{m}_T^l$ is any other lower confidence bound, and similarly:

$$P_\beta^{S|s_C}\left\{m_T^u \geq \beta'\right\} \leq P_\beta^{S|s_C}\left\{\tilde{m}_T^u \geq \beta'\right\} \qquad \forall \beta' \geq \beta$$

for the uniformly most accurate upper bound, where $\tilde{m}_T^u$ denotes any other upper bound. Consider the lower bound for a moment. Lehmann (1997, Problem 3.21) shows that m_T^l will minimize $E_\beta^{S|s_C}\left[L_l\left(m_T^l,\beta\right)\right]$ at its level α_l for any function L_l that is nonincreasing in m_T^l for $m_T^l < \beta$ and 0 for $m_T^l \geq \beta$. The argument goes as follows. First, define two cumulative distribution functions $F^\dagger$ and $F^\ddagger$ by:

$$F^\dagger(u) = \frac{P_\beta^{S|s_C}\left\{m_T^l \leq u\right\}}{P_\beta^{S|s_C}\left\{\tilde{m}_T^l \leq \beta\right\}},\ F^\ddagger(u) = \frac{P_\beta^{S|s_C}\left\{\tilde{m}_T^l \leq u\right\}}{P_\beta^{S|s_C}\left\{\tilde{m}_T^l \leq \beta\right\}},\ u < \beta$$

and $F^\dagger(u) = F^\ddagger(u) = 1$ for $u \geq \beta$. Then clearly $F^\dagger(u) \leq F^\ddagger(u)$, $\forall u$ and hence $E_\beta^\dagger[f(U)] \leq E_\beta^\ddagger[f(U)]$ for any nonincreasing function f, where notation of the expectation operators relates to the corresponding cdfs. In particular, we have:

$$\begin{aligned} E_\beta^{S|s_C}\left[L_l\left(m_T^l,\beta\right)\right] &= P_\beta^{S|s_C}\left\{\tilde{m}_T^l \leq \beta\right\}\int L_l(u,\beta)\,dF^\dagger(u) \\ &\leq P_\beta^{S|s_C}\left\{\tilde{m}_T^l \leq \beta\right\}\int L_l(u,\beta)\,dF^\ddagger(u) = E_\beta^{S|s_C}\left[L_l\left(\tilde{m}_T^l,\beta\right)\right]. \end{aligned}$$

Similar argument shows that m_T^u will minimize $E_\beta^{S|s_C}\left[L_u\left(m_T^u,\beta\right)\right]$ for any L_u that is nondecreasing in m_T^u for $m_T^u > \beta$ and 0 for $m_T^u \leq \beta$.

Until now we have considered confidence intervals constructed conditionally on $S_C = s_C$. Note, however, that by measurability of $F_\beta^{S|s_C}(\cdot)$ over the Borel sets induced by S we can fuse (26) and (16)-(17) to find:

$$\begin{aligned} F_\beta^S(m_T^l) &= \int P_\beta^{S|s_C}\{m_T^l \leq \beta\}\,dP_{\beta,\gamma}^{S_C}(s_C) = \alpha_l, \\ F_\beta^S(m_T^u) &= \int P_\beta^{S|s_C}\{m_T^u \leq \beta\}\,dP_{\beta,\gamma}^{S_C}(s_C) = 1-\alpha_u, \end{aligned}$$

for all $\beta \in \mathbb{R}$. Hence $C_T^* = \left(m_T^l, m_T^u\right)$ retains its nominal level unconditionally.

In what follows we extend the conditional optimality statement to the class of confidence intervals that are conditionally unbiased at level $1 - \alpha_l - \alpha_u$, where conditioning refers to $\left(S_{\beta\beta}, S_{\gamma\gamma}\right) = \left(s_{\beta\beta}, s_{\gamma\gamma}\right)$. First, for any $\left(S_{\beta\beta}, S_{\gamma\gamma}\right) = \left(s_{\beta\beta}, s_{\gamma\gamma}\right)$, we take an arbitrary confidence set C_T of the specified level and show that it retains its level on the partition $S_C = s_C$ and hence, because of conditional optimality of C_T^*, is inferior on that partition. Optimality follows by integration with respect to $P_{\beta,\gamma}^{S_C|s_{\beta\beta},s_{\gamma\gamma}}$.

First, for any confidence set $C_T\left(y, x\right)$ of level $1 - \alpha_l - \alpha_u$ we have:

$$P_{\beta,\gamma}^{Y,X}\left\{\beta \in C_T\left(y,x\right)\right\} = 1 - \alpha_l - \alpha_u \qquad \forall \left(\beta, \gamma\right) \in \mathbb{R}^2. \tag{27}$$

Since conditioning on sufficient statistics leaves the resulting function independent of parameters, we can rewrite this random interval as a function of sufficient statistics, say $C_T\left(s\right)$. Specifically, if we define $1_A\left(z\right) = 1$ if $z \in A$ and 0 otherwise, we have:

$$\begin{aligned}
P_{\beta,\gamma}^{Y,X}\left\{\beta \in C_T\left(y,x\right)\right\} &= \int P^{Y,X|s}\left\{\beta \in C_T\left(y,x\right)\right\} dP_{\beta,\gamma}^S\left(s\right) \\
&= \int 1_{\left\{\beta \in C_T(y,x)\right\}}\left(y,x\right) dP_{\beta,\gamma}^S\left(s\right) \\
&= P_{\beta,\gamma}^S\left\{\beta \in C_T\left(y,x\right)\right\},
\end{aligned}$$

where the first transformation is based on the law of iterated expectations, in the spirit of (25) and (26), and the second comes from the fact that the conditional probability inside the integral is independent of either γ or β. Finally, since $dP_{\beta,\gamma}^S$ is generated by S, we can rewrite the arbitrary set $C_T\left(y,x\right)$ as a function of s, $C_T\left(s\right)$, without altering the probability statement. Hence the property (27) can alternatively be represented in the space of sufficient statistics as

$$P_{\beta,\gamma}^S\left\{\beta \in C_T\left(s\right)\right\} = 1 - \alpha_l - \alpha_u \qquad \forall \left(\beta, \gamma\right) \in \mathbb{R}^2.$$

Using (25) and (26) we can write:

$$\int \left[\int P_{\beta}^{S|s_C} \left\{ \beta \in C_T(s) \right\} dP_{\beta,\gamma}^{S_C|s_{\beta\beta},s_{\gamma\gamma}}(s_\gamma) \right] dP_{\beta,\gamma}^{S_{\beta\beta},S_{\gamma\gamma}}\left(s_{\beta\beta},s_{\gamma\gamma}\right) = P_{\beta,\gamma}^{S}\left\{ \beta \in C_T(s) \right\}.$$

Since we are conditioning on $\left(S_{\beta\beta},S_{\gamma\gamma}\right) = \left(s_{\beta\beta},s_{\gamma\gamma}\right)$, we must have:

$$\int P_{\beta}^{S|s_C}\left\{ \beta \in C_T(s) \right\} dP_{\beta,\gamma}^{S_C|s_{\beta\beta},s_{\gamma\gamma}}(s_\gamma) = 1-\alpha_l-\alpha_u \qquad \forall\, (\beta,\gamma) \in \mathbb{R}^2,$$

which, by completeness of the family $P_{\beta,\gamma}^{S_C|s_{\beta\beta},s_{\gamma\gamma}}$ (which follows from (24) and Lehmann (1997, Thm. 4.3.1.)), further reduces to:

$$P_{\beta}^{S|s_C}\left\{ \beta \in C_T(s) \right\} = 1-\alpha_l-\alpha_u \qquad a.e.\ s_\gamma \in \mathbb{R},$$

with $\left(s_{\beta\beta},s_{\gamma\gamma}\right)$ fixed, which means that $C_T(s)$ is unbiased at level $1-\alpha_l-\alpha_u$ on the partition $S_C = s_C$, for all $\beta \in \mathbb{R}$. Let N_β denote the exceptional null set in the last display, and $N = \cup N_\beta$. Then for any s_C such that $s_\gamma \notin N$ and for any $\beta \in \mathbb{R}$ we have:

$$P_{\beta}^{S|s_C}\left\{ \beta \in C_T^*(s) \right\} = P_{\beta}^{S|s_C}\left\{ \beta \in C_T(s) \right\}$$

since both have level $1-\alpha_l-\alpha_u$ on the partition $S_C = s_C$. But we know it is $C_T^*(s)$ that is optimal on this partition, in a sense that:

$$E_{\beta}^{S|s_C}\left[L_{CI}\left(C_T^*(s),\beta\right)\right] \leq E_{\beta}^{S|s_C}\left[L_{CI}\left(C_T(s),\beta\right)\right].$$

This argument, we may recall, rests on C_T^* being the uniformly most accurate unbiased confidence interval for any $\beta \in \mathbb{R}$. Since N is a $P_{\beta,\gamma}^{S_C} - \textit{null set}$ we can integrate this inequality with respect to $P_{\beta,\gamma}^{S_C|s_{\beta\beta},s_{\gamma\gamma}}$ and use (25)-(26) to conclude

$$E_{\beta,\gamma}^{S|s_{\beta\beta},s_{\gamma\gamma}}\left[L_{CI}\left(C_T^*(s),\beta\right)\right] \leq E_{\beta,\gamma}^{S|s_{\beta\beta},s_{\gamma\gamma}}\left[L_{CI}\left(C_T(s),\beta\right)\right],$$

which is equivalent to the statement of the theorem. $\square$

REFERENCES

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59(3), 817–858.

BILLINGSLEY, P. (1999): *Convergence of Probability Measures*. Wiley, New York, second edn.

BROWN, L. D. (1990): "The 1985 Wald Memorial Lectures. An Ancillarity Paradox which Appears in Multiple Linear Regression," *The Annals of Statistics*, 18(2), 471–538, With discussion and a reply by the author.

BROWN, L. D., A. COHEN, AND W. E. STRAWDERMAN (1976): "A Complete Class Theorem for Strict Monotone Likelihood Ratio with Applications," *The Annals of Statistics*, 4(4), 712–722.

CAMPBELL, J. Y., AND M. YOGO (2003): "Efficient Tests of Stock Return Predicatbility," NBER Working Paper 10026.

CAVANAGH, C. L., G. ELLIOTT, AND J. H. STOCK (1995): "Inference in Models with Nearly Integrated Regressors," *Econometric Theory*, 11(5), 1131–1147.

COX, D. D., AND I. LLATAS (1991): "Maximum Likelihood Type Estimation for Nearly Autoregressive Time Series," *Annals of Statistics*, 19, 1109–1128.

ELLIOTT, G. (1998): "On the Robustness of Cointegration Methods when Regressors Almost have Unit Roots," *Econometrica*, 66(1), 149–158.

ELLIOTT, G., I. KOMUNJER, AND A. TIMMERMANN (2004): "Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?," UC San Diego.

ELLIOTT, G., T. J. ROTHENBERG, AND J. H. STOCK (1996): "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64(4), 813–836.

ELLIOTT, G., AND J. H. STOCK (2001): "Confidence Intervals for Autoregressive Coefficients Near One," *Journal of Econometrics*, 103(1-2), 155–181.

JANSSON, M., AND M. J. MOREIRA (2004): "Optimal Inference in Regression Models with Nearly Integrated Regressors," UC Berkeley and Harvard University.

JEGANATHAN, P. (1995): "Some Aspects of Asymptotic Theory with Applications to Time Series Models," *Econometric Theory*, 11, 818–887.

——— (1997): "On Asymptotic Inference in Linear Cointegrated Time Series Systems," *Econometric Theory*, 13(5), 692–745.

JOHANSEN, S. (1988): "Statistical Analysis of Cointegrating Vectors," *Journal of Economic Dynamics & Control*, 12(2-3), 231–254.

LANNE, M. (2002): "Testing the Predictability of Stock Returns," *The Review of Economics and Statistics*, 84(3), 407–415.

LE CAM, L., AND G. L. YANG (2000): *Asymptotics in Statistics: Some Basic Concepts*. Springer-Verlag, New York, second edn.

LEHMANN, E. L. (1997): *Testing Statistical Hypotheses*, Springer Texts in Statistics. Springer-Verlag, New York, second edn.

LEHMANN, E. L., AND H. SCHEFFÉ (1950): "Completeness, Similar Regions, and Unbiased Estimation. I," *Sankhyā*, 10, 305–340.

PFANZAGL, J. (1979): "On Optimal Median Unbiased Estimators in the Presence of Nuisance Parameters," *The Annals of Statistics*, 7(1), 187–193.

——— (1985): *Asymptotic Expansions for General Statistical Models*, vol. 31 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, With the assistance of W. Wefelmeyer.

PHILLIPS, P. C. B. (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.

PHILLIPS, P. C. B., AND M. LORETAN (1991): "Estimating Long-Run Economic Equilibria," *Review of Economic Studies*, 58(3), 407–436.

PLOBERGER, W. (2004): "A Complete Class of Tests when the Likelihood is Locally Asymptotically Quadratic," *Journal of Econometrics*, 118(1-2), 67–94.

POLK, C., S. THOMPSON, AND T. VUOLTEENAHO (2004): "Cross-sectional Forecasts of the Equity Premium," Harvard University.

SAIKKONEN, P. (1991): "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory*, 7(1), 1–21.

SHAFFER, J. P. (1991): "The Gauss-Markov Theorem and Random Regressors," *The American Statistician*, 45(4), 269–273.

SHIRYAEV, A. N., AND V. G. SPOKOINY (2000): *Statistical Experiments and Decisions*. World Scientific Publishing Co. Inc., River Edge, NJ.

STOCK, J. H., AND M. W. WATSON (1993): "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61(4), 783–820.

——— (1996): "Confidence Sets in Regressions with Highly Serially Correlated Regressors," Unpublished.

THOMPSON, S. B. (2003): "Optimal Versus Robust Inference in Nearly Integrated non Gaussian Models," *Econometric Theory*.

TOROUS, W., R. VALKANOV, AND S. YAN (2001): "On Predicting Stock Returns with Nearly Integrated Explanatory Variables," UCLA.

DEPARTMENT OF ECONOMICS, PRINCETON UNIVERSITY

E-mail address: peliasz@princeton.edu