Optimal Instruments in Time Series: A Survey

Stanislav Anatolyev*
New Economic School
Nakhimovsky Prospekt, 47, Moscow, 117418 Russia
E-mail: sanatoly@nes.ru

*This version: October 2005. Address: Stanislav Anatolyev, New Economic School, Nakhimovsky Prospekt, 47, Moscow, 117418 Russia. E-mail: sanatoly@nes.ru. I would like to thank the editor and two referees whose comments helped greatly improve the presentation.

Optimal Instruments in Time Series: A Survey

Abstract

This article surveys estimation in stationary time series models using the approach of optimal instrumentation. We review tools that allow construction and implementation of optimal instrumental variables estimators in various circumstances – in single- and multiperiod models, in the absence and presence of conditional heteroskedasticity, by considering linear and nonlinear instruments. We also discuss issues adjacent to the theme of optimal instruments. The article is directed primarily towards practitioners, but also may be found useful by econometric theorists and teachers of graduate econometrics.

Keywords

Instrumental variables estimation; Moment restrictions; Optimal instrument; Efficiency bounds; Stationary time series.

1 Introduction

In stationary time series models, achieving consistent estimation is rarely a difficult issue. As a rule, there is abundance of valid instrumental variables (IV) that may be used for such estimation; these are usually taken from the set of historical variables and possibly their functions. Achieving asymptotically most efficient estimation, however, is a more challenging task, given that the set of valid instruments to choose from is typically infinite, and even "double-infinite" both spreading into an infinite past and embracing all nonlinear functions. In this survey, we gather recent results pertaining to efficient estimation by instrumental variables of time series models that are characterized by an infinite set of potential instruments, with a hope that the appropriate methods will gain more popularity among applied econometricians. After all, even though achieving more efficiency may be effortful and may come at certain computational cost in most complex situations, consistency of estimation is not lost, and a practitioner does not risk anything, so why not trying it?

Applied time series econometricians are nowadays accustomed[1] to using the generalized method of moments (GMM) of Hansen (1982), which can be recast as instrumental variables estimation, and more often think of ways to use GMM most optimally. The available surveys devoted to GMM in time series only marginally discuss the issue of optimal instrumentation (see, for example, Harris, 1999; Hansen L.P., 2002; Hansen B. and West, 2002)[2]. Section 7.2 in the recent manuscript Hall (2005) handles this topic, but it does not contain most recent results, particularly applicable for the general case characterized by serial correlation and conditional heteroskedasticity. The present review fills this gap. As it is directed primarily towards practitioners, the narration is least rigorous as it can probably be, given the technicality of this topic. To facilitate comprehensibility, in Section 2 we list several examples from applied econometric practice where the idea of optimal instrumentation may be implemented, while in Section 11 we illustrate some of techniques using a macroeconomic example and real data. We also want to emphasize that the approach being surveyed is *constructing* instruments that are *a priori* known to be optimal, rather than *picking* or *selecting* instruments from a list, which is, of course, a valid and interesting topic on its own (we list several articles representing the latter research direction at the end of this survey). For simplicity and transparency, linear single equation models are treated throughout.

We start in Section 3 by considering optimal instrumentation for unconditional moment restrictions. The problems considered here are theoretically, rather than empirically, motivated; this material is included in order to introduce certain concepts in a simplest framework. In contrast, Sections 4 through 9 are devoted to conditional moment restrictions that do arise in practice. In Section 4 we introduce the *optimality condition* derived

[1] This is clearly evidenced by the *JBES* 20th anniversary issue on the generalized method of moments.

[2] Lars Peter Hansen eloquently described the evolution of his ideas on efficiency bounds in a *JBES* interview (Ghysels and Hall, 2002).

in Hansen (1985), the key tool to finding the form of the optimal instrument in various circumstances. In Section 5 we consider the case of *single-period* moment restrictions, where the moment function becomes in the information set in the next period (i.e., is a martingale difference) and hence is serially uncorrelated. In a sense, this case is analogous to that in an IID environment, as the form of the optimal instrument is the same *static* function of certain conditional expectations, *viz.* the conditional expectation of the gradient of the moment function predivided by its conditional variance matrix (cf. Chamberlain, 1987), and further simplifies in a conditionally homoskedastic case (Amemiya, 1974). A variety of intertemporal macroeconomic and financial models instead give rise to *multiperiod* moment restrictions, where the moment function is serially correlated to a finite known order. The GMM estimation procedure in these situations does not change dramatically, but the conditions for an instrument to be optimal become more complicated. In Sections 6 and 7 we review the cases of conditional homoskedasticity and conditional heteroskedasticity, respectively. Under conditional homoskedasticity, Hansen (1985) found that the optimal instrument takes a *dynamic* form of a certain recursive relation. Under conditional heteroskedasticity, Hansen, Heaton and Ogaki (1988) presented an elegant characterization of the efficiency bound (i.e., a minimal attainable asymptotic variance). This theory leads to the optimal instrument whose form described in Anatolyev (2003b) has an additional difficulty that its dynamics is implicitly defined. In Section 9, we review an alternative route in the optimal IV literature where one artificially contracts the set of allowable instruments thus sacrificing efficiency in order to attain the efficiency bound relative to this narrowed set of instruments. The motivation behind such a seemingly strange move is relative easiness of attaining the new efficiency bound. In Section 10 we make numerical comparisons of asymptotic efficiency for optimal in various senses IV estimators as well as some of their competitors in a simple example, in order to get a feel of efficiency–complexity trade-offs.

Throughout, we also discuss ways of implementing the optimal IV estimators, in most complex cases having to adapt approximations. Apart from a complicatedness of the form of the optimal instrument, this step intrinsically contains other difficulties specific to the time series context. While in cross-sections the state vector is typically finite dimensional, here it may contain all infinite past, which makes nonparametric estimation of the needed conditional expectations more problematic. An alternative direction entails making auxiliary parameterizations, and may be viewed as non-comforting by an econometric theorist; the reasoning against this direction is a resulting reduction of robustness of the GMM approach. A practitioner, however, may be comfortable with fully or partially parameterizing the auxiliary processes (e.g., the conditional variance by a parametric GARCH), with an understanding that the efficiency may be lost and that more robust inference methods are needed ("sandwich" estimation of asymptotic variance, etc.). Unfortunately, there are few examples of application in empirical work, and the guidance to feasible implementation provided by the theoretical literature is rather limited. However, we devote

separate Section 8 to a discussion of issues of estimation of conditional expectations listing helpful references to a recent literature on nonparametric estimation.

Last, but not least, in Section 12 we briefly discuss issues adjacent to the theme of optimal instrumentation in order to provide a more complete characterization of the place of this interesting topic in a broader context of econometric theory and practice. This includes, among other aspects, applicability of similar ideas to panel data and nonstationary time series, and a connection to notions of instrument redundancy, semiparametric efficiency bounds, and empirical likelihood estimation.

2 Econometric examples

To begin with, we list several examples from applied econometric practice where the idea of optimal instrumentation may be used. Typically, the conditioning set is infinite, and contains the entire history of one or more variables. The question is how to optimally exploit the information in this wide set of potential instruments.

Example 1 *When one stationary variable, say x_t, is used to forecast another, say y_t, testing for forecast unbiasedness or predictive ability leads to the moment restriction*

$$E\left[y_{t+q} - \alpha - \beta x_t \,|\, x_t, y_t, x_{t-1}, y_{t-1}, \cdots\right] = 0.$$

In Fama (1975), x_t is interest rate and y_t is inflation; in Mishkin (1990), x_t is slope of term structure of nominal interest rates and y_t is either real interest rate or change in inflation; in Hansen and Hodrick (1980), x_t is forward premium and y_t is exchange rate depreciation.

Example 2 *If AR parameters in a semi-strong ARMA(p,q − 1) model are estimated by instrumental variables, the conditional moment restriction*

$$E\left[y_t - \alpha_0 - \alpha_1 y_{t-1} - \cdots - \alpha_p y_{t-p} \,|\, y_{t-q}, y_{t-q-1}, \cdots\right] = 0$$

is estimated. If the ARMA model is weak, it is the system of unconditional moment restrictions

$$E\left[\left(y_t - \alpha_0 - \alpha_1 y_{t-1} - \cdots - \alpha_p y_{t-p}\right)\left(y_{t-q}, y_{t-q-1}, \cdots\right)'\right] = 0$$

that is estimated.

Example 3 *Meddahi and Renault (2004) showed that the class of square-root stochastic autoregressive volatility (SR–SARV) models, a natural extension of weak GARCH models, is characterized by the conditional moment restriction*

$$E\left[\varepsilon_t^2 - \omega - \sum_{j=1}^{q-1} \gamma_i \varepsilon_{t-j}^2 \,|\, \varepsilon_{t-q}, \varepsilon_{t-q-1}, \cdots\right] = 0,$$

where ε_t is observable stationary SR–SARV(q − 1) process (for details, see Meddahi and Renault, 2004).

Example 4 *The Hansen and Singleton (1982) consumption-based CAPM with assets maturing in q periods leads to the moment restriction*

$$E\left[\delta^q \prod_{j=1}^{q}(1+r_{t+j})\left(\frac{c_{t+q}}{c_t}\right)^{-\gamma}-1 \mid \frac{c_t}{c_{t-1}}, r_t, \frac{c_{t-1}}{c_{t-2}}, r_{t-1}, \cdots\right]=0,$$

where c_t is a consumption, and r_t is a market return; δ is a discount factor, and γ is a coefficient of risk aversion. More complicated moment restrictions corresponding to $q=1$ are implied by models in Mark (1985), Mankiw, Rotemberg, Summers (1985), Hotz, Kydland, Sedlacek (1988), Kocherlakota (1990), Epstein and Zin (1991), Marshall (1992), Holman (1998), Smith (1999), Weber (2000), among others.

Example 5 *The Ferson and Constantinides (1991) consumption-based CAPM with habit formation leads to the moment restriction*

$$E\left[\delta\left(1+r_{t+1}\right)\left(\left(\frac{s_{t+1}}{s_t}\right)^{-\gamma}+\alpha\delta\left(\frac{s_{t+2}}{s_t}\right)^{-\gamma}\right)\right.$$
$$\left.-\left(1+\alpha\delta\left(\frac{s_{t+1}}{s_t}\right)^{-\gamma}\right) \mid \frac{c_t}{c_{t-1}}, r_t, \frac{c_{t-1}}{c_{t-2}}, r_{t-1}, \cdots\right]=0,$$

where $s_t=c_t+\alpha c_{t-1}$, c_t is a consumption, and r_t is a market return; α is a habit formation/durability parameter, δ is a discount factor, and γ is a coefficient of risk aversion. Similar examples are the models in Dunn and Singleton (1986), Eichenbaum, Hansen and Singleton (1988) and Weber (2002).

Example 6 *The Hansen and Singleton (1996) temporal aggregation model leads to a system of conditional moment restrictions*

$$E\left[\begin{pmatrix} u_{t+2} \\ u_{t+2}^2-2\sigma^2/3 \\ u_{t+2}^2/4-u_{t+2}u_{t+1} \end{pmatrix} \mid c_t, q_t, c_{t-1}, q_{t-1}, \cdots\right]=0,$$

where $u_{t+1}=\gamma\left(c_{t+1}-c_t\right)+\left(q_{t+1}-q_t\right)-\left(\delta-\sigma^2/2\right)$, c_t and q_t are consumption and asset price, γ and δ are preference parameters, and σ^2 is a variance measure of underlying Brownian motions. Simpler versions of this model were formulated earlier in Hall (1988) and Grossman, Melino and Shiller (1987).

Example 7 *The Meddahi, Renault and Werker (2003) model for ultra-high-frequency returns considered in more detail in Section 11 corresponds to the moment restriction*

$$E\left[\left(r_t^2-\theta\right)-\left(r_{t-1}^2-\theta\right)\exp\left(-\kappa d_{t-1}\right)\frac{c\left(\kappa d_t\right)}{c\left(\kappa d_{t-1}\right)} \mid r_{t-2}, d_{t-2}, r_{t-3}, d_{t-3}, \cdots\right]=0,$$

where d_t is a duration between successive trades, r_t is a scaled return from the latter of these trades, and $c(v)\equiv\left(1-\exp(-v)\right)/v$.

As can be seen, most of practical situations give rise to conditional restrictions rather than unconditional ones. However, to introduce main notions and terminology in a simplest context, we start by considering unconditional moment restrictions. Also, for the sake of simplicity we concentrate on a linear model. When the moment function is nonlinear in parameters as in many examples above, the role of right side variables ("regressors") is played by a quasi-regressor, the derivative of the moment function with respect to parameters. Since this quasi-regressor depends on parameters, usually an additional preliminary step of finding its consistent estimate is needed. Otherwise, the same tools can be applied to nonlinear models.

3 Unconditional problems

Consider a linear IV model with instruments that are uncorrelated with disturbances:

$$y_t = x_t'\beta + e_t, \quad E\left[z_t e_t\right] = 0, \tag{1}$$

where x_t is a $k \times 1$ vector of right-hand-side variables, z_t is an $\ell \times 1$ vector of *basic instruments*, and all variables are strictly stationary. Let $\ell \geq k$ (possibly, $\ell = \infty$ if emerging infinite-dimensional matrices are well-defined) and $Q_{xz} = E\left[x_t z_t'\right]$ have rank k. A (not necessarily efficient) GMM estimator of β indexed by an arbitrary data-dependent symmetric positive definite weight matrix $\hat{W}$ consistent for its population analog W, is[3]

$$\begin{aligned}\hat{\beta}_\varsigma &= \arg\min_b \left[\sum_t z_t\left(y_t - x_t'b\right)\right]' \hat{W} \left[\sum_t z_t\left(y_t - x_t'b\right)\right] \\ &= \left(\sum_t \hat{\varsigma}_t x_t'\right)^{-1} \sum_t \hat{\varsigma}_t y_t,\end{aligned}$$

where $\hat{\varsigma}_t = \hat{Q}_{xz}\hat{W}z_t$ and $\hat{Q}_{xz} = T^{-1}\sum_t x_t z_t'$. The estimator $\hat{\beta}_\varsigma$ is a just-identified IV estimator that uses $\hat{\varsigma}_t$ as an instrument, and is asymptotically equivalent to one that would use as an instrument the (usually non-feasible) population analog $\varsigma_t = Q_{xz}Wz_t$. Here premultiplication of the $\ell \times 1$ basic instrument z_t by $Q_{xz}W$ converts it to a just-identifying $k \times 1$ one. Note that $Q_{xz}W$ is as arbitrary as W is.

It is well known that the efficient GMM corresponds to the weight matrix $W = V_{ze}^{-1}$, where (Hansen, 1982)

$$V_{ze} \equiv \lim_{T\to\infty}\left[\frac{1}{T} var\left(\sum_t z_t e_t\right)\right] = \sum_{j=-\infty}^{+\infty} cov\left(z_t e_t, z_{t-j}e_{t-j}\right).$$

It follows that the best choice for a just-identifying instrument is $\zeta_t = Q_{xz}V_{ze}^{-1}z_t$, or its asymptotically equivalent feasible version $\hat{\zeta}_t = \hat{Q}_{xz}\hat{V}_{ze}^{-1}z_t$, where $\hat{V}_{ze}$ is a consistent

[3]Unless otherwise designated, the index of summation signs is t that runs from the beginning to the end of the sample whose length equals T. All sums are truncated to exclude components outside the sample.

estimate of V_{ze}. We call ζ_t the *optimal instrument* in the *class of allowable instruments* $\mathcal{Z}_t$, which in the present case are k linearly independent linear combinations of the basic instrument z_t. Formally,

$$\mathcal{Z}_t = \left\{ \varsigma_t : \ \varsigma_t = Az_t \quad \text{for some } k \times \ell \text{ matrix } A \right\}.$$

Remark 3.1 Premultiplication of an instrument by an arbitrary non-singular conformable constant matrix does not change an associated IV estimator. Therefore, the optimal instrument is defined up to such premultiplication. In the special case of exact identification ($\ell = k$), $Q_{xz}V_{ze}^{-1}$ is $k \times k$ and thus can be removed, so the basic instrument is trivially optimal.

Remark 3.2 Note that for convenience it is not required that an allowable instrument be relevant. Instead, irrelevant instruments are associated with infinite asymptotic variances.

Remark 3.3 The optimal instrument ζ_t satisfies the *optimality condition*

$$E\left[x_t\varsigma_t'\right] = \sum_{j=-\infty}^{+\infty} cov\left(\zeta_t e_t, \varsigma_{t-j}e_{t-j}\right) \quad \text{for all } \varsigma_t \in \mathcal{Z}_t.$$

This condition will be of great help in more complex situations and will be properly introduced in due time. The asymptotic variance of the IV estimator that employs the optimal instrument is *the efficiency bound* relative to $\mathcal{Z}_t$

$$V_{\hat{\beta}_\zeta} = Q_{\zeta x}^{-1}V_{\zeta e}Q_{\zeta x}'^{-1} = \left(Q_{xz}V_{ze}^{-1}Q_{xz}'\right)^{-1}. \tag{2}$$

Remark 3.4 It is easy to see that the optimal instrument $\zeta_t = Q_{xz}V_{ze}^{-1}z_t$ is such that any other instrument $\varsigma_t = Az_t$, exactly identifying or overidentifying, is redundant given ζ_t, no matter what A is. This directly follows from the application of Theorem 1 of Breusch, Qian, Schmidt and Wyhowski (1999).

Schematically, the implementation of the idea of optimal instruments is done in the following three distinct steps.

1. Find the *form* of the optimal instrument. In the present situation the form is $\zeta_t = Q_{xz}V_{ze}^{-1}z_t$.

2. Construct a *feasible* optimal instrument. In the present situation it entails consistent estimation of matrices Q_{xz} and V_{ze} and calculation of $\hat{\zeta}_t$ for all t.

3. *Use* the constructed instrument to estimate β and the estimate's asymptotic variance.

4 Conditional problems: optimality condition

Consider a linear IV model with the disturbances that have mean zero conditional on history of instruments:

$$y_t = x_t'\beta + e_t, \quad E\left[e_t|\Im_t\right] = 0, \tag{3}$$

where x_t is a $k \times 1$ vector of right-hand-side variables, $\Im_t \equiv \sigma\left(z_t, z_{t-1}, \ldots\right)$, where z_t is an $\ell \times 1$ vector of basic instruments[4], and all variables are strictly stationary. In a typical macroeconomic application, z_t includes x_t and a lag of y_t of certain order. When the order of this lag is $q = 1$, the restriction (3) is *single-period*, otherwise (when $q > 1$) – *multiperiod*. This terminology is justified by the typically q-step-ahead prediction context of (3) as in example 1 above; see also Hansen and West (2002, Section 2). The ARMA models of example 2 and SR–SARV models of example 3 are also multiperiod unless $q = 1$. A number of models such as those in examples 5, 6 and 7, give rise to two-period moment restrictions.

For the model (3), any stationary $\Im_t$-measurable $k \times 1$ vector ς_t with finite fourth moment is a valid instrument. That is, the broadest class of admissible instruments is

$$\mathcal{Z}_t = \left\{ \varsigma_t : \ \varsigma_t = f\left(z_t, z_{t-1}, \ldots\right) \text{ for some measurable } f \text{ s.t. } E\left[|\varsigma_t|^4\right] < \infty \right\}. \tag{4}$$

Obviously, there is no need to consider instruments that are overidentifying. As follows from the previous Section, the optimal use of such instrument would entail premultiplication by a certain matrix, while this transformation might be embedded in f in the first place. Again, we associate irrelevant instruments with infinite asymptotic variances.

Hansen (1985) derived the optimality condition which may be used to find the optimal instrument. Under suitable conditions placed on the set of instruments $\mathcal{Z}_t$ (not necessarily (4)[5]) the optimal relative to $\mathcal{Z}_t$ instrument ζ_t satisfies the *optimality condition*

$$E\left[x_t\varsigma_t'\right] = \sum_{j=-\infty}^{+\infty} cov\left(\zeta_t e_t, \varsigma_{t-j} e_{t-j}\right) \quad \text{for all } \varsigma_t \in \mathcal{Z}_t. \tag{5}$$

Indeed, for any ς_t define $Q_{\varsigma x} = E\left[\varsigma_t x_t'\right]$ and $V_{\varsigma e} = \sum_{j=-\infty}^{+\infty} cov\left(\varsigma_t e_t, \varsigma_{t-j} e_{t-j}\right)$, then from the inequality

$$\sum_{j=-\infty}^{+\infty} cov\left(Q_{\varsigma x}^{-1}\varsigma_t e_t - V_{\zeta e}^{-1}\zeta_t e_t, Q_{\varsigma x}^{-1}\varsigma_{t-j} e_{t-j} - V_{\zeta e}^{-1}\zeta_{t-j} e_{t-j}\right) \geq 0 \quad \text{for all } \varsigma_t \in \mathcal{Z}_t$$

it follows using (5) that $Q_{\varsigma x}^{-1} V_{\varsigma e} Q_{\varsigma x}'^{-1} \geq V_{\zeta e}^{-1} = Q_{\zeta x}^{-1} V_{\zeta e} Q_{\zeta x}'^{-1}$ for all $\varsigma_t \in \mathcal{Z}_t$, that is, ζ_t is optimal relative to $\mathcal{Z}_t$. In single-period problems the long-run covariance on the right side of (5) is just a short-run covariance; in multiperiod problems with horizon q the

[4]As long as β is identified by some instrument from $\Im_t$, the dimensionality of the basic instrument does not matter for the rest of discussion except for that in Section 9.

[5]The set $\mathcal{Z}_t$ needs to be a linear space; (4) is.

summation index j runs from $-(q-1)$ to $q-1$. The system (5) may be viewed as a first order condition for a problem of minimization of the asymptotic variance $Q_{\varsigma x}^{-1} V_{\varsigma e} Q_{\varsigma x}^{\prime -1}$ with respect to ς_t. Under the condition (5) the variance sandwich collapses, which is an indication of efficiency.

In constructing the optimal instrument ζ_t, the serial correlation structure of the disturbance e_t is critical. When the problem is single-period so that the error has a martingale difference structure, the optimal instrument has Chamberlain's (1987) static form. When the problem is multiperiod so that the error is serially correlated, the optimal instrument has a complicated dynamic structure which can be characterized differently depending on whether the disturbance e_t is homoskedastic or heteroskedastic conditional on $\Im_t$.

5 Single-period conditional problems

When the disturbance term is a martingale difference relative to $\Im_t \vee \sigma\left(e_{t-1}, e_{t-2}, \cdots\right)$, the optimal instrument has the following simple form:

$$\zeta_t = \frac{E\left[x_t | \Im_t\right]}{E\left[e_t^2 | \Im_t\right]}. \tag{6}$$

This form is familiar from the literature on optimal IV estimation in problems with cross-sectional data (Chamberlain, 1987). It is static in the sense that its formula does not explicitly contain lagged values of the optimal instrument, in contrast to dynamic forms for multiperiod restrictions below (which does not preclude the dependence of ζ_t on the whole infinite past). Note that the form (6) easily follows from the optimality condition:

$$\begin{aligned} E\left[x_t \varsigma_t'\right] &= E\left[\zeta_t \varsigma_t' e_t^2\right] = E\left[\zeta_t \varsigma_t' E\left[e_t^2 | \Im_t\right]\right] \quad \text{for all } \varsigma_t \in \mathcal{Z}_t \\ &\Rightarrow \quad E\left[\left(E\left[x_t | \Im_t\right] - \zeta_t E\left[e_t^2 | \Im_t\right]\right) \varsigma_t'\right] = 0 \quad \text{for all } \varsigma_t \in \mathcal{Z}_t \\ &\Rightarrow \quad E\left[x_t | \Im_t\right] - \zeta_t E\left[e_t^2 | \Im_t\right] = 0. \end{aligned}$$

Implementation of the instrument (6) reduces to estimation of the conditional score $E\left[x_t | \Im_t\right]$ and conditional variance $E\left[e_t^2 | \Im_t\right]$.

6 Multiperiod conditional problems with conditional homoskedasticity

Hansen (1985, Lemma 5.7) showed that under conditional homoskedasticity (meaning that the conditional variance and all conditional autocovariances are constant), the optimal instrument follows the recursion

$$\Theta(L)\zeta_t \propto E\left[\Theta(L^{-1})^{-1} x_t | \Im_t\right], \tag{7}$$

where $\Theta(L) = \theta_0 + \theta_1 L + \cdots + \theta_q L^q$ is a qth order moving average polynomial in the Wold representation of e_t:

$$e_t = \Theta(L)\varepsilon_t,$$

where ε_t is a (weak) white noise with unit variance. Let us check that (7) satisfies the optimality condition. The right hand side of (5) is

$$\begin{aligned}
\sum_{j=-q}^{+q} cov\left(\zeta_t e_t, \varsigma_{t-j} e_{t-j}\right) &= E\left[\varsigma_t \sum_{j=-q}^{+q} \zeta_{t-j}\Theta(L)\varepsilon_{t-j} \cdot \Theta(L)\varepsilon_t\right] \\
&= E\left[\varsigma_t \left(\theta_0\varepsilon_t\left(\theta_0\varepsilon_t\zeta_t + \cdots + \theta_q\varepsilon_t\zeta_{t+q}\right)\right.\right. \\
&\qquad +\theta_1\varepsilon_{t-1}\left(\theta_0\varepsilon_{t-1}\zeta_{t-1} + \cdots + \theta_q\varepsilon_{t-1}\zeta_{t+q-1}\right) \\
&\qquad + \cdots \\
&\qquad \left.\left. +\theta_q\varepsilon_{t-q}\left(\theta_0\varepsilon_{t-q}\zeta_{t-q} + \cdots + \theta_q\varepsilon_{t-q}\zeta_t\right)\right)\right] \\
&= E\left[\varsigma_t \cdot \Theta(L)\Theta(L^{-1})\zeta_t\right] = E\left[\varsigma_t \cdot \Theta(L^{-1})\Theta(L)\zeta_t\right].
\end{aligned}$$

Now, with (7), it is proportional to

$$\begin{aligned}
E\left[\varsigma_t \cdot \Theta(L^{-1})E\left[\Theta(L^{-1})^{-1}x_t|\Im_t\right]\right] &= E\left[\varsigma_t E\left[\Theta(L^{-1})E\left[\Theta(L^{-1})^{-1}x_t|\Im_t\right]|\Im_t\right]\right] \\
&= E\left[\varsigma_t E\left[E\left[\Theta(L^{-1})\Theta(L^{-1})^{-1}x_t|\Im_t\right]|\Im_t\right]\right] \\
&= E\left[\varsigma_t x_t\right],
\end{aligned}$$

the left hand side of (5).

The optimal instrument (7) was used in empirical work by West and Wilcox (1996) for a first-order condition characterized by conditional homoskedasticity, from a dynamic inventory model. In West and Wilcox (1996) small sample properties of the optimal estimator are investigated; West, Wong and Anatolyev (2002) also contains asymptotic computations and simulation evidence. Hansen and Singleton (1996) applied the instrument (7) for a variant of the CAPM in both conditionally homo- and heteroskedastic environments, with the understanding that in the latter case it is not optimal. Hansen and Singleton (1991) propose implementation algorithms based on state space methods.

Remark 6.1 Hayashi and Sims (1983) warn that backward filtering of the equation in order to whiten the error term, and subsequently applying IV estimation leads to inconsistent estimation. This occurs because the Wold innovation ε_t is not necessarily a martingale difference and hence is not orthogonal to the instruments from the past; see also Hansen and West (2002, Section 2) for more on this point. Hayashi and Sims (1983) instead suggest forward filtering that also whitens the error term and preserves orthogonality to past data.

Remark 6.2 Hansen and Singleton (1991, 1996) give the following interpretation of the formula (7): apply the Hayashi and Sims (1983) forward filter $\Theta(L^{-1})^{-1}$ to the original

equation to get rid of serial correlation, project the forward filtered right-hand-side variable x_t onto $\Im_t$ to get a valid instrument, and then filter this instrument back with the filter $\Theta(L)^{-1}$ to the terms of the original equation. It turns out that under conditional homoskedasticity, the instrument so constructed is optimal.

7 Multiperiod conditional problems with conditional heteroskedasticity

Hansen (1985) and Hansen, Heaton and Ogaki (1988), using Gordin's (1969) martingale difference approximation and Hayashi and Sims' (1983) forward filtering idea, presented a characterization of the efficiency bound for IV estimators that correspond to a given system of conditional moment restrictions. Building on this characterization, Anatolyev (2003b) shows that the form of the process followed by the optimal instrument is

$$\Phi_t(L)\zeta_t = E\left[\Psi_t(L^{-1})x_t|\Im_t\right], \tag{8}$$

where, in contrast to (7), the filters $\Phi_t(L)$ and $\Psi_t(L)$ are time varying and tied to heteroskedasticity and serial correlation characteristics of errors via an implicit system of nonlinear stochastic equations.

To be more specific, we will consider the case of first-order conditional serial correlation, i.e. when $E[e_t e_{t-j}|\Im_t] = 0$ for $j > 1$. Denote $\omega_t = E\left[e_t^2|\Im_t\right]$, $\gamma_t = E\left[e_t e_{t-1}|\Im_t\right]$, the conditional variance and conditional first-order autocovariance of the errors. Then the optimal instrument ζ_t follows (Anatolyev, 2003b)

$$\zeta_t = \phi_t\zeta_{t-1} + \rho_t\delta_t, \tag{9}$$

where the auxiliary processes ϕ_t, ρ_t and δ_t are determined from the following system:

$$\gamma_t + \phi_t\left(\omega_t + E_t\left[\phi_{t+1}\gamma_{t+1}|\Im_t\right]\right) = 0, \tag{10}$$

$$\rho_t = -\phi_t\gamma_t^{-1}, \tag{11}$$

$$\delta_t = E\left[x_t + \phi_{t+1}\delta_{t+1}|\Im_t\right], \tag{12}$$

$$E[\log|\phi_t|] < 0, \tag{13}$$

provided that the solution of the system exists and has finite forth order moment.

Remark 7.1 The linear autoregressive-like structure of the process ζ_t does not mean that the instrument is optimal in the class of instruments having additive structure. This structure is a result of the linearity of the optimality condition which stems from the quadratic nature of GMM, or, to be more precise, from the asymptotic linearity of the GMM first-order conditions.

Remark 7.2 Recalling the Hansen and Singleton (1991, 1996) interpretation of (7), note from (9) and (12) that to attain optimality under conditional heteroskedasticity, *time varying* filters, correspondingly forward $(1 - L^{-1}\phi_t)^{-1}$ and backward $1 - \phi_t L$, must be applied, along with the appropriate time varying scaling by ρ_t.

Remark 7.3 The condition (13) rules out unstable solutions of the nonlinear equation (10). For example, if $e_t = \varepsilon_{t+1} - \theta\varepsilon_t$ for some $\theta \in (-1, 1)$, with $\{\varepsilon_t\}$ being a martingale difference relative to $\Im_t$, with $\sigma_t^2 \equiv E_t[\varepsilon_t^2]$, then $\omega_t = E_t[\sigma_{t+1}^2] + \theta^2\sigma_t^2$, $\gamma_t = -\theta\sigma_t^2$, equation (10) can be rewritten as $(1 - \theta\phi_t)\sigma_t^2 + \phi_t E_t[(\phi_{t+1} - \theta^{-1})\sigma_{t+1}^2] = 0$, and one can notice that $\phi_t = \theta^{-1}$ for all t is a solution and is unstable. Unfortunately, the stable one is *not* (except under conditional homoskedasticity) simply $\phi_t = \theta$.

Remark 7.4 Although in general $|\phi_t|$ may exceed unity for a set of realizations of $\Im_t$ of nonzero measure, a sufficient condition for stability is $\sup_t |\phi_t| < 1$, which may sometimes be shown without knowledge of the solution to (10). For example, let the innovation ε_t follow the $GARCH(1,1)$ process: $\sigma_t^2 = \sigma^2(1 - a - b) + a\varepsilon_{t-1}^2 + b\sigma_{t-1}^2$, where σ^2 is the unconditional variance of ε_t, and $0 < \theta < a + b < 1$. Then $\kappa \equiv \theta \sup_t \left\{\sigma_t^2 / E_t[\sigma_{t+1}^2]\right\} < 1$ and it is easy to show by induction over iterations in

$$\phi_t^{(0)} = 0 \; \forall t, \quad \phi_t^{(j+1)} = \left\{\theta + E_t\left[\frac{\sigma_{t+1}^2}{\sigma_t^2}(\theta^{-1} - \phi_{t+1}^{(j)})\right]\right\}^{-1}, \quad j = 0, 1, \cdots,$$

that $0 \leq \phi_t \leq \kappa < 1$. If in addition ε_t has bounded support, the condition $\theta < a + b$ may be relaxed.

The major trouble that distinguishes the present situation from that with single-period restrictions or under conditional homoskedasticity is that in general the system (10)–(13) cannot be solved analytically for the auxiliary processes. The Heaton and Ogaki (1991) example (see the bottom of this Section) is a notable exception, although in order to accomplish the goal, one has to assume normality of the fundamental process, which nullifies this example's practical significance. Therefore, if the above result is used as a point of departure in constructing the feasible instrument, there seem to be two ways to proceed: one is to estimate the auxiliary processes directly from the system by designing an iterative scheme, the other is to approximate the system in such a way that an explicit though approximate solution can be obtained and estimated.

7.1 Iterative optimal instrument

The following is based on ideas in Anatolyev (1999). Note that ϕ_t is defined implicitly by (10) accompanied by (13). A serious annoyance is the presence of unstable solutions along with the stable one sought for. One may think of designing an iterative scheme that would lead to the appropriate stable solution of (10) at the population level. The obvious

candidate is

$$\phi_t^{(j)} = -\left(\omega_t + E\left[\phi_{t+1}^{(j-1)}\gamma_{t+1}|\Im_t\right]\right)^{-1}\gamma_t, \quad j = 1, 2, \cdots, \tag{14}$$

until convergence is attained, if it is attained. The iterative scheme (14) will work if it embeds contraction, i.e. if $||\phi_t^{(j+1)} - \phi_t^{(j)}|| \leq \kappa||\phi_t^{(j)} - \phi_t^{(j-1)}||$ for some $\kappa < 1$ and some norm $||\cdot||$ over infinite sequences. The iterative scheme (14) is reminiscent of one way of numerically finding the stable root of the quadratic equation

$$ax^2 + bx + c = 0, \quad a > 0, \quad b^2 > 4ac,$$

using the iterative scheme

$$x^{(j)} = -\frac{c}{ax^{(j-1)} + b}, \quad j = 1, 2, \cdots,$$

which does embed contraction in the vicinity of the stable root. In order to fall in its vicinity, and not in the vicinity of the unstable root, one may start the iterations from, say, zero. In fact, the example with the quadratic equation corresponds to the case of conditional homoskedasticity, with $a = c = \theta\sigma^2$, $b = -\left(1 + \theta^2\right)\sigma^2$, where θ and σ^2 are parameters of the Wold decomposition. In more general cases, one can start the iterations from $\phi_t^{(0)} = 0$ for all t, or, alternatively, from the "conditionally homoskedastic solution" $\phi_t^{(0)} = \theta$ for all t.

In addition to the relatively trivial case of conditional homoskedasticity (and, by continuity, cases of small departures from conditional homoskedasticity), contraction embedding can be shown to hold for a pretty general class of error structures. Assume that

$$e_t = g_0 w_{t+1} + g_1 w_t,$$

where g_0 and g_1 are $1 \times p$ row vectors, and w_t is a $p \times 1$ column vector of disturbances with the properties $E[w_t|\Im_t \vee \sigma(w_{t-1}, w_{t-2}, \cdots)] = 0$, $E[w_t w_t'|\Im_t \vee \sigma(w_{t-1}, w_{t-2}, \cdots)] = \Sigma_t$. Then, provided that $\phi_t^{(0)} = 0$, it is possible to show that $\sup_t |\phi_t^{(j)} - \phi_t^{(j-1)}| \to 0$ as $j \to \infty$. It follows that the iterative scheme converges to the stable fixed point ϕ_t in supnorm.

Unfortunately, it does not seem possible to verify contraction embedding for a fully general case, so the iterative scheme may in principle explode or converge to an unstable solution. In practice, if there is no enough information to prove contraction embedding, one may check the stability of the emerged estimated series provided that the algorithm did converge. Alternatively, trimming $\phi_t^{(j)}$ away from large absolute values at each iteration j may help arrive at an estimated series having a property of a stable solution.

An analogous iterative scheme may be set up for δ_t based on equation (12):

$$\delta_t^{(j)} = E\left[x_t + \phi_{t+1}\delta_{t+1}^{(j-1)}|\Im_t\right], \quad j = 1, 2, \cdots,$$

starting from, say, $\delta_t^{(0)} = 0$, if it embeds contraction. If contraction embedding is suspect, an alternative way is to approximate δ_t by

$$\delta_t^{\approx} = E\left[x_t + \phi_{t+1}x_{t+1} + \phi_{t+2}\phi_{t+1}x_{t+2} + \cdots + \phi_{t+J}\phi_{t+J-1}\cdots\phi_{t+1}x_{t+J}|\Im_t\right] + \frac{\theta^{J+1}}{1-\theta}E\left[x_t\right]$$

for sufficiently big J.

The iterative schemes described above may be used to make up analogous schemes at the sample level, at each iteration nonparametrically estimating needed conditional expectations. After an estimate $\hat{\phi}_t$ is computed for all t, estimates $\hat{\delta}_t$ are found using already computed $\hat{\phi}_t$; estimates $\hat{\rho}_t$ can be calculated using (11). The feasible *iterative optimal instrument* $\hat{\zeta}_t$ is then constructed recursively using (9):

$$\hat{\zeta}_t = \hat{\zeta}_{t-1}\hat{\phi}_t + \hat{\rho}_t\hat{\delta}_t, \quad t = 1, 2, \cdots, T,$$

starting from, say, $\hat{\zeta}_0 = 0$.

7.2 Approximately optimal instrument

Note that in the absence of conditional heteroskedasticity and serial correlation, the ideal instrument is proportional to $E\left[x_t|\Im_t\right]$. Recall the Hansen (1985) instrument (7) that would be optimal if there were no conditional homoskedasticity. In the case of first-order serial correlation it is

$$\zeta_t = \zeta_{t-1}\theta + \frac{1}{\sigma^2}E\left[\sum_{i=0}^{\infty}\theta^i x_{t+i}|\Im_t\right], \tag{15}$$

where σ^2 is the variance of the implied Wold innovation of e_t, and θ is a negative of its first-order autocovariance coefficient. Note that the instrument (15) reduces to $E\left[x_t|\Im_t\right]$ by setting $\theta = 0$, hence let $\phi_t^0 = \theta$ be a benchmark when we look for an approximation for ϕ_t in the face of conditional heteroskedasticity. Now recall that when the error is conditionally serially uncorrelated, the optimal instrument is that of Chamberlain (1987):

$$\zeta_t = \frac{E\left[x_t|\Im_t\right]}{\omega_t}. \tag{16}$$

Note that the instrument (16) reduces to $E\left[x_t|\Im_t\right]$ by setting $\omega_t = \text{const}$, hence let $\rho_t^0 = \omega_t^{-1}$ be a benchmark when we look for an approximation for ρ_t in the face of serial correlation.

The idea of constructing an approximation to the optimal instrument is to find an explicit solution of the system that emerges after the equation for ϕ_t

$$\gamma_t + \phi_t\left(\omega_t + E\left[\phi_{t+1}\gamma_{t+1}|\Im_t\right]\right) = 0,$$

and the equation for ρ_t

$$\rho_t(\omega_t - E\left[\rho_{t+1}\gamma_{t+1}^2|\Im_t\right]) = 1,$$

are linearized with respect to ϕ_t and ρ_t around the mentioned benchmarks. The details of derivations of the approximations can be found in Anatolyev (2002b). The solution of this approximated system results in the following *approximately optimal instrument* $\bar{\zeta}_t$:

$$\bar{\zeta}_0 = 0, \quad \bar{\zeta}_t = \left\langle\bar{\phi}_t\right\rangle\bar{\zeta}_{t-1} + \bar{\rho}_t\sum_{i=0}^{\infty}E\left[\prod_{j=0}^{i-1}\bar{\phi}_{t+j}x_{t+i}|\Im_t\right], \quad t = 1, 2, \cdots, T, \tag{17}$$

where

$$\bar{\phi}_t = \theta + \frac{1}{\sigma^2}\left(\gamma_t - \sum_{i=0}^{\infty}\theta^{2i}E\left[\theta\omega_{t+i} + 2\gamma_{t+i}|\Im_t\right]\right), \tag{18}$$

$$\bar{\rho}_t = \frac{1}{\omega_t}, \tag{19}$$

and $\langle\cdot\rangle$ indicates trimming to ensure stability of the recursion and existence of fourth moments of $\bar{\zeta}_t$. A simple trimming device is

$$\langle\varsigma\rangle \equiv \min\{1-\epsilon, \max\{-1+\epsilon, \varsigma\}\}$$

for a small positive number ϵ.

An interesting question is how much is lost in terms of asymptotic efficiency from the use of an approximately optimal instrument in place of the truly optimal one. Naturally, to answer this question we need to have exact analytical expressions for both instruments. The only non-trivial (i.e. conditionally heteroskedastic) example known in the literature where the optimal instrument can be analytically derived is the Heaton and Ogaki (1991) problem. Unfortunately, in order to accomplish this goal, one has to assume normality of the fundamental process, which nullifies this example's practical significance. The example, however, can be used to get a feel of asymptotic efficiency gains and losses; it is, of course, unclear if the relative figures carry over to other problems.

We briefly outline the Heaton and Ogaki (1991) example without giving explicit forms for the instruments as the analytic expressions are quite messy, and then proceed with calibration of asymptotic variances. Let w_t be a serially independent standard normal series, and u_t be a two-period ahead forecast error with the Wold representation $u_t = \nu_0 w_t + \nu_1 w_{t-1}$. Observable at time t is z_t, and the space of instruments is $\Im_t = \sigma(z_t, z_{t-1}, \dots)$. Let u_t be connected to z_t via $u_t = z_t + \beta z_{t-1}$, where β is a scalar parameter of interest. The rational expectations hypothesis imposes the conditional moment restriction

$$E_t\left[u_{t+2}\right] = 0.$$

Under the assumptions made, the error in this equation is conditionally homoskedastic. There is conditional heteroskedasticity in another conditional moment restriction, a conditional analog of the Working (1960) result on temporal aggregation:

$$E_t\left[u_{t+2}\left(\rho u_{t+2} - u_{t+1}\right)\right] = 0,$$

where $\rho \equiv \nu_0\nu_1/\left(\nu_0^2 + \nu_1^2\right) = \frac{1}{4}$. To calibrate asymptotic variances, let $\nu_0 = 1$, $\nu_1 = 2-\sqrt{3}$. The Table below presents asymptotic variances of some IV estimators for a number of values of β. The optimal IV estimator is most efficient, and significantly beats the optimal IV estimator that ignores the second equation ("first equation optimal"), especially when β is close to ν_1. The "homoskedasticity optimal" instrument that would be optimal if there were no conditional heteroskedasticity captures much of the efficiency gains. However, the

approximately optimal instrument captures an overwhelming part of the further efficiency gains provided by the optimal instrument. Thus, the efficiency losses arising from the approximation error turn out to be small (only $1 \div 8\%$ for the shown values of β), and show that the approximately optimal instrument is able to nearly attain the efficiency bound.

β	-0.8	-0.3	0	$+0.3$	$+0.8$
Optimal	0.360	0.910	1.000	0.910	0.360
Approximately optimal	0.365	0.920	1.012	0.924	0.390
Homoskedasticity optimal	0.399	1.070	1.313	1.235	0.430
First equation optimal	0.466	3.293	13.93	749.4	0.786

8 On estimation of conditional expectations

In previous Sections various complex forms of instruments are presented. Barring special cases, these forms necessarily contain conditional expectations (conditional variances, conditional autocovariances, conditional scores, etc.), sometimes an infinite number thereof. As mentioned in the introduction, "parametric" feasible estimation requires extraneous knowledge of the functional forms for these conditional expectations. If a researcher is willing to impose such auxiliary parameterizations, the "parametric optimal instrument" yields an estimator that will likely be consistent, but their misspecification will lead to asymptotic inefficiency in the class of allowable estimators. At the end of the day, such an estimator may be no better (while being harder to implement) than, say, a quasi-maximum likelihood (QML) estimator using same or different parameterizations (see comparisons in Section 10). To really attain the IV efficiency bound, a researcher is expected to go all the way and estimate the involved conditional expectations non-parametrically. To allow doing that, and to allow assessing gains from doing that, a sample size should be plentiful to allow precise non-parametric estimation, and the problem's characteristics should be such that the precision of the optimal IV estimator is likely to overweight those of competitors even in a given finite sample. Usually, this means the presence of strong heteroskedasticity, and in multiperiod problems – in addition the presence of noticeable serial correlation. More precise characterization, unfortunately, is not possible in general.

"Nonparametric" feasible versions of previously considered instruments entails replacement of conditional expectations in a formula for the optimal instrument by their nonparametric estimates. For example, the auxiliary process ϕ_t of the iterative optimal instrument is estimated by repeating the scheme (14) at the sample level:

$$\begin{aligned}
\hat{\phi}_t^{(0)} &= \hat{\theta}, \quad t = 1, 2, \cdots, T, T+1; \\
\hat{\phi}_{T+1}^{(j)} &= \hat{\theta}, \quad j = 1, 2, \cdots; \\
\hat{\phi}_t^{(j)} &= -\left(\hat{E}\left[\hat{e}_t^2 + \hat{\phi}_{t+1}^{(j-1)} \hat{e}_{t+1}\hat{e}_t | \Im_t\right]\right)^{-1} \hat{E}\left[\hat{e}_t\hat{e}_{t-1} | \Im_t\right], \quad t = 1, 2, \cdots, T, \quad j = 1, 2, \cdots,
\end{aligned}$$

where $\hat{\theta}$ and $\hat{e}_t$ are an MA(1) coefficient estimate and residuals from a preliminary estimation step, and $\hat{E}$ denotes an estimator of the conditional expectation operator. Next, some of expressions for various components of complex instruments contain infinite summations of certain conditional expectations. For example, (18) contains

$$\gamma_t - \sum_{i=0}^{\infty} \theta^{2i} E\left[\theta\omega_{t+i} + 2\gamma_{t+i} | \Im_t\right].$$

This does not mean that one has to evaluate an infinite number of conditional expectations; instead, one may collect all terms under a single expectation sign and trim the summation. For example, the preceding expression may be approximated with an arbitrary precision by

$$E\left[e_t e_{t-1} - \sum_{i=0}^{J} \theta^{2i}\left(\theta e_{t+i}^2 + 2e_{t+i}e_{t+i-1}\right) | \Im_t\right] - \frac{\theta^{2J+2}}{1-\theta^2} E\left[\theta e_t^2 + 2e_t e_{t-1}\right]$$

for some large J.

In the rest of this Section, we discuss some aspects of non-parametric estimation of conditional expectations. In an IID framework, to construct feasible GLS estimators in regression models, Carroll (1982) proposes kernel methods for their estimation, Robinson (1987) suggests a nearest neighbors approach, Newey (1990) considers nearest neighbors and series approximations, Donald, Imbens and Newey (2003) use splines; see also Pagan and Ullah (1999) for a survey. In time series, the matter is additionally complicated by the fact that the composition of the state vector is unknown *a priori*, and the state vector may potentially involve an infinite number of variables. In practice, the IID techniques may be applied for time series, provided that the state vector is finite dimensional (and small-dimensional, too!). Intuitively, the state vector may be made expanding with a sufficiently slow rate to reach complete conditioning asymptotically, but no formal results are established in this regard. Some promising ongoing work (Pinkse, 2005) seeks ways to reduce the dimensionality of the state vector using dimension reduction techniques, but to date this work deals only with cross-sections and panels.

More recent statistics and econometrics literature, however, does provide some practical nonparametric and semiparametric tools.

Tschernig and Yang (2000) propose a criterion called "nonparametric corrected asymptotic final prediction error" (CAFPE) to simultaneously select a composition of the state vector (so called "significant lags") in conditioning and optimal bandwidth values in the context of kernel estimation. When the basic instrument is a single variable, the criterion to be minimized is

$$CAFPE = \left(\hat{A} + \frac{2}{T - i_m + 1}\left(\frac{k(0)}{h_{opt}}\right)^m \hat{B}\right)\left(1 + \frac{m}{(T - i_m + 1)^{4/(m+4)}}\right), \qquad (20)$$

where $k(u)$ is a kernel function, m is a number of employed lags of the basic instrument, and i_m is its maximal employed lag. Here, $\hat{A}$ and $\hat{B}$ are the following nonparametric

estimates of ingredients of the asymptotic final prediction error:

$$\begin{aligned}\hat{A} &= \frac{1}{T-i_m+1}\sum_{t=i_m}^{T}\left(y_t-\hat{m}_y\left(z_{t-i_1},\cdots,z_{t-i_m}\right)\right)^2 w_t,\\ \hat{B} &= \frac{1}{T-i_m+1}\sum_{t=i_m}^{T}\frac{\left(y_t-\hat{m}_y\left(z_{t-i_1},\cdots,z_{t-i_m}\right)\right)^2}{\hat{\mu}\left(z_{t-i_1},\cdots,z_{t-i_m}\right)}w_t,\end{aligned}$$

where $\hat{m}_y\left(z_{t-i_1},\cdots,z_{t-i_m}\right)$ is a Nadaraya–Watson estimate of the regression function of y_t on the included lags of the basic instrument, $\hat{\mu}\left(z_{t-i_1},\cdots,z_{t-i_m}\right)$ is a Nadaraya–Watson estimate of the joint density of the vector of included lags, and w_t equals zero if the associated joint density estimate $\hat{\mu}\left(z_{t-i_1},\cdots,z_{t-i_m}\right)$ is among the lowest 5% over the values of $\left(z_{t-i_1},\cdots,z_{t-i_m}\right)$ in the sample, and unity otherwise (such screening off extreme observations is conventional in nonparametric estimation literature; see also Tjøstheim and Auestad, 1994). The optimal bandwidth h_{opt} is determined via a grid search procedure. The first term in (20) is a nonparametric estimate of the asymptotic final prediction error, the second term in (20) is a correction aimed at penalizing lag overfitting (i.e. choosing superfluous lags in addition to correct ones). When a number of significant lags is likely to be large, a search across all lag combinations may take enormous computational time. To speed up the search, Tjøstheim and Auestad (1994) suggest a shortcut by picking significant lags one by one, taking already selected ones as included.

To increase a precision of nonparametric estimation, Glad (1998) and Fan and Ullah (1999) suggest a "combined" estimator by first estimating a parametric possibly nonlinear projection (hopefully capturing a large part of the dependence) and then estimating a multiplicative residual nonparametrically. Mishra, Su and Ullah (2004) extended this approach to estimation of conditional variance. Likewise, other similar objects like conditional autocovariances may be estimated in a similar two-step manner.

When there is high persistence in the processes of conditional estimators of interest (say, in conditional volatility), the true state vector may be prohibitively large leading to the curse of dimensionality for the unconstrained nonparametric regression. In such cases, constraints of the kind $\sigma_t^2=\varphi\sigma_{t-1}^2+g\left(\Im_t\right)$, where σ_t^2 is (say) a conditional variance, and $g\left(\Im_t\right)$ is an unspecified function of a *small number* of arguments belonging to $\Im_t$, may be imposed in order to semiparametrically estimate the conditional expectations; see the semiparametric ARCH models of Linton and Mammen (2005).

9 Linear subclass of instruments

The theory presented in previous Sections is difficult. As a result, much greater popularity has been gained by the theory of *linear optimal instruments* because of its simplicity and tractability. One searches for the optimal instrument within a narrower subclass of allowable instruments than (4), the *linear subclass* consisting of only linear functions of

the basic instruments:

$$\mathcal{Z}_t = \left\{ \varsigma_t : \ \varsigma_t = \sum_{i=0}^{\infty} \kappa_i z_{t-i} \text{ for some } k \times \ell \text{ weights } \kappa_i \text{ s.t. } \sum_{i=0}^{\infty} |\kappa_i| < \infty \right\}. \tag{21}$$

The theory and feasible implementation are developed in West (2001), West, Wong and Anatolyev (2002), and Kuersteiner (2001, 2002). A convenient tool is the optimality condition (5).

Generally, employing the linear subclass of instruments delivers efficiency gains, often substantial, compared to the use of the basic instrument or a finite number of its lags (Stambaugh, 1993; Kuersteiner, 2002), especially in multiperiod problems (Kuersteiner, 2001, West, Wong and Anatolyev, 2002). Sometimes a special structure of conditional heteroskedasticity may kill any the gains though. An example of redundancy of all lags can be found in Anatolyev (2003a) and Carrasco (2004), of redundancy of all lags but their small number – in West (2002). However, in spite of likely efficiency gains, the linear subclass is significantly narrower than the entire class of allowable instruments. Nevertheless, one may be willing to sacrifice potential efficiency gains provided by conditional moment restrictions for the sake of simplicity and tractability of the resulting theory. This can be observed, for instance, in the eagerness to use OLS in place of more efficient GLS when testing for forecast unbiasedness (as in Example 1 of Section 2). In Section 10 we will present some numerical evidence revealing whether this choice is justified.

Example: optimal instrument in conditionally heteroskedastic AR(1) model Consider a zero mean AR(k) model

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \cdots + \rho_k y_{t-k} + e_t, \tag{22}$$

where the martingale difference innovations e_t, conditionally on $\Im_t$, are symmetrically distributed and have variance σ_t^2. In this case, $\beta = (\rho_1 \ \rho_2 \ \cdots \ \rho_k)'$, $x_t = (y_{t-1} \ y_{t-2} \ \cdots \ y_{t-k})'$. Let us see how the linear optimal instrument is constructed for a conditionally heteroskedastic AR(1) model. Conditional on $\Im_t$, let the innovations ε_t be symmetrically distributed and have variance σ_t^2. The linear subclass of allowable instruments is

$$\mathcal{Z}_t = \left\{ \varsigma_t : \ \varsigma_t = \sum_{i=1}^{\infty} \kappa_i y_{t-i} \text{ for some scalar weights } \kappa_i \text{ s.t. } \sum_{i=1}^{\infty} |\kappa_i| < \infty \right\}.$$

Using the West (2001) trick, we let the optimal instrument be $\zeta_t = \sum_{i=1}^{\infty} \phi_i \varepsilon_{t-i}$ and rewrite the optimality condition as

$$E\left[y_{t-1}\varepsilon_{t-r}\right] = cov\left(\sum_{i=1}^{\infty} \phi_i \varepsilon_{t-i}\varepsilon_t, \varepsilon_{t-r}\varepsilon_t\right),$$

for all $r \geq 1$. See also West (2002). When e_t is conditionally homoskedastic, it follows that $\phi_r \propto \rho^r$, i.e. $\zeta_t \propto x_t$, and OLS is efficient.[6] In the general case of conditional

[6]Even more easily this result can be obtained by seeing that the optimal combination of unconditional moment restrictions corresponding to most recent $\ell \geq k$ lags of y_t as instruments is

heteroskedasticity,

$$\phi_r = \frac{\rho^{r-1}}{\tau_r},$$

where $\tau_r \equiv \left(E\left[\varepsilon_t^2\right]\right)^{-2} E\left[\varepsilon_t^2 \varepsilon_{t-r}^2\right]$. Suppose a consistent estimate $\hat{\rho}$ of ρ and consistent estimates $\hat{\tau}_r$ for τ_r for all $r \geq 1$ are available. Then feasible $\hat{\phi}_r$ for all $1 \leq r \leq T-1$ along with estimates $\hat{\varepsilon}_t$ for $1 \leq t \leq T-1$ can be computed; a feasible optimal instrument $\hat{\zeta}_t$ can be constructed as $\sum_{i=1}^{t-1} \hat{\phi}_i \hat{\varepsilon}_{t-i}$, where $\hat{\varepsilon}_t = y_t - \hat{\rho} y_{t-1}$ are residuals.

Example: optimal instrument in conditionally heteroskedastic ARMA$(1,1)$ model Consider the following ARMA(1,1) model:

$$y_t = \rho y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}, \tag{23}$$

where the martingale difference innovations ε_t, conditionally on $\Im_t$, are symmetrically distributed and have variance σ_t^2. In this case, $\beta = \rho$, $x_t = y_{t-1}$, and $e_t = \varepsilon_t - \theta \varepsilon_{t-1}$. Assume that $|\theta| < 1$, $\theta \neq 0$, $\theta \neq \rho$. Dolado (1990) proves the intuitive fact that among all y_{t-j}, $j \geq 2$, the instrument y_{t-2} is best. However, using y_{t-j} for some $j \geq 3$ along with y_{t-2} in general increases efficiency.

The linear subclass of allowable instruments is

$$\mathcal{Z}_t = \left\{ \varsigma_t : \ \varsigma_t = \sum_{i=1}^{\infty} \kappa_i y_{t-1-i} \text{ for some scalar weights } \kappa_i \text{ s.t. } \sum_{i=1}^{\infty} |\kappa_i| < \infty \right\}.$$

Again using the West (2001) trick, we let the optimal instrument be $\zeta_t = \sum_{i=1}^{\infty} \phi_i \varepsilon_{t-1-i}$ and rewrite the optimality condition as

$$E\left[y_{t-1} \varepsilon_{t-1-r}\right] = \sum_{j=-1}^{+1} cov\left(\sum_{i=1}^{\infty} \phi_i \varepsilon_{t-1-i} \left(\varepsilon_t - \theta \varepsilon_{t-1}\right), \varepsilon_{t-1-r-j} \left(\varepsilon_{t-j} - \theta \varepsilon_{t-1-j}\right) \right),$$

for all $r \geq 1$. Upon simplification we obtain a second-order linear difference equation for ϕ_i with time-varying coefficients:

$$-\theta \tau_{r+1} \phi_{r+1} + \left(\tau_{r+1} + \theta^2 \tau_r\right) \phi_r - \theta \tau_r \phi_{r-1} \propto \rho^r, \quad \text{for all } r \geq 1, \tag{24}$$

subject to two boundary conditions $\phi_0 = \phi_\infty = 0$.

Consider first the case of conditional homoskedasticity, in which case τ_r does not depend of r. Then the solution of the difference equation (24) is

$$\phi_r \propto \rho^r - \theta^r,$$

$Q_{xz} V_{ze}^{-1} \propto \left[I_k \ O_{k \times (\ell - k)}\right]$ (see section 3, and also Kim, Qian and Schmidt (1999) and Anatolyev (2002a)), or by considering optimal instrumentation for the conditional moment restriction $E\left[e_t | y_{t-1}, y_{t-2}, \cdots\right] = 0$ (see section 5), or simply by comparing GMM and 2SLS (coinciding with OLS) estimators which are asymptotically equivalent under the circumstances.

and the optimal instrument is

$$\zeta_t \propto \sum_{i=1}^{\infty} \left(\rho^i - \theta^i\right) \varepsilon_{t-1-i} \propto \sum_{i=1}^{\infty} i\theta^{i-1} y_{t-1-i}.$$

One can see that the optimal weights at the lags die out a bit slower than exponentially. Another, more useful (anticipating further discussion), representation of ζ_t is

$$(1-\theta L)\,\zeta_t \propto \frac{y_{t-2}}{1-\theta L}.$$

This formula is a special case of an expression for the optimal instrument for estimation of parameters of the finite-lag polynomial $\mathrm{P}(L)$ in the general ARMA model $\mathrm{P}(L)y_t = \Theta(L)\varepsilon_t$ with zero mean IID innovations is

$$\Theta(L)\zeta_t \propto \mathrm{proj}\left[\Theta(L^{-1})^{-1} x_t | \mathcal{Z}_t\right]. \tag{25}$$

In turn, (25) is an application of Hansen's (1985) formula for the optimal instrument in conditionally homoskedastic models (cf. (7)). Optimal instrumentation in homoskedastic ARMA models was also independently considered in the engineering literature, see Stoica, Söderström and Friedlander (1985, 1987).

In the general case of conditional heteroskedasticity, the solution of the difference equation (24) is[7]

$$\phi_r = \sum_{i=0}^{\infty} \theta^i \frac{\rho^{j+i} - \theta^{j+i}\left(1+\theta\phi_1\tau_1\right)}{\tau_{r+i+1}},$$

where

$$\phi_1 = \left(1 + \sum_{i=1}^{\infty} \theta^{2i} \frac{\tau_1}{\tau_{i+1}}\right)^{-1} \sum_{i=1}^{\infty} \theta^{i-1} \frac{\rho^i - \theta^i}{\tau_{i+1}}.$$

Suppose consistent estimators $\hat{\rho}$ and $\hat{\theta}$ of ρ and θ are available, along with consistent estimates $\hat{\tau}_r$ for τ_r for all $r \geq 1$. Then feasible $\hat{\phi}_r$ for all $1 \leq r \leq T-2$ along with estimates $\hat{\varepsilon}_t$ for $1 \leq t \leq T-2$ can be computed; a feasible optimal instrument $\hat{\zeta}_t$ can be constructed as $\sum_{i=1}^{t-2} \hat{\phi}_i \hat{\varepsilon}_{t-1-i}$.

More generally, suppose that we want to determine $k \times \ell$ optimal weights in the representation of the optimal instrument $\zeta_t = \sum_{i=0}^{T-1} \phi_i \varepsilon_{t-i}$ for the equation (3) with $E\left[e_t e_{t-q} | \Im_t\right] = 0$ for finite integer $q > 0$, where ε_t is the Wold innovation in the $\ell \times 1$ basic instrument z_t (simplifying that there is no deterministic part). Then the optimality condition looks as follows:

$$\Psi = S\Phi,$$

where $\Psi \equiv E\left[\mathcal{E}_t x_t'\right]$, $S \equiv \sum_{j=-q}^{+q} E\left[\mathcal{E}_t \mathcal{E}_{t-j}' e_t e_{t-j}\right]$, and $\mathcal{E}_t \equiv vec\left(\varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_{t-T+1}\right)$. The matrices Ψ and S are $T\ell \times k$ and $T\ell \times T\ell$, respectively, and the matrix of weights $\Phi \equiv vec\left(\phi_0', \phi_1', \cdots, \phi_{T-1}'\right)$ sought for is $T\ell \times k$. The solution is, of course, $\Phi = S^{-1}\Psi$, with

[7] This solution is due to Kenneth West and Ka-fu Wong.

the feasible version being $\hat{\Phi} = \hat{S}^{-1}\hat{\Psi}$, where $\hat{\Psi}$ and $\hat{S}$ are estimates of Ψ and S, respectively. West, Wong and Anatolyev (2002) proceed by imposing auxiliary parametric structures on the moments that figure in the matrices Ψ and S. In their leading example, they impose an assumption of symmetry of fourth-order moments of innovations hence reducing the dimensionality of the matrix S.

In the above two examples, postulating and estimating an auxiliary parametric model for σ_t^2 like GARCH(1,1): $\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$, yields

$$\tau_r = 1 + \alpha\left(1 + \frac{\alpha\beta}{1 - 2\alpha\beta - \beta^2}\right)(\alpha + \beta)^{r-1}(\tau_0 - 1).$$

This step obviously requires one to go beyond the initially given model and can induce misspecification. If the auxiliary model is misspecified, the feasible instrument will no longer be optimal. However, it is likely that even a misspecified equation will yield quite reliable estimates of fourth moments, and the instrumental variables estimator will be good, although not fully efficient. West, Wong and Anatolyev (2002) also show that neither trimming in the summation $\zeta_t = \sum_{i=0}^{T-1}\phi_i\varepsilon_{t-i}$ at some moderate J or at $t-1$, nor minor misspecification of auxiliary parameterizations have a significant effect on the performance of the optimal IV estimator.

Kuersteiner (2002) specializes to a class of semi-strong conditionally heteroskedastic AR models possessing the symmetry condition mentioned above, and proposes a (frequency domain) nonparametric estimator. Kuersteiner (2001) considers more flexible linear autoregressive models possibly of infinite order where the innovation sequence has a martingale difference property (this class includes semi-strong ARMA models in particular).

10 Some asymptotic comparisons

In this Section, we make asymptotic comparisons for some IV estimators and their competitors, to get a feel of efficiency–complexity trade-offs. The central questions are how much efficiency is gained by the linearly optimal instrument, how much efficiency is further gained by the non-linearly optimal instrument, and how big the relative size of these gains is. We use a simple AR(1)–GARCH(1,1) example to that end, where it is possible to compute asymptotic variances of all estimators of interest.

The results for $\alpha = \beta = 0.4$ and various values of ρ are contained in the table below. The conditional distribution of errors is standard normal, standardized Student's t with 6 degrees of freedom (so that the coefficient of kurtosis equals 6), or standardized and recentered Chi-square with 4 degrees of freedom (so that the coefficient of skewness equals $\sqrt{2}$, and the coefficient of kurtosis equals 6). Asymptotic variances of four estimators are compared. The first is the OLS estimator, an IV estimator with the basic instrument y_{t-1}. The second one, labeled "LOIV", is an IV estimator based on the linear optimal

instrument. The third one, labeled "NLOIV", is an IV estimator based on the non-linear optimal instrument. The fourth one, labeled "QML", is a quasi-maximum likelihood estimator based on the normal conditional density and correct specification of the variance dynamics. The table contains figures for a relative asymptotic efficiency with respect to the OLS estimator, i.e. for a ratio of the asymptotic variance of a particular estimator to the asymptotic variance of OLS.

	Normal			Student, $df = 6$			Chi-square, $df = 4$		
ρ	LOIV	NLOIV	QML	LOIV	NLOIV	QML	LOIV	NLOIV	QML
0.1	99.9	24.3	19.0	99.9	7.09	7.47	99.9	6.15	6.67
0.3	99.6	22.1	17.4	99.5	7.52	7.71	99.5	5.18	5.76
0.5	98.6	25.1	19.8	97.9	2.83	3.06	98.0	6.47	7.43
0.8	87.7	19.7	16.0	78.1	2.23	2.27	77.6	1.77	2.10

One can see that the linearly optimal instrument is not able to yield substantial efficiency gains relative to the basic instrument, unless the latter has a very persistent dynamics. Roughly consistent with the above evidence, figures for the relative efficiency for the linear optimal IV estimation with respect to OLS reported in Kuersteiner (2001, 2002) for AR(1)–ARCH(1) and ARMA(1,1)–ARCH(1) models vary between 0.8 (which is not that low) and 1.0. Furthermore, asymptotic calculations in West, Wong and Anatolyev (2002, Section 3) specializing on multiperiod problems give enough evidence that the use of infinite past of basic instruments tends to yield much more efficiency gains in multiperiod models than in single-period ones, these gains tending to be higher the higher is the degree of serial correlation in the error term.

It is clear that both nonlinear optimal instrument and quasi-maximum likelihood allow one to extract much more efficiency, with comparable relative efficiency figures. Naturally, the normal maximum likelihood is asymptotically more efficient in the conditionally normal model, while nonlinear optimal instrumentation is asymptotically more efficient when the conditional density is misspecified. Further, QML estimation when the variance equation is incorrectly specified may yield arbitrary asymptotic losses (for example, a conditional variance thought to be constant renders QML equivalent to OLS).

11 Illustrative application

In this Section we show a small application demonstrating some of techniques described above. We consider the Hansen and Singleton (1982) and Ferson and Constantinides (1991) consumption-based capital asset pricing models (CAPM), with the former being characterized by a one-period conditional restriction, and the latter being characterized by a two-period conditional restriction, i.e. by the presence of serial correlation of first order.

The Ferson and Constantinides (1991) CAPM with habit formation leads to the fol-

lowing conditional moment restriction:

$$E\left[\frac{\delta}{1+\alpha\delta}R_{t+1}\left(\left(\frac{s_{t+1}}{s_t}\right)^{-\gamma}+\alpha\delta\left(\frac{s_{t+2}}{s_t}\right)^{-\gamma}\right)\right.$$
$$\left.-\frac{1}{1+\alpha\delta}\left(1+\alpha\delta\left(\frac{s_{t+1}}{s_t}\right)^{-\gamma}\right)\mid\frac{c_t}{c_{t-1}},R_t,\frac{c_{t-1}}{c_{t-2}},R_{t-1},\cdots\right]=0,$$

where $s_t = c_t + \alpha c_{t-1}$, c_t is a consumption, and R_t is a (gross) market return; α is a habit formation/durability parameter, δ is a discount factor, and γ is a coefficient of risk aversion. The standard division by $1+\alpha\delta$ appears in order to avoid the trap $\gamma = 0$, $\alpha\delta = -1$. When α is zero, the model reduces to the Hansen and Singleton (1982) CAPM characterized by the following conditional moment restriction:

$$E\left[\delta R_{t+1}\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}-1\mid\frac{c_t}{c_{t-1}},R_t,\frac{c_{t-1}}{c_{t-2}},R_{t-1},\cdots\right]=0.$$

Note that in both problems the moment function is nonlinear in the parameter vector.

We use monthly consumption and return data from 1959.1 to 2005.7, totaling to $T = 555$ observations on c_t/c_{t-1} and R_t (monthly instead of quarterly data are used for the sake of precision of nonparametric estimates). For c_t, we use seasonally adjusted personal consumption expenditures published by the Bureau of Economic Analysis of the US Department of Commerce, for R_t – 3-month treasury bill secondary market rate. All data are taken from the FRED database at `http://research.stlouisfed.org/fred2`.

For conventional IV implementation, we choose the instrument vector

$$z_t=\left(1,\frac{c_t}{c_{t-1}},\cdots,\frac{c_{t-\ell+1}}{c_{t-\ell}},R_t,\cdots,R_{t-\ell+1}\right)'$$

for $\ell = 1$ and 2, and use the identity weight matrix at the preliminary stage. The residuals corresponding to $\ell = 2$ are used in constructing more complex IV estimators. Asymptotic variances are estimated using the Smith (2005) automatic formula with the lag truncation parameter 3. The linearly optimal instrument is implemented according to the algorithm in West, Wong and Anatolyev (2002, Additional Appendix, Section 5).

In implementing the iterative optimal and approximately optimal IV estimation, to estimate nonparametrically various conditional expectations we employ the Nadaraya–Watson estimator using a product kernel with Epanechnikov marginals and the CAFPE criterion of Tschernig and Yang (2000) to select significant lags and an optimal value of the bandwidth (see Section 8). In every loop, we make the bandwidths for c_t/c_{t-1} and R_t take one of 5 values on a two dimensional grid $[.5h_c, 1.5h_c] \times [.5h_R, 1.5h_R]$, where $h_\varsigma = (4/(m+2))^{1/(m+4)}(T - i_m + 1)^{-1/(m+4)}$, and ς is c or R, to determine optimal bandwidth values $h_{opt,c}$ and $h_{opt,R}$. In the standard formula (20) we change h_{opt}^{-m} to $h_{opt,c}^{-m_c}h_{opt,R}^{-m_R}$, where m_ς is a number of employed lags of the basic instrument ς, ς being c or R, and

$m_c + m_R = m$. The maximal m_c and m_R are set at 3; trimming of $|\phi_t^{(j)}|$ at 0.95 is used to ensure stability.

The results for both variations are presented below, with standard errors in parentheses. In the Ferson–Constantinides model, the degree of serial correlation is given by the estimate $\hat{\theta} = -0.188$ of the MA coefficient in a Wold representation $\varepsilon_{t+1} - \theta\varepsilon_t$. In both models, all conventional GMM J-tests do not reject the null of correct specification at the 5% significance level.

	Ferson–Constantinides			Hansen–Singleton	
Method	α	δ	γ	δ	γ
GMM with $\ell = 1$	0.201	1.193	41.62	1.075	21.36
	(0.070)	(0.101)	(16.84)	(0.034)	(5.80)
GMM with $\ell = 2$	0.194	1.160	35.89	1.061	19.08
	(0.066)	(0.086)	(14.49)	(0.031)	(5.21)
Linearly optimal	0.176	1.170	37.78	1.119	30.16
	(0.089)	(0.031)	(4.75)	(0.009)	(1.22)
Nonlinearly optimal	0.458	0.975	4.07	0.967	3.28
	(0.226)	(0.014)	(2.63)	(0.002)	(0.37)

In both models, the conventional GMM and linearly optimal IV estimates turns out to be close, the latter seeming to be more precise (though, the confidence intervals for conventional GMM may not be reliable, as documented in Tauchen, 1986 and West, Wong and Anatolyev, 2002). The nonlinearly optimal IV estimates are even more precise, but differ in value quite sizably from linear IV estimates. In particular, the discount factor is estimated to be larger than unity by linear IV methods but smaller than unity by the nonlinear method; the nonlinear optimal IV estimates of the degree of risk aversion are much smaller than those given by linear IV methods. The high variability of point estimates of risk aversion across models and estimation methods is traditional in the literature.

12 Special issues and related research

Redundancy of moment conditions and instruments Breusch, Qian, Schmidt and Wyhowski (1999) introduced a concept of redundancy of instruments and, more generally, of moment conditions, and derived criteria for redundancy in the context of both unconditional and conditional moment restrictions. A set of instruments is called redundant relative to a given set of instruments if expanding the latter set to include the former does not increase asymptotic efficiency. The criteria for redundancy may help

in particular examples to show the optimality of an instrument; compare, for instance, Dhaene's (2004) and Carrasco's (2004) solutions of the problem by Anatolyev (2003a).

Unit roots and nonstationarity When the error term is conditionally homoskedastic and has a finite number of isolated unit roots in its moving average representation, Hansen, Heaton and Ogaki (1988) show that it is still possible to attain the efficiency bound, and, consequently, construct the optimal instrument. The complication involved is the following. The unit root in the polynomial $\Theta(L)$ of the Wold decomposition $e_t = \Theta(L)\varepsilon_t$ makes the coefficients implied by $\Theta(L)^{-1}$ and $\Theta(L^{-1})^{-1}$ not converge to zero; however, the right hand side of (7) may still converge to a well defined random variable if the projections of future x_t onto $\Im_t$ fall sufficiently fast. However, the optimal instrument ζ_t will not be well defined because of further filtration by $\Theta(L)^{-1}$. Hansen, Heaton and Ogaki (1988, section 4.2) show that the efficiency bound is attained if a truncated version of the polynomial $\Theta(L)^{-1}$ is used. Under conditional heteroskedasticity, this issue is yet unexplored. Although the unconditional unit root is of no concern, a similar problem arises when all candidates for $\Phi_t(L)$ in (8) do not satisfy the stability condition.

Bates and White (1988) extend the econometrics of efficiency bounds to "heterogenous", i.e. nonstationary but ergodic, data. Of course, some amount of heterogeneity in the data is allowed as long as the conventional asymptotic theory still applies.

Semiparametric efficiency bounds Suppose the true conditional probability distribution is parameterized subject to the moment restrictions implied by the available set of instruments, and the maximum likelihood estimation is applied to this system. The least upper bound for the asymptotic variance of such an estimator for β is named the *semiparametric efficiency bound*. The fact is that in many circumstances this bound exactly equals the asymptotic variance of the optimal IV estimator. This is proved in Chamberlain (1987) for general nonlinear conditional moment restrictions in an IID environment, and one may conjecture that this also holds in time series conditional problems with a martingale difference structure of the moment function. Hansen and Singleton (1991) argue that the relation must also hold in conditionally homoskedastic time series models with serial correlation, and Hansen, Heaton and Ogaki (1988) conjecture that the same must be also true in multiperiod problems under conditional heteroskedasticity.

Single equation vs. system estimation In the "parametric" approach to implementation of the optimal instrument, auxiliary parameterizations for dynamics of right hand side variables and basic instruments are assumed. A different approach would take these auxiliary parameterizations seriously and construct the optimal IV estimator for the system containing the equations of interest and the auxiliary equations. If the latter are correctly specified, the resulting estimator for the parameters of interest is expected to be asymptotically more efficient than the one designed for a single equation. However, if the

auxiliary equations are misspecified, not only asymptotic efficiency, but also consistency may be lost. In the context of a log-linear asset pricing model, Hansen and Singleton (1996) evaluate the asymptotic efficiency gains from the system optimal IV estimator (which is asymptotically equivalent to the normal MLE) over the single equation optimal IV estimator. These gains turns out to be very small.

Panel data The Hayashi and Sims (1983) idea of forward filtering has found its way to some methods of estimation of panel data models. Keane and Runkle (1992) suggested removing possible serial correlation from the error by applying a forward filter, and using the instruments that are presumed weakly exogenous, to the transformed equation. The serial correlation in the error typically arises after first differencing. Schmidt, Ahn and Wyhowski (1992) show that the forward filtered IV estimator does not increase asymptotic efficiency in comparison with already known estimators as long as all available instruments are used, and that there exist more efficient IV estimators. Thus, the advantages of the proposed estimator may lie only in its possibly better finite sample properties. Besides, the forward filtering idea rests on the large-T asymptotics, which is not the case for usual short panels.

The authors mentioned above make assumptions that imply conditional homoskedasticity. Hence they exploit identification information only in expected cross-products of instruments and errors, thus considering the linear subclass of the space of potential instruments. In the context of more general relationship between errors and instruments, Chamberlain (1992) expands the subclass to take full advantage of conditional moment restrictions aiming at reaching the semiparametric efficiency bound. He exploits conditional forward filtering similar to that of Hansen, Heaton and Ogaki (1988), and suggests constructing the optimal instrument sequentially using nonparametric techniques for estimating conditional expectations. Hahn (1997) proposes a feasible estimator that attains the efficiency bound using an asymptotically expanding set of instruments.

Instrument selection and alternative asymptotics An alternative way to exploit information in an infinite set of instruments asymptotically efficiently is by using a portion of them for a fixed sample size but letting this portion expand asymptotically. This literature essentially approaches the issue of "too many instruments" by *picking* instruments from some set rather than *constructing* instruments as in the featured optimal instrument approach. Koenker and Machado (1999) establish an acceptable rate of this expansion; Andrews (1999) and Hall and Peixe (2003) develop methods to consistently form the "best" set of valid and relevant instruments. Hahn (2002) derives an efficiency bound when the number of instruments grows with the same rate as the sample size. Unfortunately, most of the work done concerns estimation in an IID environment.

Another strand of literature appeals to higher order asymptotic properties of IV estimators to select an instrument set from a larger list; see, for example, Donald and Newey

(2001) for IID models, and Kuersteiner (2006) for time series.

Empirical likelihood and related methods The method of maximum empirical likelihood (EL) is an alternative method of estimation of parameters in moment restriction based systems; see the overview in Imbens (2002). One of main advantages of EL over GMM is its ability to reduce the finite sample bias of GMM estimates, see Newey and Smith (2004) for IID data and Anatolyev (2005) for time series data. The adaptation to conditional moment restrictions is made in Donald, Imbens and Newey (2003), but only for an IID envoronment.

References

Amemiya, T. (1974) The Nonlinear Two-Stage Least-Squares Estimator. *Journal of Econometrics* 2, 105–110.

Anatolyev, S. (1999) Nonparametric estimation of nonlinear rational expectations models. *Economics Letters* 62, 1–6.

Anatolyev, S. (2002a) Autoregression and redundant instruments. *Econometric Theory* 18, Problem 02.6.2; 19, Solution 02.6.2.

Anatolyev, S. (2002b) Approximately optimal instrument for multiperiod conditional moment restrictions. Manuscript, New Economic School.

Anatolyev, S. (2003a) Redundancy of lagged regressors in a conditionally heteroskedastic time series regression. *Econometric Theory* 18, Problem 03.1.2.

Anatolyev, S. (2003b) The form of the optimal nonlinear instrument for multiperiod conditional moment restrictions. *Econometric Theory* 19, 602–609.

Anatolyev, S. (2005) GMM, GEL, serial correlation and asymptotic bias. *Econometrica* 73, 983–1002.

Andrews, D.W.K. (1999) Consistent Moment Selection Procedures for Generalized Method of Moment Estimation. *Econometrica* 67, 543–564.

Bates, C.E. and H. White (1988) Efficient Instrumental Variables Estimation of Systems of Implicit Heterogeneous Nonlinear Dynamic Equations with Nonspherical Errors. In: Barnett, Berndt and White, eds., *Dynamic Econometric Modeling*. Cambridge University Press: Cambridge.

Bates, C.E. and H. White (1993) Determination of Estimators with Minimum Asymptotic Covariance Matrices. *Econometric Theory* 9, 633–48.

Bougerol, P. and N. Picard (1992) Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20, 1714–1730.

Brandt, A. (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Advances in Applied Probability* 18, 211–220.

Breusch, T., Qian, H., Schmidt, P., and D. Wyhowski (1999) Redundancy of Moment Conditions. *Journal of Econometrics* 91, 89–111.

Broze, L., C. Francq and J.-M. Zakoïan (2001) Non-redundancy of high order moment conditions for efficient GMM estimation of weak AR processes. *Economics Letters* 71, 317–322.

Carrasco, M. (2004) Redundancy of lagged regressors in a conditionally heteroskedastic time series regression. *Econometric Theory* 19, Solution 03.1.2.

Carroll, R.J. (1982) Adapting for Heteroscedasticity in Linear Models. *Annals of Statistics* 10, 1224–1233.

Chamberlain, G. (1987) Asymptotic e ciency in estimation with conditional moment restrictions. *Journal of Econometrics* 34, 305–334.

Chamberlain, G. (1992) Comment: Sequential moment restrictions in panel data. *Journal of Business and Economic Statistics* 10, 20–26.

Dhane, G. (2004) Redundancy of lagged regressors in a conditionally heteroskedastic time series regression. *Econometric Theory* 19, Solution 03.1.2.

Dolado, J. (1990) Optimal instrumental variable estimator of the AR parameter of an ARMA(1,1) process. *Econometric Theory* 6, 117–119.

Donald, S.G., G.W. Imbens and W.K. Newey (2003) Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics* 117, 55–93.

Donald, S.G. and W.K. Newey (2001) Choosing the number of instruments. *Econometrica* 69, 1161–1191.

Dunn, K. and K. Singleton (1986) Modeling the term structure of interest rates under nonseparable utility and durability of goods. *Journal of Financial Economics* 17, 27–55.

Eichenbaum, M.S., L.P. Hansen and K.J. Singleton (1988) A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *Quarterly Journal of Economics* 103, 51–78.

Epstein, L.G. and S.E. Zin (1991) Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of Political Economy* 99, 263–286.

Fama, E. (1975) Short-Term Interest Rates as Predictors of Inflation. *American Economic Review* 65, 269–282.

Fan, J. and A. Ullah (1999) Asymptotic nomality of a combined regression estimator. *Journal of Multivariate Analysis* 85, 191–240.

Ferson, W.E. and G.M. Constantinides (1991) Habit persistence and durability in aggregate consumption. *Journal of Financial Economics* 29, 199–240.

Ghysels, E. and A.R. Hall (2002) Interview with Lars Peter Hansen. *Journal of Business and Economic Statistics* 20, 442–447.

Glad, I.K. (1998) Parametrically guided non-parametric regression. *Scandinavian Journal of Statistics* 25, 649–668.

Gordin, M.I. (1969) The central limit theorem for stationary processes. *Soviet Mathematics Doklady* 10, 1174–1176.

Grossman, S., Melino, A., Shiller, R. J. (1987). Estimating the Continuous Time Consumption-Based Asset Pricing Model. *Journal of Business and Economic Statistics* 5, 315–327.

Hall, A.R. (2005) *Generalized Method of Moments.* Oxford University Press.

Hall, A.R. and F.P.M. Peixe (2003) A Consistent Method for the Selection of Relevant Instruments. *Econometric Reviews* 22, 269–287.

Hall, R.E. (1988) Intertemporal substitution in consumption. *Journal of Political Economy* 96, 339–357.

Hahn, J. (1997) Efficient Estimation of Panel Data Models with Sequential Moment Restrictions. *Journal of Econometrics* 79, 1–21.

Hahn, J. (2002) Optimal inference with many instruments. *Econometric Theory* 18, 140–168.

Hansen, B. and K.D. West (2002) Generalized method of moments and macroeconomics. *Journal of Business and Economic Statistics* 20, 460–469.

Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029–1054.

Hansen, L.P. (1985) A method for calculating bounds on the asymptotic variance-covariance matrices of generalized method of moments estimators. *Journal of Econometrics* 30, 203–228.

Hansen, L.P. (2002) Methods of Moments. In: N. J. Smelser and P. B. Bates , eds., *International Encyclopedia of the Social & Behavioral Sciences*, Elsevier Publishing.

Hansen, L.P., J.C. Heaton and M. Ogaki (1988) Efficiency bounds implied by multiperiod conditional moment restrictions. *Journal of the American Statistical Association* 83, 863–871.

Hansen, L.P and R.J. Hodrick (1980) Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis. *The Journal of Political Economy* 88, 829–853.

Hansen, L.P. and K.J. Singleton (1991) Computing semiparametric efficiency bounds for linear time series models. In: Barnett, Powell and Tauchen, eds., *Semiparametric and Nonparametric Methods in Statistics and Econometrics*, Cambridge University Press.

Hansen, L.P. and K.J. Singleton (1996) Efficient estimation of linear asset pricing models with moving-average errors. *Journal of Business and Economic Statistics* 14, 53–68.

Harris, D. (1999) GMM Estimation of Time Series Models. In: László Mátyás, ed., *Generalized Method of Moments Estimation*, Cambridge University Press.

Hayashi, F., and C. Sims (1983) Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments. *Econometrica* 51, 783–798.

Holman, J. (1998) GMM estimation of a money-in-the-utility-function model: The implications of functional forms. *Journal of Money, Credit, and Banking* 30, 679–698.

Hotz, V.J., F.E. Kydland and G.L. Sedlacek (1988) Intertemporal preferences and labor supply. *Econometrica* 56, 335–360.

Imbens, G. (2002) Generalized method of moments and empirical likelihood. *Journal of Business and Economic Statistics* 20, 493–506.

Keane, M.P. and D.E. Runkle (1992) On the estimation of panel data models with serial correlation when instruments are not strictly exogenous. *Journal of Business and Economic Statistics* 10, 1–10.

Kim, Y., H. Qian and P. Schmidt (1999) Efficient GMM and MD estimation of autoregressive models. *Economics Letters* 62, 265–270.

Kocherlakota, N.R. (1990) On tests of representative consumer asset pricing models. *Journal of Monetary Economics* 26, 285–304.

Koenker, R. and J.A.F. Machado (1999) GMM Inference When the Number of Moment Conditions is Large. *Journal of Econometrics* 93, 327–344.

Kuersteiner, G.M. (2001) Optimal instrumental variables estimation for ARMA models. *Journal of Econometrics* 104, 359–405.

Kuersteiner, G.M. (2002) Efficient IV estimation for autoregressive models with conditional heteroskedasticity. *Econometric Theory* 18, 547–583.

Kuersteiner, G.M. (2006) Moment Selection and Bias Reduction for GMM in Conditionally Heteroskedastic Models. In: D. Corbea, S. Durlauf and B.E. Hansen, eds. *Econometric Theory and Practice: Frontiers of Analysis and Applied Research: Essays in Honor of Peter C.B. Phillips*, Cambridge University Press.

Linton, O. and E. Mammen (2005) Estimating semiparametric ARCH(∞) models by kernel smoothing methods. *Econometrica* 73, 771–836.

Mankiw, N.G., J.J. Rotemberg and L.H. Summers (1985) Intertemporal substitution in macroeconomics. *Quarterly Journal of Economics* 100, 225–251.

Marshall, D.A. (1992) Inflation and asset returns in a monetary economy. *Journal of Finance* 47, 1315–1342.

Meddahi, N. and E. Renault (2004) Temporal aggregation of volatility models. *Journal of Econometrics* 119, 355–379.

Meddahi, N., E. Renault and B. Werker (2003) GARCH and Irregularly Spaced Data. CIREQ and Université de Montréal.

Mishkin, F. (1990) What Does the Term Structure Tell Us About Future Inflation? *Journal of Monetary Economics* 25, 77–95.

Mishra, S., L. Su and A. Ullah (2004) Combined estimator of time series conditional heteroskedasticity. Presented at Econometric Society World Congress in London, 2005.

Newey, W.K. (1990) Efficient Instrumental Variables Estimation of Nonlinear Models. *Econometrica* 58, 809–837.

Newey, W.K. (1993) Efficient Estimation of Models with Conditional Moment Restrictions, in *Handbook of Statistics*, Vol. 11, Ch. 16, Elsevier Science, North Holland.

Newey, W.K. and R.J. Smith (2004) Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72, 219–255.

Pagan, A. and A. Ullah (1999) Nonparametric Econometrics. Cambridge University Press.

Pinkse, J. (2005) Heteroskedasticity correction and reinstrumenting. Working Paper, Pennsylvania State University.

Robinson, P. (1987). Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form. *Econometrica* 55, 875–891.

Schmidt, P., S.C. Ahn and D. Wyhowski (1992) Comment. *Journal of Business and Economic Statistics* 10, 10–14.

Smith, D.C. (1999) Finite sample properties of tests of the Epstein–Zin asset pricing model. *Journal of Econometrics* 93, 113–148.

Smith, R. (2005) Automatic Positive Semidefinite HAC Covariance Matrix and GMM Estimation. *Econometric Theory* 21, 158–170.

Stambaugh, R.F. (1993) Estimating Conditional Expectations when Volatility Fluctuates. NBER Working Paper #t0140.

Stoica, P., T. Söderström, and B. Friedlander (1985) Optimal instrumental variable estimates of the AR parameters of an ARMA process. *IEEE Transactions on Automatic Control* 30, 1066–1074.

Stoica, P., T. Söderström, and B. Friedlander (1987) Optimal instrumental variable multistep algorithms for estimation of the AR parameters of an ARMA process. *International Journal of Control* 45, 2083–2107.

Tauchen, G. (1986) Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data. *Journal of Business and Economic Statistics* 4, 397–416.

Tjøstheim, D. and B. Auestad (1994) Nonparametric Identification of Nonlinear Time Series: Selecting Significant Lags. *Journal of American Statistical Association* 89, 1410–1419.

Tschernig, R. and L. Yang (2000) Nonparametric Lag Selection for Time Series. *Journal of Time Series Analysis* 21, 457–487.

Uhlig, H. (1999) A toolkit for analyzing nonlinear dynamic stochastic models easily. In: R. Marimon and A. Scott, eds., *Computational Methods for the Study of Dynamic Economies*, 30–61, Oxford University Press.

Weber, C.E. (2000) "Rule-of-thumb" consumption, intertemporal substitution, and risk aversion. *Journal of Business and Economic Statistics* 18, 497–502.

Weber, C.E. (2002) Intertemporal non-separability and "rule of thumb" consumption. *Journal of Monetary Economics* 49, 293–308.

West, K.D. (2001) On optimal instrumental variables estimation of stationary time series models. *International Economic Review* 42, 1043–1050.

West, K.D. (2002) Efficient GMM estimation of weak AR processes. *Economics Letters* 75, 415–418.

West, K. D., K.-f. Wong and S. Anatolyev (2002) Instrumental variables estimation of heteroskedastic linear models using all lags of instruments. Working Paper, University of Wisconsin–Madison.

Working, H. (1960) Note on the Correlation of First Differences of Averages in a Random Chain. *Econometrica* 28, 916–918.