

NBER WORKING PAPER SERIES

ESTIMATION AND EVALUATION OF CONDITIONAL ASSET PRICING MODELS

Stefan Nagel  
Kenneth J. Singleton

Working Paper 16457  
<http://www.nber.org/papers/w16457>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
October 2010

We are grateful to seminar participants at Baruch College, the Berkeley-Stanford joint finance seminar, London Business School, Northwestern University, Princeton University, UC San Diego, the National Bureau of Economic Research, and the Western Finance Association Meetings, as well as to Foussemi Chabi-Yo, Wayne Ferson, Lars Hansen, the Editor, and two anonymous referees, for helpful comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2010 by Stefan Nagel and Kenneth J. Singleton. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Estimation and Evaluation of Conditional Asset Pricing Models  
Stefan Nagel and Kenneth J. Singleton  
NBER Working Paper No. 16457  
October 2010  
JEL No. G12

**ABSTRACT**

We find that several recently proposed consumption-based models of stock returns, when evaluated using an optimal set of managed portfolios and the associated model-implied conditional moment restrictions, fail to capture key features of risk premiums in equity markets. To arrive at these conclusions, we construct an optimal GMM estimator for models in which the stochastic discount factor (SDF) is a conditionally affine function of a set of priced risk factors. Further, for the (often relevant) case where a researcher is proposing a generalized SDF relative to some null model, we show that there is an optimal choice of managed portfolios to use in testing the null against the proposed alternative.

Stefan Nagel  
Stanford University  
Graduate School of Business  
518 Memorial Way  
Stanford, CA 94305  
and NBER  
nagel\_stefan@gsb.stanford.edu

Kenneth J. Singleton  
Graduate School of Business  
Stanford University  
Stanford, CA 94305  
and NBER  
ken@future.stanford.edu

There is a large and growing literature that explores the goodness-of-fit of dynamic asset pricing models in which the stochastic discount factor (*SDF*) takes the *conditionally* affine form  $m_{t+1}(\theta_0) = \phi_t^0(\theta_0) + \phi_t^{f'}(\theta_0)f_{t+1}$ , where  $f$  is the vector of observed “priced” risk factors, the factor weights  $(\phi_t^0, \phi_t^{f'})$  are in the modeler’s information set  $\mathcal{J}_t$ , and  $\theta_0$  is an unknown vector of parameters. *SDFs* of this form are implicit in conditional versions of the classical *CAPM* and its multifactor extensions (as posited, for example, in Fama and French (1996), Jagannathan and Wang (1996), and explored empirically in Hodrick and Zhang (2001)). They also arise from linearized consumption-based asset pricing models in which  $m_{t+1}$  is a representative agent’s marginal rate of substitution (e.g., Lettau and Ludvigson (2001b), and Santos and Veronesi (2006)).

To evaluate the fits of their candidate *SDFs*, researchers typically posit an  $R$ -vector of “test-asset” returns  $r_{t+1}$ , construct *GMM* estimators  $\theta_T$  of  $\theta_0$ , and then examine whether the test asset payoffs are correctly priced by the candidate *SDF*; that is, whether  $T^{-1} \sum_{t=1}^T (m_{t+1}(\theta_T)r_{t+1} - p)$  is close to zero, where  $p$  is an  $R$ -vector of prices. Based on these assessments, several candidate *SDFs* have been found to adequately describe the unconditional expected returns on common stocks. This lack of discrimination between models, some with very different economic underpinnings, is why Daniel and Titman (2006) and Lewellen, Nagel, and Shanken (2010), among others, have questioned the statistical power of extant tests.

A key premise of this paper is that considerable latitude remains for enhanced model discrimination by more efficiently exploiting the economic content of the dynamic pricing relation<sup>1</sup>

$$E[m_{t+1}(\theta_0)r_{t+1}|\mathcal{J}_t] = p. \tag{1}$$

Any model satisfying (1) must not only fit the cross-section of average returns, but also

the potentially more informative and demanding implied restrictions on the conditional moments of  $(m_{t+1}, r_{t+1})$ . We explore the fit of (1) by examining whether  $m_{t+1}(\theta_0)$ , evaluated at a *GMM* estimator  $\theta_T$  of  $\theta_0$ , reliably prices managed portfolio payoffs of the form  $B_t r_{t+1}$ , where  $B_t \in \mathcal{J}_t$  is a state-dependent matrix of portfolio weights.

Heuristically, assessments of whether a candidate *SDF* accurately prices the payoffs  $B_t r_{t+1}$  will be more reliable the more precise are the estimates of  $\theta_0$ . Yet in practice instrument selection for *GMM* estimation has not been tied to the specific formulation of the *SDF*, other than to include lagged values of returns, consumption growth, and other variables in  $\mathcal{J}_t$  that enter  $m_{t+1}$ . In this paper we draw upon the work of Hansen (1985) and Chamberlain (1987) to show that there is an *optimal* choice of instruments in the sense that the resulting *GMM* estimator has the smallest asymptotic covariance matrix among all admissible *GMM* estimators based on the conditional moment restrictions (1). Importantly, the optimal instruments are *not* lagged values of returns or of the variables comprising the *SDF*. Rather, we will show that they are nonlinear functions of the conditioning information  $\mathcal{J}_t$  that are related to the first and second moments of products of returns and factors,  $r_{t+1} f'_{t+1}$ , as suggested by the restrictions (1) on the conditional distribution of  $m_{t+1}(\theta_0) r_{t+1}$ .

Equipped with the efficient *GMM* estimator  $\theta_T^*$ , we proceed to construct chi-square goodness-of-fit tests based on the implication of (1) that a candidate *SDF* should price any pre-specified *M*-vector of managed payoffs  $B_t r_{t+1}$ :

$$E[m_{t+1}(\theta_0) B_t r_{t+1} - B_t p] = 0. \tag{2}$$

This approach enhances the *GMM*-based inference strategies used by Hodrick and Zhang (2001), Lettau and Ludvigson (2001b), and Roussanov (2009), among many

others, by using the asymptotically efficient estimator  $\theta_T^*$  of  $\theta_0$ .

Specializing further, we formalize the connection between maximal efficiency of the *GMM* estimator and maximal power of goodness-of-fit tests for the situation where a researcher is proposing a generalized *SDF*

$$m_{t+1}^{\mathcal{G}}(\theta_0) = \phi^0(z_t; \beta_0, \gamma_0) + \phi^{f'}(z_t; \beta_0, \gamma_0)f_{t+1}, \quad (3)$$

where  $z_t \in \mathcal{J}_t$ ,  $f_{t+1}$  is a vector of risk factors, and the null specification  $m_{t+1}^{\mathcal{N}}(\beta_0)$  is the nested special case with  $\gamma_0 = 0$ ;  $m_{t+1}^{\mathcal{N}}(\beta_0) = m_{t+1}^{\mathcal{G}}(\beta_0, 0)$ . Examples include the conditional consumption *CAPM* examined by Lettau and Ludvigson (2001b) ( $z_t = CAY_t$ ) where  $m_{t+1}^{\mathcal{N}}$  is the pricing kernel induced by constant relative risk averse preferences. Also included are the conditional *CAPMs* of Santos and Veronesi (2006) ( $z_t =$  the ratio of labor income to total income) and Jagannathan and Wang (1996) ( $z_t =$  the spread on high-yield bonds) where  $m_{t+1}^{\mathcal{N}}$  is the *SDF* induced by a classical *CAPM* in which expected returns are affine functions of their associated unconditional betas. Similarly, we subsume explorations of the economic significance of expanding the set of risk factors that are priced. This includes extensions of the conditional *CAPM* [e.g., the inclusion of returns to human capital in Jagannathan and Wang (1996)] or of the three-factor Fama and French (1992) model [e.g., the inclusion of momentum (Carhart (1997)) or liquidity (Pastor and Stambaugh (2003)) factors], as well as a linearized version of the model in Lustig and Van Nieuwerburgh (2006) with preferences defined over aggregate consumption and housing services.

We show that the *Wald* and Lagrange-multiplier (*LM*) tests of the null  $\gamma_0 = 0$  based on the optimal *GMM* estimator  $\theta_0^*$  are the (locally) most powerful chi-square tests against the alternative hypothesis that the pricing kernel is  $m_{t+1}^{\mathcal{G}}$ . Moreover, these op-

timal tests can be reinterpreted as tests of the null hypothesis  $E[B_t^*(m_{t+1}^{\mathcal{N}}(\beta_0)r_{t+1}-p)] = 0$ , for suitably chosen  $B_t^* \in \mathcal{J}_t$ . In this manner we derive an optimal set of managed portfolios  $B_t^*$  that maximize the power of our proposed chi-square tests of  $m_{t+1}^{\mathcal{N}}$  against the alternative  $m_{t+1}^{\mathcal{G}}$ . The portfolio weights  $B_t^*$  take an economically intuitive form: letting  $h_{t+1}(\theta_0) = (m_{t+1}^{\mathcal{G}}(\theta_0)r_{t+1} - p)$  denote the population pricing errors for the test asset returns  $r_{t+1}$ ,  $B_t^*$  is proportional to the component of  $E[\partial h_{t+1}(\theta_0)/\partial \gamma | \mathcal{J}_t]$ —the expected sensitivity of pricing errors to changes in the parameters governing the extended  $m_{t+1}^{\mathcal{G}}$ —that is conditionally orthogonal to its counterpart for the parameters  $\beta$  of the null specification,  $E[\partial h_{t+1}(\theta_0)/\partial \beta | \mathcal{J}_t]$ . Thus, the test statistics effectively check whether the pricing errors in the null model are forecastable using the incremental information contained in the additional factors of the generalized alternative model. Maximal power is achieved by using the optimal portfolio weights  $B_t^*$  and evaluating  $m_{t+1}$  at the efficient *GMM* estimator  $\theta_T^*$ .

The remainder of this paper is organized as follows. Section I reviews some of the key properties of conditional affine pricing models that will be needed in subsequent discussions. In Section II we outline the standard inference strategy of evaluating dynamic asset pricing models based on the pricing of managed portfolios as in (2). Then we construct optimal *GMM* estimators for conditionally affine *SDF*s. The characterization of the optimal choice of managed-portfolio weights  $B_t^*$  for maximizing the power of tests of  $m_{t+1}^{\mathcal{N}}$  against the alternative  $m_{t+1}^{\mathcal{G}}$  is developed in Section III.

We then turn to empirical implementations of our proposed methods in Sections IV and V. Two different constructions of the optimal instruments and portfolio weights are explored. One is a nonparametric estimation strategy in which we use local polynomial regressions to approximate conditional moments as a function of the source  $z_t$  of the state-dependence of the *SDF* weights  $\phi^f(z_t, \theta_0)$ . The other is a sieve method in

which we approximate conditional moments with a (global) polynomial function of  $z_t$ , consumption growth, and  $r_t$ . The results suggest that there are substantial gains in efficiency from using the optimal *GMM* estimator over other standard *GMM* estimators that have been used in previous studies. Additionally none of the models examined pass standard diagnostic chi-square tests when the test assets are portfolios sorted by firm size and book-to-market and conditional moment restrictions are used in estimation. While these models seemingly do quite well in fitting unconditional moments, the *SDF* parameter estimates at which the models produce these small *average* pricing errors imply counterfactual variation in conditional moments, which manifests itself as large and volatile *conditional* pricing errors. Model estimation and evaluation with conditional moment restrictions reveals that the models are unable to simultaneously fit the cross section and time series of asset returns.

Proofs as well as some Monte-Carlo evidence on the small-sample properties of the optimal *GMM* estimator are provided in the Internet Appendix.

## I Conditional Factor Models

A now standard approach to testing the cross-sectional implications of (1) is to assume that the pricing kernel has the conditionally affine structure (3), often with the factor weights  $\tilde{\phi}'_t = (\phi_t^0, \phi_t^{f'}) \in \mathcal{J}_t$  also being affine functions of an underlying vector of conditioning variables  $z_t$ . Letting  $\tilde{f}'_t = (1, f'_t)$  and “conditioning down” to the modeler’s information set  $\mathcal{J}_t$  leads to the following conditional “beta” representation of returns,<sup>2</sup>

$$E[r_{t+1}^i | \mathcal{J}_t] - r_t^f = \beta_{i,t}^{\mathcal{J}} \lambda_t^{\mathcal{J}}, \quad (4)$$

$$r_t^f = 1/E[m_{t+1}(\theta_0) | \mathcal{J}_t], \quad (5)$$

where  $\beta_{i,t}^{\mathcal{J}} = \text{Cov}(f_{t+1}, f'_{t+1} | \mathcal{J}_t)^{-1} \text{Cov}(f_{t+1}, r_{t+1}^i | \mathcal{J}_t)$  and  $\lambda_t^{\mathcal{J}} = -r_t^f \text{Cov}(f_{t+1}, \tilde{f}'_{t+1} | \mathcal{J}_t) \tilde{\phi}_t$ . Both  $\beta_{i,t}^{\mathcal{J}}$  and  $\lambda_t^{\mathcal{J}}$  are in general state-dependent, and  $\lambda_t^{\mathcal{J}}$  depends on the factor weights  $\phi_t$  when not all of the factors are returns or excess returns on traded portfolios. Therefore, many have followed Cochrane (1996) and imposed special structure on the pricing kernel that leads to a convenient *unconditional* factor model for returns.

Specifically, supposing that  $\tilde{\phi}_t$  is an affine function of  $z_t$ ,  $m_{t+1}$  can be expressed as

$$m_{t+1}(\theta_0) = \theta' f_{t+1}^{\#}. \quad (6)$$

The  $K \times 1$  vector of risk factors  $f_{t+1}^{\#}$  is built up from  $z_t$  and  $f_{t+1}$  and products of the elements of these vectors. Thus the pricing kernel can be thought of as arising from a  $K$ -factor model with constant factor weights (with factors that are dated both at dates  $t$  and  $t + 1$ ) and where  $K$  is larger (potentially much larger) than the number of factors in the underlying conditional model,  $F$ .

Furthermore, substituting (6) into  $E[h_{t+1}(\theta_0)] = 0$  gives the moment equations

$$E[\theta' f_{t+1}^{\#} r_{t+1}^i] = 1, \quad i = 1, \dots, R. \quad (7)$$

By the same reasoning leading to (4), but with  $\mathcal{J} = \emptyset$ , there exists a scalar  $\mu^0$  and constant  $K \times 1$  vectors  $\beta_i^{\#}$  and  $\lambda^{\#}$  such that

$$E[r_{t+1}^i] - \mu^0 = \beta_i^{\#} \lambda^{\#}, \quad i = 1, \dots, R, \quad (8)$$

where  $\beta_i^{\#} = \text{Cov}(f_t^{\#}, f_t^{\#\prime})^{-1} \text{Cov}(f_t^{\#}, r_t^i)$ , and  $\lambda^{\#} = -\mu^0 \text{Cov}(f_{t+1}^{\#}, m_{t+1})$ . Expression (8) imposes (relatively) easily testable restrictions on the cross-section of expected excess returns on the  $R$  test assets.



Tests based on the unconditional moment restriction (8) are omitting two potentially important sources of information about the validity of the underlying conditional asset pricing models. First the conditional moment restriction (1) leads to the expression (4) for conditional expected excess returns, with potentially state-dependent factor betas and market prices of risk. That is, potentially informative restrictions across the conditional first and second moments of the returns and risk factors are being omitted from assessments of goodness-of-fit. Second, implicit in (1) are the links between  $r_t^f$  and the conditional mean of  $m_{t+1}(\theta_0)^3$  (see (5)) and between  $\lambda_t^{\mathcal{J}}$ , the conditional second moments of  $f_{t+1}$ , and the factor weights  $\phi_t$  that determine the pricing kernel. When  $f_{t+1}$  is a vector of returns or excess returns on traded portfolios, then the latter restrictions imply a direct link between  $\lambda_t^{\mathcal{J}}$  and the excess returns on these portfolios.

A key premise of our analysis is that examination of the conditional pricing relations (4) and (5) jointly is potentially more revealing about the strengths and weaknesses of *SDFs* as descriptions of history, and about the features of *SDFs* that are needed to better match the historical, conditional distribution of returns. Examination of the joint restriction (4)-(5) is equivalent to examination of the conditional moment restriction (1). Thus, optimal tests based on (1) will be (asymptotically) at least as powerful as those based on (4), because the former incorporates more of the economic content of the conditional pricing model. Moreover, (1) embodies substantially more information than does the orthogonality of  $m_{t+1}$  and excess returns,  $E[m_{t+1}(\theta_0)(r_{t+1} - r_t^f) | \mathcal{J}_t] = 0$ . The latter expression implicitly relaxes the constraint (5) on the conditional mean of the pricing kernel and, hence, the scale of the pricing kernel cannot be identified.

## II Efficient *GMM* Estimation of Factor Models

Model assessment has frequently focused on whether a candidate *SDF*  $m_{t+1}(\theta_0)$  accurately prices the portfolio payoffs  $B_t r_{t+1}$ —that is, whether  $H_0 : E[B_t h_{t+1}(\theta_0)] = 0$  is satisfied—for a pre-specified set of managed portfolio weights  $B_t \in \mathcal{J}_t$ . This null hypothesis cannot be examined directly, because  $\theta_0$  (and hence  $B_t h_{t+1}(\theta_0)$ ) is unknown. Standard practice is to first construct a *GMM* estimator  $\theta_T$  of  $\theta_0$ , and then use the sample mean of  $\{B_t h_{t+1}(\theta_T)\}$  to construct a chi-square test of  $H_0$ . Owing to the first-stage estimation of  $\theta_0$ , this inference strategy involves the joint hypothesis that  $B_t r_{t+1}$  is accurately priced by  $m_{t+1}(\theta_0)$  and that the moment conditions underlying the construction of the *GMM* estimator of  $\theta_0$  are satisfied. Accordingly, we begin our discussion of the estimation of  $\theta_0$  by briefly reviewing the large-samples properties of chi-square tests constructed in this manner.

Suppose that a *GMM* estimator of the  $K$ -dimensional vector of unknown parameters  $\theta_0$  governing the *SDF* is constructed from the moment condition<sup>4</sup>

$$E[A_t h_{t+1}(\theta_0)] = 0, \tag{9}$$

for some  $K \times R$  matrix  $A_t$  with entries in  $\mathcal{J}_t$ . Since (9) constitutes  $K$  equations in the  $K$  unknowns  $\theta_0$ , we can define the *GMM* estimator  $\theta_T^A$  of  $\theta_0$ , indexed by the modeler's choice of instrument process  $\{A_t\}$ , as the value of  $\theta$  that solves

$$\frac{1}{T} \sum_{t=1}^T A_t (m_{t+1}(\theta_T^A) r_{t+1} - p) = \frac{1}{T} \sum_{t=1}^T A_t h_{t+1}(\theta_T^A) = 0. \tag{10}$$

Under regularity, the asymptotic covariance matrix of  $\theta_T^A$  is (Hansen (1982))

$$\Omega_0^A = E \left[ A_t \frac{\partial h_{t+1}(\theta_0)}{\partial \theta} \right]^{-1} \Sigma_0^A E \left[ \frac{\partial h_{t+1}(\theta_0)'}{\partial \theta} A_t' \right]^{-1}, \quad (11)$$

where<sup>5</sup>

$$\Sigma_0^A = E[A_t h_{t+1}(\theta_0) h_{t+1}(\theta_0)' A_t']. \quad (12)$$

With the *GMM* estimator in hand, assessment of whether a candidate *SDF* accurately prices the payoffs  $B_t r_{t+1}$  typically involves the computation of a chi-square statistic based on the sample pricing errors

$$\frac{1}{T} \sum_{t=1}^T B_t (m_{t+1}(\theta_T^A) r_{t+1} - p) = \frac{1}{T} \sum_{t=1}^T B_t h_{t+1}(\theta_T^A). \quad (13)$$

In the Internet Appendix A we show that

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T B_t h_{t+1}(\theta_T^A) \xrightarrow{\mathcal{D}} N(0, \Gamma_0^A), \quad \Gamma_0^A = E[C_t^A \Sigma_t C_t^{A'}], \quad (14)$$

where  $\xrightarrow{\mathcal{D}}$  denotes convergence in distribution,  $\Sigma_t = E[h_{t+1}(\theta_0) h_{t+1}(\theta_0)' | \mathcal{J}_t]$ , and

$$C_t^A = B_t - E \left[ B_t \frac{\partial h_{t+1}(\theta_0)}{\partial \theta} \right] E \left[ A_t \frac{\partial h_{t+1}(\theta_0)}{\partial \theta} \right]^{-1} A_t. \quad (15)$$

The form of  $C_t^A$  reflects the fact that pre-estimation of  $\theta_0$  using the instruments  $A_t$  affects the asymptotic distribution of the sample mean (13). It follows that

$$\tau_T(B, A) \equiv \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T h_{t+1}(\theta_T^A)' B_t' \right) (\Gamma_T^A)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T B_t h_{t+1}(\theta_T^A) \right) \quad (16)$$

$$\stackrel{a}{=} \left( \frac{1}{\sqrt{T}} \sum_t h_{t+1}(\theta_0)' C_t^{A'} \right) (\Gamma_T^A)^{-1} \left( \frac{1}{\sqrt{T}} \sum_t C_t^A h_{t+1}(\theta_0) \right), \quad (17)$$

where  $\stackrel{a}{\approx}$  means “asymptotically equivalent to.” By standard arguments  $\tau_T(B, A) \xrightarrow{\mathcal{D}} \chi^2(M)$ , where the degrees of freedom  $M$  is determined by the row dimension of the test matrix  $B_t$ .

The joint nature of the null hypothesis that is effectively being tested with the statistic  $\tau(B, A)$  is immediately apparent from (17). For  $\tau(B, A)$  to have an asymptotic chi-square distribution, it must be the case that

$$H_0 : E \left[ \left( B_t - E \left[ B_t \frac{\partial h_{t+1}(\theta_0)}{\partial \theta} \right] E \left[ A_t \frac{\partial h_{t+1}(\theta_0)}{\partial \theta} \right]^{-1} A_t \right) h_{t+1}(\theta_0) \right] = 0. \quad (18)$$

The first part of this joint null is accurate pricing:  $E[B_t h_{t+1}(\theta_0)] = 0$ . The second piece,  $E[A_t h_{t+1}(\theta_0)] = 0$ , ensures that  $\theta_T^A$  is a consistent estimator of  $\theta_0$ . The sample counterpart of the left-hand side of (18) is (13), because  $\theta_T^A$  satisfies the first-order conditions (10). We subsequently exploit the dependence of the power function of this chi-square test on the choice of  $(A_t, B_t)$  to derive optimal choices of these matrices.

## A The Optimal GMM Estimator

If we index each estimator  $\theta_T^A$  by its associated instrument matrix  $A_t$ , then we can define the admissible class of GMM estimators as<sup>6</sup>

$$\mathcal{A} = \left\{ A_t \in \mathcal{J}_t, \text{ such that } E \left[ A_t \frac{\partial h_{t+1}(\theta_0)}{\partial \theta} \right] \text{ has full rank} \right\}. \quad (19)$$

Researchers have considerable latitude in selecting the sequence of matrices  $\{A_t\}$  to construct a consistent estimator of  $\theta_0$ . Elements of  $A_t$  are typically built up from linear combinations of lagged returns, consumption growth rates, or other macroeconomic constructs underlying the pricing kernel. We seek the choice of  $A_t \in \mathcal{A}$  that gives

rise to the asymptotically most efficient estimator of  $\theta_0$ . In so doing, we ensure that our estimator is at least as efficient as any *GMM* estimator based on a given set of instruments  $w_t$  of *any* dimension  $L$  and the associated  $L \times R$  orthogonality conditions  $E[h_{t+1}(\theta_0) \otimes w_t] = 0$ . This is because the sample moment conditions for any such “fixed-instrument” *GMM* estimator (Hansen and Singleton (1982)) can be written in the form of (10) for an appropriate choice of  $A_t \in \mathcal{A}$ .<sup>7</sup>

The most efficient *GMM* estimator is the one that produces the smallest  $\Omega_0^A$  by choice of  $\{A_t\} \in \mathcal{A}$ . Fortunately, the solution to this minimization problem has been characterized (for our case of errors that follow a martingale difference sequence) by Hansen (1985), Chamberlain (1987), and Hansen, Heaton, and Ogaki (1988). Specifically, the optimal choice is

$$A_t^* = \Psi_t^{\theta'} \Sigma_t^{-1}, \text{ where } \Psi_t^\theta \equiv E \left[ \frac{\partial h_{t+1}(\theta_0)}{\partial \theta} \middle| \mathcal{J}_t \right], \quad (20)$$

and the associated asymptotic covariance matrix is

$$\Omega_0^* = (E [\Psi_t^{\theta'} \Sigma_t^{-1} \Psi_t^\theta])^{-1}. \quad (21)$$

The first term in the definition of  $A_t^*$ ,  $\Psi_t^{\theta'}$ , captures the sensitivity of  $h_{t+1}(\theta_0)$  to changes in the parameters. Since, in general,  $\partial h_{t+1}(\theta_0)/\partial \theta \notin \mathcal{J}_t$ , the role of the conditional expectation is to project these partial derivatives onto the econometrician’s information set (thereby giving admissible instruments).<sup>8</sup> The post-multiplication by  $\Sigma_t^{-1}$  serves to adjust for conditional heteroskedasticity, in a manner exactly analogous to the scaling of both regressors and errors in the implementation of *GLS* estimators.

Though at first glance the structure of  $A_t^*$  may appear to be intractable,<sup>9</sup> for models with conditionally affine pricing kernels of the form (3), the building blocks of  $A_t^*$  take

tractable forms. Specifically, writing  $m_{t+1}(\theta_0) = \tilde{\phi}(z_t, \theta_0)' \tilde{f}_{t+1}$ , a typical element of the first term in (20) takes the form

$$E \left[ \frac{\partial h_{i,t+1}(\theta_0)}{\partial \theta_{0j}} \Big| \mathcal{J}_t \right] = \frac{\partial \tilde{\phi}(z_t, \theta_0)'}{\partial \theta_{0j}} E \left[ \tilde{f}_{t+1} r_{i,t+1} \Big| \mathcal{J}_t \right]. \quad (22)$$

The functional form of  $\tilde{\phi}(z_t, \theta_0)$  is known from the specification of the pricing kernel and, hence, so are its partial derivatives. Therefore computation of (22) involves computing the conditional moments of cross-products of asset returns  $r_{i,t+1}$  and the elements of  $\tilde{f}_{t+1}$ . When the factors themselves are excess returns, we are computing conditional first and second moments of returns. Otherwise we are computing the conditional first moment of returns, risk factors, and their cross-products. Similarly,

$$\begin{aligned} E \left[ h_{i,t+1}(\theta_0) h_{j,t+1}(\theta_0) \Big| \mathcal{J}_t \right] &= \tilde{\phi}(z_t, \theta_0)' E \left[ r_{i,t+1} r_{j,t+1} \tilde{f}_{t+1} \tilde{f}_{t+1}' \Big| \mathcal{J}_t \right] \tilde{\phi}(z_t, \theta_0) \\ &\quad - \tilde{\phi}(z_t, \theta_0)' E \left[ \tilde{f}_{t+1} r_{i,t+1} \Big| \mathcal{J}_t \right] - \tilde{\phi}(z_t, \theta_0)' E \left[ \tilde{f}_{t+1} r_{j,t+1} \Big| \mathcal{J}_t \right] + 1. \end{aligned} \quad (23)$$

The first term on the right-hand side of (23) requires the computation of conditional second moments of returns and cross fourth moments of returns and factors (conditional means of terms like  $r_{i,t+1} r_{j,t+1} f_{k,t+1} f_{\ell,t+1}$ ).

The tractability of implementing the optimal *GMM* estimator for conditionally affine pricing models warrants special emphasis. There is substantial evidence that fixed-instrument *GMM* estimators based on the orthogonality conditions  $E[h_{t+1}(\theta_0) \otimes w_t] = 0$  exhibit asymptotic bias as the number of moment conditions grows.<sup>10</sup> Intuitively, the sources of this bias are two-fold: (i) the need to pre-estimate the optimal distance matrix for two-step *GMM* estimation, and (ii) the fact that the implied matrix  $A_t(\theta_T^\#)$  of instruments, evaluated at the first-stage estimator  $\theta_T^\#$ , may be correlated with the pricing errors  $h_{t+1}(\theta_T^A)$  evaluated at the second-stage *GMM* estimator (see,

e.g., Newey and Smith (2004)).

Our optimal *GMM* estimator avoids these sources of bias, because there is no first-stage estimation of a (potentially large) distance matrix. Moreover, once we have estimated the conditional moments of the data underlying the components of  $A^*$ , we proceed to find the  $\theta_T^*$  that solves the sample moment equations (10) with  $A_t = A_t^*$ . That is, we implement what is effectively a continuously-updated *GMM* estimator (Hansen, Heaton, and Yaron (1996)). It follows that, by construction,  $A_t^*(\theta_T^*)$  is orthogonal to  $h_{t+1}(\theta_T^*)$ , thereby removing a key source of bias in *GMM* estimation.

The conditionally affine structure of the pricing kernel also means that we have considerable latitude in specifying the functional form for the factor weight  $\tilde{\phi}(z_t, \theta_0)$ . Typically linearized versions of consumption-based pricing models assume that  $\tilde{\phi}(z_t)$  is an affine function of  $z_t$ . More generally, our approach to model evaluation applies without modification to cases where  $\tilde{\phi}(z_t)$  is a flexible function of  $z_t$ , represented for example using Hermite polynomials or Fourier approximations.

The dependence of  $A^*$  on conditional moments does raise the practical question of whether, in deriving the large-sample distribution of  $\theta_T^*$ , it is presumed that (a) the components of  $A_t^*$  (see (20)) are correctly specified, or (b) they are approximated with a scheme that becomes increasingly accurate as the sample size increases. The first case arises when a researcher adopts parametric models of  $\Psi_t^\theta$  and  $\Sigma_t$ . In this case, the asymptotic covariance matrix of  $\theta_T^*$  is (21).

The second case arises when either nonparametric or semi-nonparametric methods are used to estimate conditional moments. For a given degree of flexibility in the approximating scheme for the optimal instrument matrix  $A_t^*$ , our *GMM* estimators are consistent and asymptotically normal. Valid inference is possible even if our approximation scheme is not exact by relying on the robust version of the asymptotic covariance

in (11) (which is valid for a generic instrument matrix) instead of (21) (which presumes that the instrument matrix is equal to  $A_t^*$ ). To investigate the sensitivity of our empirical findings we consider two approximation schemes: local polynomial regression and a sieve method that uses a global polynomial approximation.

Evaluating  $\tau(B, A)$  in (16) at the optimal *GMM* estimator  $\theta_T^*$  gives

$$\tau_T(B, A^*) = \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T h_{t+1}(\theta_T^*)' B_t' \right) (\Gamma_T^{A^*})^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T B_t h_{t+1}(\theta_T^*) \right), \quad (24)$$

where  $\Gamma_T^{A^*}$  is a consistent estimator of  $\Gamma_0^{A^*} = E[C_t^{A^*} \Sigma_t C_t^{A^*}']$ . The robust version of this chi-square statistic uses a consistent estimator of  $\Gamma_0^A = E[C_t^A h_{t+1}(\theta_0) h_{t+1}(\theta_0)' C_t^{A'}]$  without presuming that  $h_{t+1}(\theta_0) h_{t+1}(\theta_0)'$  can be replaced by  $\Sigma_t$ .

## B The Wald Test with Maximal Power

Consider again the case where the goal is an evaluation of the improvement in fit of  $m_{t+1}^G(\beta_0, \gamma_0)$ , as given by (3), relative to the null specification  $m_{t+1}^N(\beta_0)$  obtained as the special case with  $\gamma_0 = 0$ . Suppose that  $\theta_0$  is estimated by *GMM* by solving the sample moment equations (10), for some sequence of  $K \times R$  instrument matrices  $\{A_t\}$  with  $A_t \in \mathcal{J}_t$ . Under regularity, the asymptotic covariance matrix of  $\theta_T^A$  is given by (11). Letting  $\Omega_{\gamma\gamma}^A$  denote the lower-diagonal  $G \times G$  block of  $\Omega_0^A$ , where  $G$  is the dimension of  $\gamma_0$ , it follows under  $H_0 : \gamma_0 = 0$  that

$$\varsigma_T^W(A) \equiv T \gamma_T' (\Omega_{\gamma\gamma}^A)^{-1} \gamma_T \xrightarrow{\mathcal{D}} \chi^2(G). \quad (25)$$

The power of the *Wald* test based on  $\varsigma_T^W(A)$  depends on the choice of instrument matrix  $A$ , consistent with our motivating heuristic that precision in estimation of  $\theta_0$



affects the power of tests of fit. In order to explore this dependence we focus on the *local* alternative  $H_{1T} : m_{t+1}^G(\beta_0, \gamma = \gamma_T^L)$ , for which the parameter sequence  $\gamma_T^L$  converges to the null of  $\gamma_0 = 0$  at the rate  $\sqrt{T}$ :  $\gamma_T^L = \delta/\sqrt{T}$ , for some nonzero  $G \times 1$  vector  $\delta$  of proportionality constants.<sup>11</sup> Under this local alternative,<sup>12</sup>  $\sqrt{T}(\gamma_T^A - \gamma_0) \xrightarrow{D} N(\delta, \Omega_{\gamma\gamma}^A)$ . It follows that the asymptotic distribution of  $\zeta_T^W(A)$  is that of a non-central chi-square distribution with  $G$  degrees of freedom and non-centrality parameter

$$\mathcal{NC}(A) = \delta' (\Omega_{\gamma\gamma}^A)^{-1} \delta. \quad (26)$$

The power of a chi-square test against a specific alternative is governed by the magnitude of the non-centrality parameter: the larger the value of  $\mathcal{NC}(A)$ , the more powerful is the test. An implication of (11) is that  $\mathcal{NC}(A)$  depends on the choice of instrument matrix  $A$  through the asymptotic covariance matrix of  $\gamma_T^A$ . The more econometrically efficient is the estimator  $\gamma_T^A$  of  $\gamma_0$ , the smaller is this covariance matrix and the higher is the power of the associated test based on  $\zeta_T^W(A)$ . Thus, we are led immediately to the conclusion that *GMM* estimation using the optimal instruments  $A_t^*$  gives the asymptotically (locally) most powerful *Wald* test of the null specification  $m_{t+1}^N$  against the alternative specification  $m_{t+1}^G$ .

### III Portfolio Selection for Maximal (Local) Power

Though the construction of the *Wald* statistic  $\zeta_T^W(A^*)$  might seem far removed from the discussion in the literature about how to best construct test portfolios in order to have power against alternative formulations of the pricing kernel, there is in fact an intimate connection to this issue. Indeed, tests based on  $\zeta_t^W(A^*)$  can be reinterpreted as tests based on an optimal set of test portfolios.

Specifically, using the superscript  $\mathcal{G}$  to indicate constructs evaluated at the unconstrained  $\theta_0$  governing  $m_{t+1}^{\mathcal{G}}$ , the *Wald* statistic  $\varsigma_T^W(A^*)$  can be expressed in the asymptotically equivalent form (see Internet Appendix B)

$$\varsigma_T^W(A^*) \stackrel{a}{=} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T h_{t+1}(\theta_0)' \Sigma_t^{\mathcal{G}-1} \mathcal{H}_t^{\mathcal{G}} \right) \Omega_{\gamma\gamma}^* \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathcal{H}_t^{\mathcal{G}'} \Sigma_t^{\mathcal{G}-1} h_{t+1}(\theta_0) \right), \quad (27)$$

where

$$\Psi_t^\gamma \equiv E \left[ \frac{\partial h_{t+1}(\beta_0, \gamma_0)}{\partial \gamma} \middle| \mathcal{J}_t \right], \quad \Psi_t^\beta \equiv E \left[ \frac{\partial h_{t+1}(\beta_0, \gamma_0)}{\partial \beta} \middle| \mathcal{J}_t \right],$$

$\mathcal{K}^{\beta\gamma} \equiv E \left[ \Psi_t^{\beta'} \Sigma_t^{-1} \Psi_t^\gamma \right]$ , and  $\mathcal{H}_t \equiv \Psi_t^\gamma - \Psi_t^\beta (\mathcal{K}^{\beta\beta})^{-1} \mathcal{K}^{\beta\gamma}$ . Asymptotic equivalence holds not only under  $H_0$  but under local alternatives as well.

An immediate implication of (27) is that the (locally) most powerful *Wald* test of  $H_0 : \gamma_0 = 0$  (against the alternative  $\gamma_0 \neq 0$ ) can be viewed as a test of

$$E \left[ \mathcal{H}_t^{\mathcal{G}'} \Sigma_t^{\mathcal{G}-1} h_{t+1}(\theta_0) \right] = 0; \quad (28)$$

that is, the *Wald* test evaluates whether the managed portfolio returns  $\mathcal{H}_t^{\mathcal{G}'} \Sigma_t^{\mathcal{G}-1} r_{t+1}$  are priced by  $m_{t+1}^{\mathcal{G}}$ . Factoring  $\Sigma_t^{-1}$  as  $D_t^{-1/2'} D_t^{-1/2}$ , the component  $D_t^{-1/2} \mathcal{H}_t^{\mathcal{G}}$  of the portfolio weights represents the part of  $D_t^{-1/2} \Psi_t^\gamma$  that is orthogonal to  $D_t^{-1/2} \Psi_t^\beta$ . Thus, it is as if  $E[D_t^{-1/2} \Psi_t^{\beta'} \Sigma_t^{\mathcal{G}-1} h_{t+1}(\theta_0)] = 0$  captures the economic content of the null specification  $m_{t+1}^{\mathcal{N}}$ , and the *Wald* test uses the part of  $D_t^{-1/2} \Psi_t^\gamma$  that is orthogonal to this null information to evaluate whether  $m_{t+1}^{\mathcal{G}}$  adds incrementally to pricing performance.

As an illustration of this optimality result, consider an extended consumption-based pricing kernel in which  $c_t$  denotes the logarithm of consumption and

$$m_{t+1}^{\mathcal{G}}(\theta_0) = (\beta_1 + \gamma_1 z_t) + (\beta_2 + \gamma_2 z_t) \Delta c_{t+1}. \quad (29)$$

The model in Lettau and Ludvigson (2001b), for example, is the special case with  $z_t$  equal to *cay*. These extensions add no explanatory power to the (linearized) consumption-based model with constant relative risk aversion if  $(\gamma_1, \gamma_2) = 0$ . For this setup,

$$E \left[ \frac{\partial h_{t+1}}{\partial \beta_1}(\theta_0) | \mathcal{J}_t \right] = E[r_{t+1} | \mathcal{J}_t], \quad E \left[ \frac{\partial h_{t+1}}{\partial \beta_2}(\theta_0) | \mathcal{J}_t \right] = E[\Delta c_{t+1} r_{t+1} | \mathcal{J}_t], \quad (30)$$

$$E \left[ \frac{\partial h_{t+1}}{\partial \gamma_1}(\theta_0) | \mathcal{J}_t \right] = E[r_{t+1} z_t | \mathcal{J}_t], \quad E \left[ \frac{\partial h_{t+1}}{\partial \gamma_2}(\theta_0) | \mathcal{J}_t \right] = E[\Delta c_{t+1} r_{t+1} z_t | \mathcal{J}_t], \quad (31)$$

where  $r_{t+1}$  is the vector of test assets used to estimate and evaluate the fit of the pricing model. Thus the optimal dynamic trading strategies are constructed using the components of the  $E[r_{t+1} z_t | \mathcal{J}_t]$  and  $E[\Delta c_{t+1} r_{t+1} z_t | \mathcal{J}_t]$  that are orthogonal (in a linear projection sense) to the information contained in  $E[r_{t+1} | \mathcal{J}_t]$  and  $E[\Delta c_{t+1} r_{t+1} | \mathcal{J}_t]$ .<sup>13</sup>

Our construction of optimal test portfolios differs from strategies typically employed in testing *unconditional* factor models based on the vector of pseudo-factors  $(z_t, \Delta c_{t+1}, \Delta c_{t+1} z_t)$  (see Section I) in several important respects. The construction of portfolio weights  $\mathcal{H}_t$  is explicitly linked to the contribution of new (pseudo) factors  $z_t$  and  $\Delta c_{t+1} z_t$  to the reduction in the model's pricing errors. In the sense made precise by the form of  $\mathcal{H}_t$  only the new information in these factors over and above what is already captured by the extant factor  $\Delta c_{t+1}$  is examined. Equally importantly, it is not the projection of the factors themselves onto  $\mathcal{J}_t$  that is relevant for portfolio construction, but rather the return-augmented projections  $E[r_{t+1} z_t | \mathcal{J}_t]$  and  $E[\Delta c_{t+1} r_{t+1} z_t | \mathcal{J}_t]$  are used. Among other considerations, this observation leads us to examine the conditional second moment  $E[\Delta c_{t+1} r_{t+1} | \mathcal{J}_t]$  when constructing  $\mathcal{H}_t$ . It is these interaction effects that tie  $\mathcal{H}_t$  to the model's pricing errors and lead to the dynamic test portfolios that maximize power against the proposed alternative model with  $(\gamma_1, \gamma_2) \neq 0$ .

As a second illustration, suppose that a researcher is interested in evaluating the

incremental contribution of a new risk factor  $f$  to the pricing of the test assets with returns  $r_{t+1}$ . A very simple version of this scenario has

$$m_{t+1}(\theta_0) = \beta_1 + \beta_2 \Delta c_{t+1} + \gamma_1 f_{t+1}. \quad (32)$$

For this example, the relevant expressions related to  $\beta_0$  are identical to (30) and

$$E \left[ \frac{\partial h_{t+1}}{\partial \gamma_1}(\theta_0) | \mathcal{J}_t \right] = E[r_{t+1} f_{t+1} | \mathcal{J}_t]. \quad (33)$$

Thus, the optimal dynamic test portfolio is constructed by examining the component of  $E[r_{t+1} f_{t+1} | \mathcal{J}_t]$  that is orthogonal to  $E[r_{t+1} | \mathcal{J}_t]$  and  $E[\Delta c_{t+1} r_{t+1} | \mathcal{J}_t]$ . Again this construction calls for an exploration of the conditional second-moment properties of the returns and risk factors (both  $\Delta c_{t+1}$  and the new factor  $f_{t+1}$ ).

## A Optimal Test Portfolios as Lagrange Multipliers

An alternative approach to deriving the optimal test portfolios starts with constrained estimates using  $m_{t+1} = m_{t+1}^{\mathcal{N}}$ , and then inquires whether adding additional risk factors or conditioning information in the factor weights improves pricing. This question can be addressed with the *LM* test.

In Internet Appendix C we show that the Lagrange multiplier for the constraints  $\gamma_T = 0$  can be expressed as

$$\lambda_T = \frac{1}{T} \sum_t \Psi_t' \Sigma_t^{\mathcal{N}-1} h_{t+1}^{\mathcal{N}}(\beta_T) \stackrel{a}{=} \frac{1}{T} \sum_t \mathcal{H}_t^{\mathcal{N}'} \Sigma_t^{\mathcal{N}-1} h_{t+1}^{\mathcal{N}}(\beta_0), \quad (34)$$

where  $\mathcal{H}_t^{\mathcal{N}}$  is the matrix  $\mathcal{H}_t$  evaluated at the constrained  $(\beta_0, \gamma_0 = 0)$ . Therefore, under  $H_0$ , the asymptotic distribution of  $\lambda_T$  is normal with mean zero and covariance matrix

$E[\mathcal{H}_t^{\mathcal{N}'} \Sigma_t^{\mathcal{N}-1} \mathcal{H}_t^{\mathcal{N}}]$ , from which it follows that

$$\varsigma_T^{LM}(A^*) = T \lambda_T' \left( \frac{1}{T} \sum_t \mathcal{H}_t^{\mathcal{N}'} (\beta_T^{\mathcal{N}}) \Sigma_t^{\mathcal{N}-1} (\beta_T^{\mathcal{N}}) \mathcal{H}_t^{\mathcal{N}} (\beta_T^{\mathcal{N}}) \right)^{-1} \lambda_T \xrightarrow{\mathcal{D}} \chi^2(G). \quad (35)$$

Summarizing our results,

$$\begin{aligned} \varsigma_T^W(A^*) & \text{ is asymptotically equivalent to } \tau(\mathcal{H}_t^{\mathcal{G}'}(\theta_0) \Sigma_t^{\mathcal{G}-1}(\theta_0), A^*) \\ \varsigma_T^{LM}(A^*) & \text{ is asymptotically equivalent to } \tau(\mathcal{H}_t^{\mathcal{N}'}(\beta_0) \Sigma_t^{\mathcal{N}-1}(\beta_0), A^*). \end{aligned}$$

Both tests effectively assess whether the managed portfolio returns  $\mathcal{H}_t^{\mathcal{G}'} \Sigma_t^{\mathcal{G}-1} r_{t+1}$  are correctly priced by  $m_{t+1}$ . The difference is that the (locally) most powerful, managed portfolio weights  $\mathcal{H}_t^{\mathcal{G}'} \Sigma_t^{\mathcal{G}-1}$  underlying the *Wald* test are evaluated at  $\theta_0$ , whereas the weights  $\mathcal{H}_t^{\mathcal{N}'} \Sigma_t^{\mathcal{N}-1}$  used to construct the *LM* statistic are evaluated at  $\gamma_0 = 0$ . It follows immediately that the *Wald* and *LM* statistics have the same asymptotic distribution under  $H_0$  and local alternatives.

## B Wald and *LM* Tests for “Completely” Affine *SDFs*

For the special case in which the factor weights  $\phi^0(z_t, \theta_0)$  and  $\phi^f(z_t, \theta_0)$  are affine functions of  $z_t$ ,<sup>14</sup> and thus  $m_{t+1}^{\mathcal{G}}$  can be expressed as a higher dimensional factor model with constant coefficients as in (6), the *sample* optimal *Wald* and *LM* tests take a particularly revealing form that further highlights the structure of the optimal portfolio weights. Since these representations hold exactly for the sample statistics, as contrasted with results for asymptotically equivalent expansions, they are useful for interpreting the subsequent empirical examples.

Assume that the *SDF* under the alternative can be expressed as

$$m_{t+1}^{\mathcal{G}}(\theta_0) = \beta_0' f_{t+1}^{\#\mathcal{N}} + \gamma_0' f_{t+1}^{\#\mathcal{G}}, \quad (36)$$

and  $m_{t+1}^{\mathcal{N}}(\beta_0)$  is again the special case of  $\gamma_0 = 0$ . With state-dependent weights on the actual risk factors  $f_{t+1}$ , the pseudo-factors  $f^{\#\mathcal{N}}$  and  $f^{\#\mathcal{G}}$  are composed of components of  $f_{t+1}$  and the conditioning variables  $z_t$  determining the factor weights, and their cross-products. Let  $(\hat{\Sigma}_t^{\mathcal{G}}, h_{t+1}^{\mathcal{G}}(\theta_T^{\mathcal{G}}), \theta_T^{\mathcal{G}})$  and  $(\hat{\Sigma}_t^{\mathcal{N}}, h_{t+1}^{\mathcal{N}}(\beta_T^{\mathcal{N}}), \beta_T^{\mathcal{N}})$  be the estimated conditional pricing error second moment matrix, realized pricing errors, and optimal *GMM* estimates when estimation is done under the alternative ( $\mathcal{G}$ ) and with the null  $\gamma_0 = 0$  ( $\mathcal{N}$ ) imposed.

Solving for the sample moment condition defining the optimal *GMM* estimate  $\theta_T^{\mathcal{G}}$  for the  $G$ -subvector  $\gamma_T^{\mathcal{G}}$  gives<sup>15</sup>

$$\begin{aligned} \gamma_T^{\mathcal{G}} &= [0, I_G] \left( \frac{1}{T} \sum_{t=1}^T \hat{\Psi}_t^{\theta'} \hat{\Sigma}_t^{\mathcal{G}-1} r_{t+1} f_{t+1}^{\#\mathcal{N}'} \right)^{-1} \frac{1}{T} \sum_{t=1}^T \hat{\Psi}_t^{\theta'} \hat{\Sigma}_t^{\mathcal{G}-1} p \\ &= \hat{\Omega}_{\gamma\gamma}^{\mathcal{G}} \frac{1}{T} \sum_{t=1}^T \hat{\mathcal{H}}_t^{\mathcal{G}}(\theta_T^{\mathcal{G}})' \hat{\Sigma}_t^{\mathcal{G}-1} p, \end{aligned}$$

where  $\hat{\mathcal{H}}_t^{\mathcal{G}}(\theta_T^{\mathcal{G}}) \equiv \hat{\Psi}_t^{\gamma'} - \hat{\mathcal{K}}_T^{\gamma\beta} (\hat{\mathcal{K}}_T^{\beta\beta})^{-1} \hat{\Psi}_t^{\beta'}$  and it is now understood that

$$\hat{\mathcal{K}}_T^{\gamma\beta}(\theta_T^{\mathcal{G}}) \equiv \frac{1}{T} \sum_{t=1}^T \left[ \hat{\Psi}_t^{\gamma'} \hat{\Sigma}_t^{\mathcal{G}-1} r_{t+1} f_{t+1}^{\#\mathcal{N}'} \right], \quad (37)$$

the robust, sample version of  $E[\Psi_t^{\gamma'} \Sigma_t^{\mathcal{G}-1} \Psi_t^{\beta}]$ , and similarly for  $\hat{\mathcal{K}}_T^{\beta\beta}(\theta_T^{\mathcal{G}})$ . Note that, for this completely affine setting, the matrices  $\hat{\Psi}_t^{\gamma}$  and  $\hat{\Psi}_t^{\beta}$  are the same whether they are

evaluated under the null or the alternative. Substitution into (25) gives

$$\varsigma_T^W = T \left( \frac{1}{T} \sum_{t=1}^T \widehat{\mathcal{H}}_t^{\mathcal{G}} \widehat{\Sigma}_t^{\mathcal{G}-1} p \right)' \left( \frac{1}{T} \sum_{t=1}^T \widehat{\mathcal{H}}_t^{\mathcal{G}} \widehat{\Sigma}_t^{\mathcal{G}-1} \widehat{\mathcal{H}}_t^{\mathcal{G}'} \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^T \widehat{\mathcal{H}}_t^{\mathcal{G}} \widehat{\Sigma}_t^{\mathcal{G}-1} p \right). \quad (38)$$

Now, as shown in Internet Appendix D, for a completely affine *SDF*,

$$\frac{1}{T} \sum_{t=1}^T \widehat{\mathcal{H}}_t^{\mathcal{G}} \widehat{\Sigma}_t^{\mathcal{G}-1} p = \frac{1}{T} \sum_{t=1}^T \widehat{\mathcal{H}}_t^{\mathcal{G}} \widehat{\Sigma}_t^{\mathcal{G}-1} h_{t+1}^{\mathcal{N}}(\beta_T^{\mathcal{N}}). \quad (39)$$

Thus, we can interpret the sample *Wald* statistic as checking whether the *SDF* under  $H_0$  prices the managed portfolios  $B_t^{Wald} = \widehat{\mathcal{H}}_t^{\mathcal{G}} \widehat{\Sigma}_t^{\mathcal{G}-1}$  evaluated at  $\theta_T^{\mathcal{G}}$ . Recall from Section A that the sample moment entering the *LM* statistic  $\varsigma_T^{LM}$  is<sup>16</sup>

$$\frac{1}{T} \sum_t \Psi_t^{\gamma'} \widehat{\Sigma}_t^{\mathcal{N}-1} h_{t+1}^{\mathcal{N}}(\beta_T^{\mathcal{N}}) = \frac{1}{T} \sum_{t=1}^T \widehat{\mathcal{H}}_t^{\mathcal{N}} \widehat{\Sigma}_t^{\mathcal{N}-1} h_{t+1}^{\mathcal{N}}(\beta_T^{\mathcal{N}}). \quad (40)$$

This expression is identical to (39), except that the managed portfolio weights  $B_t^{LM} = \widehat{\mathcal{H}}_t^{\mathcal{N}} \widehat{\Sigma}_t^{\mathcal{N}-1}$  are evaluated under the null at  $\beta_T^{\mathcal{N}}$ . Similarly the matrices that define the quadratic forms  $\varsigma_T^W$  and  $\varsigma_T^{LM}$  are identical, except again they are evaluated at  $\theta_T^{\mathcal{G}}$  and  $\beta_T^{\mathcal{N}}$ , respectively. Thus, to the extent that there are conflicts between these tests in evaluating the goodness-of-fit of an *SDF*, it is a consequence of the use of different estimates of the parameters to define the sample weights of the managed portfolios or the distance matrices in the quadratic forms. Both tests are constructed with identical pricing errors, namely those under  $H_0$ .

## IV Implementation: Methods and Data

In our empirical analysis, we consider several linearized consumption-based  $SDF$ s that have been proposed in the recent literature. The factor weights of each of these pricing kernels are affine functions of a (scalar) conditioning variable  $z_t$ ,

$$m_{t+1}^{\mathcal{G}}(\theta_0) = (\beta_1 + \gamma_1 z_t) + (\beta_2 + \gamma_2 z_t) \Delta c_{t+1}. \quad (41)$$

We consider three choices of  $z_t$ : the consumption-wealth ratio of Lettau and Ludvigson (2001a) ( $cay_t$ ), the corporate bond spread as in Jagannathan and Wang (1996) ( $def_t$ ), or the labor income-consumption ratio of Santos and Veronesi (2006) ( $yc_t$ ).<sup>17</sup>

Our sample period runs from 1952:2 to 2006:4, and we construct a quarterly log consumption growth series for this period from nondurables and services consumption, seasonally adjusted, per capita, and in 2000 chained dollars, as reported by the Bureau of Economic Analysis. We obtain a series of  $cay_t$  from Martin Lettau’s website. The  $def_t$  series is the spread in yields between Baa- and Aaa-rated bonds, obtained from the Federal Reserve Bank of St. Louis. Finally, following Santos and Veronesi (2006), we calculate  $yc_t$  using labor income defined as the labor income component of  $cay_t$  and with data from the Bureau of Economic Analysis.

The “primitive” returns that enter the construction of the portfolios with maximal power can be those on individual common stocks or portfolios of these stocks. While in principle it seems desirable to work with relatively disaggregated portfolios so that the nature of the  $SDF$  is central to determining the weights on the traded securities, computational considerations may lead one to partially aggregate assets into test portfolios and then to apply the optimal weights  $B_t^{Wald}$  or  $B_t^{LM}$  to the latter portfolios. To illustrate our methods we follow the latter approach and use the three-month Trea-



surely Bill and common stock portfolios sorted by firm size and book-to-market equity as test assets. More specifically, we choose the small-value, small-growth, large-value, and large-growth portfolios from the six portfolios of Fama and French (1993) as our equity test portfolios. Restricting the set of equity portfolios to these four allows us to keep the number of assets low (small  $R$ ), but still capture most of the cross-sectional variation in returns related to the “size” and “value” effects. Including a larger number of size and book-to-market portfolios would not add much additional return variation, due to the strong commonality in the returns of these portfolios (Fama and French (1993); Lewellen, Nagel, and Shanken (2010)). By construction of  $B_t^{Wald}$  and  $B_t^{LM}$ , we are asking candidate  $SDF$ s to explain not only the cross-section of unconditional moments of returns, but also their conditional moments.

We compound monthly stock portfolio returns to obtain quarterly returns from 1952:2 to 2006:4 (in tests that use lagged returns as instruments we also use returns from quarter 1952:1 as instruments). Nominal returns are deflated by the quarterly *CPI* inflation rate to obtain ex-post real returns. To distinguish how well the candidate models do in fitting the return on T-Bills and the return premia of stocks over and above T-Bill returns, we use returns in excess of T-Bill returns for the four equity portfolios (i.e., payoffs with a price of zero), and the gross real return for T-Bills (i.e., a payoff with price of one).

## A Estimation of Conditional Moments

Implementation of the optimal estimator requires estimates of the conditional moments

$$E \left[ \frac{\partial h_{t+1}(\theta_0)'}{\partial \theta_0} \middle| \mathcal{J}_t \right] = \frac{\partial \tilde{\phi}(z_t, \theta_0)'}{\partial \theta_0} E \left[ \begin{pmatrix} r'_{t+1} \\ \Delta c_{t+1} r'_{t+1} \end{pmatrix} \middle| \mathcal{J}_t \right]', \quad (42)$$

and

$$\text{Var} [h_{t+1}(\theta_0) | \mathcal{J}_t] = \tilde{\phi}(z_t, \theta_0)' \text{Var} \left[ \begin{pmatrix} r_{t+1} \\ \Delta c_{t+1} r_{t+1} \end{pmatrix} | \mathcal{J}_t \right] \tilde{\phi}(z_t, \theta_0), \quad (43)$$

where  $\partial \tilde{\phi}(z_t, \theta_0)' / \partial \theta_0 = (I_2 \otimes \tilde{z}'_t)$ ,  $\tilde{z}'_t = (1, z_t)$ , for the affine pricing kernels (41) that we consider here. In our empirical implementation, we work with  $\text{Var} [h_{t+1}(\theta_0) | \mathcal{J}_t]$  instead of the uncentered  $E [h_{t+1}(\theta_0) h_{t+1}(\theta_0)' | \mathcal{J}_t]$ . Both are equivalent under the null hypothesis, but the centered  $\text{Var} [h_{t+1}(\theta_0) | \mathcal{J}_t]$  should be better behaved under misspecification. To construct estimates of (42) and (43), we need estimates of the conditional moments  $E[(r'_{t+1}, \Delta c_{t+1} r'_{t+1})' | \mathcal{J}_t]$  and  $\text{Var}[(r'_{t+1}, \Delta c_{t+1} r'_{t+1})' | \mathcal{J}_t]$ . We use nonparametric local polynomial regression estimators of these moments, as well as a sieve method that uses a global polynomial approximation.<sup>18</sup>

Nonparametric estimators converge asymptotically, under regularity and as the flexibility of the approximating conditional moment functions increases with sample size, to the true moments conditional on  $\mathcal{J}_t$ . The downside is that computational considerations typically dictate that nonparametric estimation must focus on a small number of conditioning variables. In our implementation we restrict ourselves to just one conditioning variable. For each of the three pricing kernels, we condition moments on  $z_t$ , i.e., the conditioning variable  $cay_t$ ,  $def_t$ , or  $yc_t$  that appears in the pricing kernel. The dependence of the *SDF* weights on  $z_t$  means that, if these models are correctly specified, conditional moments of returns and consumption are likely to vary with  $z_t$ .

To estimate  $g(z_t) \equiv E[(r'_{t+1}, \Delta c_{t+1} r'_{t+1})' | z_t]$ , we run local linear regressions of the elements of  $y_{t+1} \equiv (r'_{t+1}, \Delta c_{t+1} r'_{t+1})'$  on  $z_t$ . Local linear regression has several desirable properties, including better behavior at the boundaries of the state space compared with fitting a local constant (Fan (1992)). To obtain the estimates  $\hat{g}(z_t)$  of the conditional mean function, a linear regression is estimated locally, with weighted least

squares in a fixed neighborhood around  $z_t$ , where the neighborhood is defined in terms of the distance  $|z_j - z_t|$ , not proximity in time. The weights are determined by the kernel function, the distance  $|z_j - z_t|$ , and the bandwidth  $b$ . The fitted value at  $z_t$  yields the conditional moment estimate  $\hat{g}(z_t)$ .

We use the Epanechnikov kernel function,

$$K(u) = \frac{3}{4} (1 - u^2) I(|u| \leq 1),$$

where  $u \equiv |z_j - z_t|/b$ . The bandwidth  $b$  determines the weighting of the neighborhood observations around each point  $z_t$ , and hence the smoothness of the estimated function. Regarding the choice of  $b$ , our experience from the simulations reported in Internet Appendix F suggests that in small samples the optimal *GMM* estimator is better behaved numerically when we impose a common bandwidth  $b_k$  for each pair  $y_{k,t+1} = (r'_{k,t+1}, \Delta c_{t+1} r'_{k,t+1})'$  corresponding to asset  $k$ . Effectively, this means that for each asset  $k$ , the two conditional moments in  $g_k(z_t) = E[(r'_{k,t+1}, \Delta c_{t+1} r'_{k,t+1})' | z_t]$  are estimated from the same local neighborhood around  $z_t$ . To determine the optimal bandwidth  $b_k^*$ , we use automatic bandwidth selection by leave-one-out cross-validation, i.e.,

$$b_k^* = \arg \min_{b_k} \frac{1}{T} \sum_{t=1}^T \{y_{k,t+1} - \hat{g}_{k,-t}(z_t)\}' V^{-1} \{y_{k,t+1} - \hat{g}_{k,-t}(z_t)\},$$

where  $\hat{g}_{k,-t}(z_t)$  denotes the local linear regression estimate of  $g_k(z_t)$  that is obtained with bandwidth  $b_k$  and with observation  $t$  excluded from the estimation.<sup>19</sup> The matrix  $V$  is diagonal, with the vector of sample variances of  $y_{k,t+1}$  on the diagonal. As  $T \rightarrow \infty$ , and more and more observations exist in the neighborhood of  $z_t$ , the optimal bandwidth shrinks, and the nonparametric regression estimates converge to the true conditional moments.

To estimate  $\Omega(z_t) \equiv \text{Var}[(r'_{t+1}, \Delta c_{t+1} r'_{t+1})' | z_t]$  we calculate the residuals  $y_{t+1} - \hat{g}(z_t)$  from the “first step” local regressions, and we use all elements of the cross-product matrix of these residuals as the dependent variables for “second step” local regressions. We make two modifications compared with the “first step” methodology to ensure that our estimated matrices  $\hat{\Omega}(z_t)$  are positive semi-definite: We fit a local constant instead of a local linear regression and we use a common bandwidth for all elements of  $\hat{\Omega}(z_t)$ . Fitting a local constant with a common bandwidth for all elements of  $\hat{\Omega}(z_t)$  is equivalent to estimating a sample covariance matrix in the usual way (albeit with weighted observations, and only those in a neighborhood of  $z_t$ ), which ensures positive semi-definiteness. Similar to the first-step estimation of  $g(z_t)$ , we also use an Epanechnikov kernel for  $\Omega(z_t)$ . The common optimal bandwidth is chosen according to a likelihood-type criterion as

$$b_{\Omega}^* = \arg \min_{b_{\Omega}} \frac{1}{T} \sum_{t=1}^T \left[ \{y_{t+1} - \hat{g}(z_t)\}' \hat{\Omega}_{-t}(z_t)^{-1} \{y_{t+1} - \hat{g}(z_t)\} + \log \left( \left| \hat{\Omega}_{-t}(z_t) \right| \right) \right],$$

where  $\hat{\Omega}_{-t}(z_t)$  denotes the estimate of  $\Omega(z_t)$  obtained with bandwidth  $b_{\Omega}$  and observation  $t$  omitted.

Figure 1 plots the nonparametric estimates of  $E[r_{t+1} | z_t]$  (a subvector of  $g(z_t)$ ), where  $z_t$  is set to  $cay_t$ ,  $def_t$ , and  $yc_t$  in the top, middle, and bottom graphs, respectively. The left-hand graphs depict the fitted conditional expected excess returns of the four stock portfolios, and the right-hand graphs show the fitted conditional expected gross return on the T-Bill. The relationships between  $cay_t$  and  $yc_t$  and the stock portfolio returns and the T-Bill return reveal some non-linearities. For  $def_t$ , the local polynomial regressions indicate only slight non-linearity. In this case, the estimated optimal bandwidths for the stock portfolio returns are sufficiently high so that

the local linear regression essentially turns into a globally linear regression.

[Figure 1 about here]

[Figure 2 about here]

Figure 2 plots the nonparametric estimates of  $E[\Delta c_{t+1} r_{t+1} | z_t]$  (a subvector of  $g(z_t)$ ). In this case there are pronounced non-linearities for all three conditioning variables.<sup>20</sup> While there are some cross-sectional differences in the relationships between returns and the predictors, most of the variation in the fitted conditional cross-products is common to the four stock portfolios.

Overall, the nonparametric regressions pick up considerable time-variation in conditional moments related to *cay*, *def*, and *yc*. This suggests that conditional moment restrictions constructed with these estimated conditional moments are likely to present a more serious challenge to the asset-pricing models than the restriction that the unconditional means of the pricing errors are zero.

Our nonparametric estimates for  $\Omega(z_t)$ , in contrast, do not pick up much time-variation. The bandwidth for  $\Omega(z_t)$  chosen by the optimal bandwidth selection algorithm is between three and four times the sample range of for all three predictors. This means that the estimated  $\Omega(z_t)$  is essentially the unconditional sample covariance matrix. Not surprisingly then, our subsequent asset-pricing results are virtually identical if one estimates  $\Omega(z_t)$  with the time-constant unconditional sample covariance matrix. The power of our optimal instruments estimator therefore derives mainly from time-variation in  $g(z_t)$ , i.e., from predictability of returns and cross-products of returns and consumption growth, not from time-variation in the higher moments captured by  $\Omega(z_t)$ .

As an alternative to the local polynomial estimates of conditional moments we

employ a sieve estimator that relies on a global polynomial approximation. For this construction we assume that  $E[r_{t+1}|\mathcal{J}_t]$  and  $E[r_{t+1}\Delta c_{t+1}|\mathcal{J}_t]$  have the functional forms of linear projections onto  $x_t \equiv (r_t, \Delta c_t, z_t, z_t^2, (z_t - \min(z_t) + 0.01)^{-1})$ .<sup>21</sup> For each of the elements of  $y_{t+1}$ , we use the Akaike Information Criterion (*AIC*) to select regressors. The regressor selection by *AIC* plays a similar role as optimal bandwidth estimation by cross-validation does in our local regression method. Both have the property that they would allow the approximation of conditional moments to become increasingly flexible with increasing sample size.

We use the sample covariance matrix of the residuals from these regressions to construct  $\text{Var}[(r'_{t+1}, \Delta c_{t+1}r'_{t+1})'|\mathcal{J}_t]$ . Thus, we assume that this conditional covariance matrix is constant. This assumption is motivated by the lack of evidence of time-variation in  $\Omega(z_t)$  in the local regression case discussed above, as well as a paucity of evidence for significant conditional heteroskedasticity in quarterly returns and consumption growth.<sup>22</sup>

While this sieve method is potentially less flexible in adapting to highly non-linear dependence on  $z_t$  than the local regression method, it allows us to condition on a broader set of instruments that includes  $(r_t, \Delta c_t)$ . The resulting estimates of  $E[r_{t+1}|\mathcal{J}_t]$  and  $E[r_{t+1}\Delta c_{t+1}|\mathcal{J}_t]$  capture well the linear, parabolic, and “S on its side” patterns displayed in Figures 1 and 2, but they also capture some additional variation in conditional moments due to the conditioning on lagged returns and consumption growth.

We emphasize again that, for valid inference, it is not necessary to assume that the approximation of  $A_t^*$  constructed from these conditional moment estimators perfectly matches the population counterpart  $A_t^*$ . In cases where one is concerned about the accuracy of these approximations in small samples, robust statistics should be used that are valid even if the approximation accuracy is poor (see Internet Appendix E).

## B Estimators and Test Statistics

We present results for four different estimators: One (denoted “unconditional”) is based on the  $R$  unconditional moment restrictions,

$$E[m_{t+1}(\theta_0) r_{t+1} - p] = 0, \quad (44)$$

where the elements of  $p$  are 1 for gross returns and 0 for excess returns. The second (denoted “fixed IV”) is based on the  $LR$  moment restrictions,

$$E[(m_{t+1}(\theta_0) r_{t+1} - p) \otimes w_t] = 0, \quad (45)$$

where  $w_t = (1, r'_t, \Delta c_t, z_t)'$  is an  $L \times 1$  vector, and  $z_t$  equals *cay*<sub>*t*</sub>, *def*<sub>*t*</sub>, or *yc*<sub>*t*</sub>, depending on the asset-pricing model. Our third estimator (denoted “optimal IV – local”) is our optimal *GMM* estimator, based on the  $K$  moment restrictions

$$E[A_t^*(m_{t+1}(\theta_0) r_{t+1} - p)] = 0, \quad (46)$$

and conditional moments estimated with local polynomial regressions. Finally, we let “optimal IV – sieve” denote the optimal *GMM* estimator that employs conditional moments estimated with the sieve method.

In the cases of the unconditional and fixed IV estimators, we iterate on the associated distance matrices until convergence. In the case of the optimal *GMM* estimators, we solve  $K$  equations in the  $K$  unknowns  $\theta_T$  with both  $A_t^*$  and  $m_{t+1}$  depending on  $\theta_T$  and, thus, this calculation is analogous to the continuously-updated *GMM* estimator. The discussion of the small-sample simulations in Internet Appendix F discusses some of the practical issues that can arise in the numerical solution of these equations.

[Table I about here]

For each of the choices of *GMM* estimator  $\theta_T^A$  we present three test statistics for model evaluation:  $\tau_T(I)$ , for the null hypothesis that the means of the “pricing errors” (44) or (45) are zero; and the *Wald* and *LM* statistics,  $\tau_T(B^{Wald})$  and  $\tau_T(B^{LM})$ , for the joint test that the *SDF* parameters  $\gamma_1 = 0$  and  $\gamma_2 = 0$ . All three of these statistics are variants of our general specification test based on a test matrix  $B_t$ ,

$$\tau_T(B, A) = \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T h_{t+1}(\theta_T^A)' B_t' \right) (\Gamma_T^A)^{-1} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^T B_t h_{t+1}(\theta_T^A) \right). \quad (47)$$

Table I summarizes the ingredients that enter into the calculation of the test statistics. Their construction differs depending on the estimator (unconditional, fixed IV, or optimal IV). For the unconditional and fixed IV estimators,  $\tau_T(I)$  represents Hansen’s *J*-test statistic. The statistics  $\tau_T(B^{Wald})$  and  $\tau_T(B^{LM})$  are calculated with unconditional moments for the unconditional and fixed IV estimators, and with conditional moments for the optimal IV estimator.

Our baseline standard errors and the test statistics are computed in their robust forms, without relying on the assumption that the conditional moments are correctly specified, but for the optimal IV estimators we also report results based on the latter assumption. Appendix E provides the details.

## V Implementation: Results

As a basis for comparing models with time-varying *SDF* factor weights, we start by estimating the constant-weight consumption *CAPM*, which is obtained by setting  $\gamma_1 = 0$  and  $\gamma_2 = 0$  in the pricing kernel (41). We focus on the conditioning variable



$z_t = cay_t$  as the estimators conditioned on  $def_t$  or  $yc_t$  give very similar results.

In the case of estimation based on unconditional moment restrictions, the estimated coefficient on consumption growth lies within the economically admissible region (Table II), but its magnitude is implausibly large in absolute value, 365. On the other hand, when estimation is based both on the cross-section of mean pricing errors and the models' restrictions on the conditional distributions of returns (fixed IV and optimal IV), the implied consumption risk premium is almost zero. This pattern is very similar to previous results from estimating consumption-based Euler equations with CRRA preferences. Grossman and Shiller (1981) find an unreasonably high relative risk aversion coefficient based on unconditional moment restrictions, while Hansen and Singleton (1982) work with conditional moment restrictions and obtain an estimate that is much closer to zero. Again, consistent with this prior literature, the test statistics  $\tau(I)$  constructed with all three estimators suggest that CRRA preferences fail to describe the real returns on common stocks and Treasury bills.

[Table II about here]

The results with time-varying  $SDF$  factor weights are displayed in Tables III, IV, and V for conditioning variables  $cay$ ,  $def$ , and  $yc$ , respectively. A common feature of the results for all three conditioning variables is that the standard errors of the  $SDF$  parameters are notably larger in the case of the unconditional estimator than for either the fixed IV or optimal IV estimators. This is reflected in the relatively small magnitudes of  $\tau_T(B^{Wald})$  and  $\tau_T(B^{LM})$  and the lack of evidence against the null hypothesis that  $(\gamma_1, \gamma_2) = 0$ , regardless of the choice of conditioning variable  $z_t$ , with the exception of  $\tau_T(B^{LM})$  for  $cay$ , which has a  $p$ -value of 0.02. Based on this evidence from the unconditional estimator, one would reasonably be led to conclude that one

cannot have much statistical confidence that the three enhanced consumption-based models improve pricing over and above the simpler model with *CRRA* preferences.

[Table III about here]

Substantially different estimates, with correspondingly smaller estimated standard errors, are obtained when conditioning information is used to construct the fixed IV and optimal *GMM* estimators. For the Lettau and Ludvigson (2001b) model in Table III with  $z_t = cay_t$ , the  $\tau_T(B^{Wald})$  and  $\tau_T(B^{LM})$  statistics provide some evidence to reject the null of the basic *CRRA* model in favor of the extended model, but more so for fixed IV than for optimal IV. With  $z_t = def_t$  and  $z_t = yc_t$  in Tables IV and V, the picture is also mixed, with some support for a rejection of  $(\gamma_1, \gamma_2) = 0$  with fixed IV and optimal IV - sieve, but not with optimal IV - local.

However this indication that conditioning the *SDF* on  $z_t$  may help in pricing the test assets must be interpreted with caution, because of the evidence from the overall goodness-of-fit statistic  $\tau_T(I)$ . For all three models, when conditioning information is incorporated in estimation, this statistic is large relative to its degrees of freedom, indicating failure of these models at conventional significance levels. Only in the case of  $z_t = cay_t$  and estimation based on unconditional moments does the evidence suggest that the pricing model adequately describes expected returns. In this case it appears to be a relative lack of power when estimation is based on unconditional moment restrictions, and not the actual success of the Lettau and Ludvigson (2001b) model, that explains their findings and ours.

The *Wald* and *LM* tests provide a complementary perspective in circumstances where power of overall goodness-of-fit tests may be an issue, as these tests may point to non-rejection of the simpler null model. This is what we find for the Lettau and Lud-

vigson (2001b) model with unconditional moment restrictions: The overall goodness-of-fit statistic  $\tau_T(I)$  does not reject the extended model, while at the same time the *Wald* test does not indicate that the extension of the model beyond the basic *CRRRA* model helps in pricing the test assets, consistent with a lack of power.

[Table IV about here]

Looking across the three models, the point estimates of the parameters based on the optimal IV - local and optimal IV - sieve estimators are quite close to each other, and the fixed IV estimates are also much closer to the optimal IV estimates than the unconditional ones. The finding that the fixed IV and optimal IV estimators produce results that are quite similar raises the question of under what circumstances the optimal IV estimator provides an efficiency gain over fixed IV estimators. In general, as in our specific application, this will depend on the choice of fixed instruments  $w_t$  (on their functional dependence on information in  $\mathcal{J}_t$ ).

To illustrate this sensitivity, recall that Lettau and Ludvigson (2001b) find that the fit of their model evaluated at the fixed IV estimator with  $w_t = (1, 1 + \frac{cay}{\sigma(cay)})'$  is comparable to the fit obtained with their unconditional estimator. A similar pattern appears in our data. The fixed IV estimator based on  $w_t = (1, 1 + \frac{cay}{\sigma(cay)})'$  yields  $\tau_T(I) = 18.48$  ( $p$ -value 0.01), which is much closer to the  $\tau_T(I)$  from the unconditional method than the  $\tau_T(I)$  from our baseline fixed IV results reported in Table III. Even though the fixed IV estimator with  $w_t = (1, 1 + \frac{cay}{\sigma(cay)})'$  conditions on the same information set as our optimal IV - local estimator, it appears to have much less power. Our strategy removes the arbitrariness of many past choices of  $w_t$  by directing attention to the choice that maximizes the (local) power of chi-square tests of fit.

In addition, even though our baseline fixed IV estimator produces *SDF* parameter

estimates that are close to those from the optimal IV estimators, the optimal estimators based on the sieve method (which use the same information set as the fixed IV estimator) often produce considerably smaller standard errors. This finding supports our premise that the incorporation of conditioning information in a manner that allows researchers to achieve the asymptotic efficiency bounds improves the reliability of estimation. The optimal IV - local estimator is more difficult to compare in this respect because it conditions on a smaller information set (only  $z_t$ ) than the fixed IV estimator.

Comparing the optimal *GMM* estimators based on the local regression and sieve methods, the similarity of the point estimates (relative to the unconditional estimates) is encouraging as there is some robustness to the precise specification of the model of the conditional moments. The lower standard errors from the sieve method could be an indication that conditioning  $E[(r'_{t+1}, \Delta c_{t+1} r'_{t+1})' | \mathcal{J}_t]$  on the history of past returns and consumption growth in addition to  $z_t$  leads to some additional efficiency gains.

It is also noteworthy that the difference between the robust standard errors and test statistics and those that assume correctly specified conditional moments is, in most cases, quite small, particularly relative to the differences in standard errors between the unconditional, fixed IV, and optimal IV estimators. This suggests that our methods of empirically approximating the conditional moments work reasonably well.

[Table V about here]

## A Conditional Pricing Errors

The main motivation for moving from simple constant-weight pricing kernels to models with time-varying weights is to obtain a more flexible asset-pricing model that is in

better accordance with the data, in the cross-section of unconditional moments, but also the time-series of conditional moments. So far the literature has focused mostly on examining the cross-section of *average* pricing errors, but Daniel and Titman (2006) and Lewellen, Nagel, and Shanken (2010) argue that this is not an informative criterion to judge these models. Examination of their *conditional* pricing errors is a natural alternative. Since our method involves explicit estimation of conditional moments, it provides a straightforward way of checking to what extent the *SDFs* estimated from unconditional moment restrictions, which produce a relatively good fit in the cross-section, actually achieve their promise of matching the conditional moment properties of the data, and how this picture changes when *SDFs* are estimated from conditional moment restrictions.

Figure 3 presents our estimates of the conditional pricing errors of the five “primitive” assets evaluated at the unconditional, fixed IV, and optimal IV - local estimators. In each case, the conditional moments are estimated with the local regression method. For the stock portfolio we look at what is perhaps the most interesting dimension: the spread between high and low B/M stocks. The plots on the left-hand side show the conditional pricing errors of a zero-investment portfolio that takes a long position in the two high B/M portfolios (each with weight one-half) and a short position in the two low B/M portfolios (each with weight one-half). The plots on the right-hand side show the conditional pricing error of the T-Bill.

The two plots in the top row illustrate that the pricing kernel estimated with unconditional moment restrictions and  $z_t = cay$  fails dramatically in matching time-variation in conditional moments. Conditional pricing errors for the high-low B/M portfolio vary between  $-0.1$  and  $0.4$ . Those for the T-Bill vary between  $-8$  and  $15$  (the most extreme peaks extend beyond the range shown in the figures). Given that the

T-Bill payoff has a constant price of 1.0, the magnitudes of this conditional mispricing is enormous. Similar patterns are evident, albeit less extreme, for  $z_t = def$  in the middle row. With  $z_t = yc$  in the bottom row, the magnitudes of the conditional pricing errors are relatively smaller, but still large in absolute terms, ranging from  $-0.05$  to  $0.05$  for the high-low B/M portfolio, and from  $-1.5$  to  $1.5$  for the T-Bill.

[Figure 3 about here]

[Figure 4 about here]

Employing conditional moment restrictions should help alleviate this mismatch between model-implied and actual variation in conditional moments. Indeed, the fixed IV and optimal IV estimates produce conditional pricing errors that are an order of magnitude smaller than those based on unconditional estimates for the stock portfolios, and several orders of magnitude smaller for the T-Bill. These IV estimators give nontrivial weight to conditional moments in estimation and, thereby, enforce consistency between the model-implied and sample conditional moments. It is important to note, though, that even for these IV estimators the conditional pricing errors are economically large. The models do not match the time-variation in the sample conditional moments. The  $SDF$  parameters we obtained with optimal IV imply a virtually constant  $SDF$  which does not help much to explain cross-sectional or time-series variation in returns. The reason why the conditional pricing errors are so much bigger with the unconditional  $SDF$  estimates is that these  $SDF$  estimates imply variation in conditional moments that is far greater than what is actually found in the data, which produces conditional pricing errors that are far in excess of what one would get by naively setting the pricing kernel to a constant, say 0.99.

Figure 4 compares the model-implied conditional pricing errors based on the two

optimal IV estimators with the axes scaled to reveal differences around zero. These optimal IV methods produce conditional pricing errors that are positively correlated with each other, but the errors from the sieve method exhibit more high-frequency variation. This is a consequence of our inclusion of lagged returns and lagged consumption growth in the conditioning set for the optimal IV-sieve estimator. In the models, the *SDF* weights vary with the relatively slow moving  $z_t$  variables. When, as with the optimal IV-sieve estimation, conditioning involves a richer information set, the limitations of the model are revealed through much greater short-run predictability of the model-implied pricing errors. If one takes the view that frictionless consumption-based asset-pricing models are not designed to explain such short-run predictability patterns, one might prefer to focus on the conditional pricing errors from the local regression method, which are conditioned only on  $z_t$ . For the T-Bill, any differences that exist between the two methods are small relative to the differences that exist between the errors based on unconditional and optimal IV estimators.

[Table VI about here]

The message from Figures 3 and 4 is also underscored by Table VI, which summarizes the time-series standard deviation (S.D.) of conditional pricing errors, and the cross-sectional root mean squared unconditional pricing errors (RMSE). As Panel A shows, the unconditional estimates with  $z_t = cay_t$  imply an enormous standard deviation of the conditional pricing errors, particularly for the T-Bill. Evidently, the model achieves a relatively good fit in the cross section at the unconditional moment restriction estimates, as in Lettau and Ludvigson (2001b), but at the price of producing wild swings in conditional pricing errors. Similar patterns, albeit somewhat less dramatic, exist in Panels B and C for  $z_t = def_t$  and  $z_t = yc_t$ . Evaluated at the unconditional

estimates, the models imply variation in conditional moments of asset returns far in excess of the variation that exists in the data. This pattern is consistent with the finding in Lewellen and Nagel (2006) that the pricing kernels estimated with unconditional moment restrictions and size- and book-to-market sorted equity portfolio returns imply excessive variation in conditional factor risk premia.

When conditioning information is introduced in estimation, variation in the conditional pricing errors shrinks, but the cross-sectional RMSE increases. Given that the motivation for models with time-varying pricing kernel weights is to match conditional moments of returns and factors, this inability to reconcile the cross section and time series of asset returns is an important failure of the model.

A key difference between the way the real returns on the T-bill and the stock portfolios enter our pricing relations is that the former enters as a gross return while the latter enter as excess returns. The model-implied price of a gross return is more sensitive to misspecification in the conditional mean of the pricing kernel than the model-implied price of an excess return, because

$$E[h_{t+1}|z_t] = E[m_{t+1}|z_t] E[r_{t+1}|z_t] + \text{Cov}[m_{t+1}, r_{t+1}|z_t] - 1.$$

Misspecification of  $E[m_{t+1}|z_t]$  has a much bigger effect on  $E[h_{t+1}|z_t]$  when  $r_{t+1}$  is 1 plus a return than when it is an excess return. This observation no doubt partially explains the finding that the T-Bill features the biggest differences in conditional pricing errors between the unconditional and the IV estimates. However it is not the T-bill *per se* that challenges these pricing kernels. We obtain similar results if we replace the gross return on the T-Bill with, for example, the gross return on a value-weighted stock market index. Rather, it is the fact that inclusion of a gross return (as contrasted with



working exclusively with excess returns) is informative about misspecification of the conditional mean of the *SDF*.

## B Time-variation of Estimated *SDF* Weights

An alternative way of evaluating the economic properties of these models is to examine the implied estimates of the time-varying pricing kernel weights,  $\phi_t^0 = \beta_1 + \gamma_1 z_t$  and  $\phi_t^f = \beta_2 + \gamma_2 z_t$ . We focus our discussion on  $\phi_t^f$ . Figure 5 plots the estimates of  $\phi_t^f$  with  $z_t$  equal to *cay*<sub>*t*</sub>, *def*<sub>*t*</sub>, or *yc*<sub>*t*</sub>.

The coefficient  $\phi_t^f$  has a close connection to the coefficient of relative risk aversion. Consider a constant-relative risk aversion pricing kernel,  $m_{t+1} = \delta_t \exp(-\gamma_t \Delta c_{t+1})$ , with time-varying relative risk aversion  $\gamma_t$  and time-discount factor  $\delta_t$ . Linearizing  $m_{t+1}$  around  $\Delta c_{t+1} = 0$ , we get  $m_{t+1} \approx \delta_t - \delta_t \gamma_t \Delta c_{t+1}$  or, in our notation,  $\phi_t^f = -\delta_t \gamma_t$ . For  $\delta_t$  close to one we get  $\phi_t^f \approx -\gamma_t$ , which means that we can interpret the plots in Figure 5 as plots of the (negative of the) estimated implied relative risk aversion coefficient. Clearly,  $\phi_t^f$  should then always be negative to make economic sense.

As an example of a *SDF* specification that produces strongly time-varying risk premia, the Campbell and Cochrane (1999) pricing kernel, linearized in a similar way, implies that the weight  $\phi_t^f$  should equal  $-\gamma [1 + \lambda(s_t)]$ , where  $\lambda(s_t)$  is the (state-dependent) sensitivity of habit to consumption (see Campbell and Cochrane's Eq. (5)). Note that  $\lambda(s_t)$  is always strictly positive in their specification, hence  $\phi_t^f$  should always be negative (at least if we ignore the approximation error in the linearization). Judging from Campbell and Cochrane's Figure 1,  $\lambda(s_t)$  is in the range of [0, 50]. Setting  $\gamma = 2$ , as in their calibrations, we get magnitudes for  $\phi_t^f \in [-100, 0]$ .

[Figure 5 about here]

[Figure 6 about here]

Focusing first on the estimates based on unconditional moment restrictions (the top graph in Figure 5), the estimates of  $\phi_t^f$  for the model with  $z_t = cay_t$  wander far outside the region of economic plausibility. Most of the time the estimates are greater than zero, implying negative relative risk aversion, and they vary far more than the range  $[-100, 0]$  suggested by the Campbell-Cochrane model (see, also, the calculations in Section 5 of Lewellen and Nagel (2006)). Consistent with our earlier analysis of conditional pricing errors, this shows that the model achieves its relatively good fit in the cross section by making risk premia counter-factually volatile. When  $z_t = def_t$  or  $z_t = yc_t$ , the estimates of  $\phi_t^f$  are much less volatile, always negative, but still outside the  $[-100, 0]$  interval, with values around  $-150$  for  $z_t = def_t$  and  $-300$  for  $z_t = yc_t$ .

Using the fixed IV estimator, as shown in the middle graph, reduces the volatility of  $\phi_t^f$  for  $z_t = cay_t$  by several orders of magnitude, but the estimated  $\phi_t^f$  are still often positive. The corresponding estimates for the model with  $z_t = yc_t$  are also much closer to zero, but are now also sometimes positive. The most volatile  $\phi_t^f$  is obtained with  $z_t = def_t$ . The statistical significance of these patterns is weak, however, as the coefficients on  $def_t$  and  $def_t \times \Delta c_{t+1}$  are estimated with relatively high standard errors (see Table IV).

Using the optimal IV-local estimator, the estimated  $\phi_t^f$  exhibit relatively little variation over time, and are close to or within the  $[-100, 0]$  range for all three for all three choices of  $z_t$ . With the sieve method, shown in Figure 6, the optimal IV estimates closely resemble those obtained with fixed IV.

Finally, it is also useful to note that the *SDFs*  $m_{t+1} = \phi_t^0 + \phi_t^f \Delta c_{t+1}$  implied by the optimal IV-local estimates (not shown) are positive throughout the entire sample for all three conditioning variables, with only a few exceptions for  $z_t = def_t$ . With

optimal IV-sieve, the estimated  $m_{t+1}$  is always greater than zero and ranges between 0.98 and 1.01. In contrast, the  $SDF$  implied by the estimates from unconditional moment restrictions frequently takes large negative values.

## VI Concluding Remarks

We explore the use of conditional moment restrictions in estimation and evaluation of asset pricing models in which the  $SDF$  is a conditionally affine function of a set of risk factors. We make two methodological advances. First, we develop and implement an optimal  $GMM$  estimator for this class of models. We thus provide some guidance in choosing from the large array of possible instruments when setting up  $GMM$  estimators. Second, we show that there is an optimal choice of managed portfolios to use in testing a generalized specification of an  $SDF$  against a more parsimonious null model. The application of these methods to several consumption-based models in the literature produces several interesting results, including (i) considerable efficiency can be gained by employing the optimal  $GMM$  estimator, and (ii) using conditional moment restrictions and optimal  $GMM$  leads to very different conclusions regarding the fit of several consumption-based models. While these models appear to do quite well in fitting the cross-section of average returns of size and book-to-market portfolios in tests based on unconditional moment restrictions, they fail to match variation in conditional moments of returns. Our methodology allows us to transparently show that the small *average* pricing errors that are obtained when estimation is based on unconditional moment restrictions hide enormous time-variation in *conditional* pricing errors.

# Notes

<sup>1</sup>Under value additivity and additional, relatively weak, regularity conditions, Hansen and Richard (1987) show that there is a unique pricing kernel  $m_{t+1}$  that prices all of the payoffs in a given payoff space according to  $E[m_{t+1}r_{i,t+1}|\mathcal{A}_t] = p$ , where  $\mathcal{A}_t$  is agents' information set. Conditioning down to the econometrician's information set  $\mathcal{J}_t$  gives this pricing relation.

<sup>2</sup>This follows from the observation that

$$E[r_{t+1}^i|\mathcal{J}_t] - \mu_t^{0\mathcal{J}} = \frac{-\text{Cov}[r_{t+1}^i, m_{t+1} | \mathcal{J}_t]}{E[m_{t+1} | \mathcal{J}_t]},$$

for a given  $r_t^i$  in the set of  $R$  test asset returns  $r_t$ . Substituting (3) and rearranging gives (4). This construction does not require the assumption that  $f_t \in \mathcal{J}_t$ . However, if  $f_t$  is not in  $\mathcal{J}_t$ , then the presumption would typically be that  $\mathcal{J}_t$  is a subset of an econometrician's information set. This is because having observations on  $f_t$  is generally required for the econometric implementation of (4)-(5).

<sup>3</sup>More generally, the links are between the return on a zero-beta portfolio and the conditional mean of  $m_{t+1}$ .

<sup>4</sup>Virtually all of the *GMM* estimators of factor models that have been implemented in the literature imply first-order conditions that are special cases of this moment condition. This includes Hansen (1982)'s fixed-instrument *GMM* estimator. Therefore, estimation based on the optimal choice of  $A_t$  determined subsequently will lead to estimators that are at least as efficient, and generally more efficient, than those employed in the extant literature.

<sup>5</sup>This form for  $\Sigma^A$  follows from the fact that  $A_t h_{t+1}(\theta_0)$  is a martingale difference sequence (see Hansen and Singleton (1982)).

<sup>6</sup>The rank condition in the definition of  $\mathcal{A}$  ensures that the model is econometrically identified. It is the counterpart to the rank condition in the classical simultaneous equations models.

<sup>7</sup>Hansen (1982)'s fixed-instrument *GMM* estimator has one minimize the quadratic form  $G_T(\theta)'W_T G_T(\theta)$ , where  $G_T(\theta) = T^{-1} \sum_t h_{t+1}(\theta) \otimes w_t$  and  $W_T$  is a  $LR \times LR$  dimensional distance matrix. The first-order conditions to this minimization problem set  $K$  linear combinations of the sample moments  $G_T(\theta_T)$  to zero. Straightforward rearrangement of these equations gives an expression of the form (10) with  $A_t$  depending on the choices of instruments  $w_t$  and distance matrix  $W$ .

<sup>8</sup>This step is exactly analogous to the projection of "right-hand-side" regressors onto the prede-

terminated variables in *2SLS* and *3SLS* estimation. In linear models, these regressors comprise the partial derivatives of the equation error with respect to  $\theta_0$ .

<sup>9</sup>In general,  $\partial h_{t+1}(\theta_0)/\partial\theta$  is nonlinear and its conditional expectation is unknown. The resulting intractability of the optimal *GMM* estimator no doubt underlies the absence of its application in financial economics. Hansen and Singleton (1996) derive and implement the optimal *GMM* estimator for a class of consumption-based pricing models with serially correlated, homoskedastic errors. The estimation problem here is fundamentally different in that we have serially uncorrelated, conditionally heteroskedastic errors.

<sup>10</sup>The potential for large biases is discussed theoretically in Newey and Smith (2004) and simulation evidence is provided by Altonji and Segal (1996), Hansen, Heaton, and Yaron (1996), and Imbens and Spady (2005), among others.

<sup>11</sup> Both the form of the pricing kernel  $m_{t+1}^G(\beta_0, \gamma_T^C)$  and the density underlying the expectation  $E[A_t h_{t+1}(\beta_0, \gamma_T^C)]$  will in general depend on  $\gamma_T^C$ .

<sup>12</sup>This form of the asymptotic distribution of  $\gamma_T^A$  under local alternatives, as well as the characterization of the non-centrality parameter in (26), follow from results in Newey and West (1987).

<sup>13</sup>More precisely, we are projecting the scaled versions of these constructs on each other, where scaling is by the square root of  $\Sigma_t^{-1}$ , as discussed above.

<sup>14</sup>We stress again that all of the derivations and results up to this point do not require that these factor weights be affine functions of  $z_t$ ; they can be any continuously differential function of  $z_t$ .

<sup>15</sup> That is, we solve (10), after substitution of the relevant special case of  $A^*$  in (20), for  $\gamma_T^G$ .

<sup>16</sup>The following equality is an immediate implication of the first-order conditions for the optimal *GMM* estimator  $\beta_T^N$  and the definition of  $\widehat{\mathcal{H}}_t^N$ .

<sup>17</sup>Jagannathan and Wang (1996) and Santos and Veronesi (2006) use these conditioning variables in  $\beta$ -style representations of excess returns, while we use them as conditioning variables in a consumption-based pricing kernel.

<sup>18</sup>Consistent with the extant literature that uses *GMM* estimators to evaluate the goodness-of-fit of asset pricing models under rational expectations, moments are estimated “in sample.” In this setting, the managed portfolio weights  $B_t$  are known to the representative agent/investor. They are not known to the econometrician assessing the model’s fit and so they are estimated using the full sample. In contrast, a “real time” investor implementing a dynamic trading strategy would be led to implement

a rolling optimal *GMM* estimator and its associated rolling portfolio weights  $B_t^*$ .

<sup>19</sup>The presence of autocorrelation does not necessarily mean that leave-one-out cross-validation will produce a suboptimal bandwidth. Autocorrelation implies dependence among neighboring observations in the time domain. Whether leave-one-out cross-validation results in under-smoothed or over-smoothed estimates depends on the dependence of observations that are neighbors in the state domain. High correlation of residuals of neighbors in time space does not necessarily translate into high correlation of residuals of neighbors in the state domain, unless  $z_t$  is very persistent and the sample short (Hart (1994); Yao and Tong (1998)).

<sup>20</sup>The conditional moment plots reveal some outliers for the lowest value of *cay* in Figure 1 and the highest value of *def* in Figure 2. Our subsequent estimation results are not sensitive to these outliers. Removal of these observations yields virtually unchanged results.

<sup>21</sup>The inclusion of this polynomial approximation to nonlinear dependence of the conditional means on  $z_t$  is motivated in part by the analysis in Ait-Sahalia (1996). This functional form is able to capture the linear, parabolic, and “S on its side” patterns evidenced in the non-parametric estimates of the conditional means displayed in Figures 1 and 2.

<sup>22</sup>We experimented with time-varying conditional covariance matrix from a dynamic conditional correlation (*DCC*) model (Engle (2002)), but allowing this flexibility had only negligible effects on our asset-pricing results. Accordingly, we proceed with the simpler specification outlined above.

## References

- Ait-Sahalia, Yacine, 1996, Testing Continuous-Time Models of the Spot Interest Rate, *Review of Financial Studies* 9, 385–426.
- Altonji, Joseph G., and Lewis M. Segal, 1996, Small Sample Bias in GMM Estimation of Covariance Structures, *Journal of Business and Economic Statistics* 14, 353–366.
- Campbell, John Y., and John H. Cochrane, 1999, By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior, *Journal of Political Economy* 107, 205–251.
- Carhart, Mark M., 1997, On Persistence of Mutual Fund Performance, *Journal of Finance* 52, 57–82.
- Chamberlain, Gary, 1987, Asymptotic Efficiency in Estimation with Conditional Moment Restrictions, *Journal of Econometrics* 34, 305–344.
- Cochrane, John H., 1996, A Cross-Sectional Test of an Investment-Based Asset Pricing Model, *Journal of Political Economy* 104, 572–621.
- Daniel, Kent D., and Sheridan Titman, 2006, Testing Factor Model Explanations of Market Anomalies, Working Paper, Northwestern University.
- Engle, Robert F., 2002, Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models, *Journal of Business and Economic Statistics* 17, 339–350.
- Fama, Eugene F., and Kenneth R. French, 1992, The Cross-Section of Expected Stock Returns, *Journal of Finance* 47, 427–465.

- Fama, Eugene F., and Kenneth R. French, 1993, Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics* 33, 23–49.
- Fama, Eugene F., and Kenneth R. French, 1996, Multifactor Explanations of Asset Pricing Anomalies, *Journal of Finance* 51, 55–87.
- Fan, Jianqing, 1992, Design-adaptive Nonparametric Regression, *Journal of the American Statistical Association* 87, 998–1004.
- Grossman, Sanford J., and Robert J. Shiller, 1981, The Determinants of the Variability of Stock Market Prices, *American Economic Review* 71, 222–227.
- Hansen, Lars P., 1982, Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* 50, 1029–1054.
- Hansen, Lars P., 1985, A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators, *Journal of Econometrics* 30, 203–238.
- Hansen, Lars P., John C. Heaton, and Masao Ogaki, 1988, Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions, *Journal of the American Statistical Association* 83, 863–871.
- Hansen, Lars P., John C. Heaton, and Amir Yaron, 1996, Finite-Sample Properties of Some Alternative GMM Estimators, *Journal of Business and Economic Statistics* 14, 262–280.
- Hansen, Lars P., and Scott F. Richard, 1987, The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models, *Econometrica* 55, 587–613.
- Hansen, Lars P., and Kenneth J. Singleton, 1982, Generalized Instrumental Variables



- Estimation of Nonlinear Rational Expectations Models, *Econometrica* 50, 1269–1286.
- Hansen, Lars P., and Kenneth J. Singleton, 1996, Efficient Estimation of Linear Asset Pricing Models with Moving Average Errors, *Journal of Business & Economic Statistics* 14, 53–68.
- Hart, Jeffrey D., 1994, Automated Kernel Smoothing of Dependent Data by Using Time Series Cross-Validation, *Journal of the Royal Statistical Society, Series B* 56, 529–542.
- Hodrick, Robert J., and Xiaoyan Zhang, 2001, Evaluating the Specification Errors of Asset Pricing Models, *Journal of Financial Economics* 62, 327–376.
- Imbens, Guido W., and Richard H. Spady, 2005, The Performance of Empirical Likelihood and its Generalizations, in Donald W.K. Andrews and J.H. Stock, eds.: *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (Cambridge University Press, New York).
- Jagannathan, Ravi, and Zhenyu Wang, 1996, The Conditional CAPM and the Cross-section of Expected Returns, *Journal of Finance* 51, 3–54.
- Lettau, Martin, and Sydney C. Ludvigson, 2001a, Consumption, Aggregate Wealth, and Expected Stock Returns, *Journal of Finance* 56, 815–849.
- Lettau, Martin, and Sydney C. Ludvigson, 2001b, Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying, *Journal of Political Economy* 109, 1238–1287.
- Lewellen, Jonathan, and Stefan Nagel, 2006, The Conditional CAPM Does Not Explain Asset Pricing Anomalies, *Journal of Financial Economics* 79, 289–314.

- Lewellen, Jonathan, Stefan Nagel, and Jay Shanken, 2010, A Skeptical Appraisal of Asset-Pricing Tests, *Journal of Financial Economics* 96, 175–194.
- Lustig, Hanno, and Stijn Van Nieuwerburgh, 2006, Housing Collateral, Consumption Insurance, and Risk Premia: An Empirical Perspective, *Journal of Finance* 60, 1167–1219.
- Newey, Whitney K., and Richard J. Smith, 2004, Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators, *Econometrica* 72, 219–255.
- Newey, Whitney K., and Kenneth D. West, 1987, Hypothesis Testing with Efficient Method of Moment Estimation, *International Economic Review* 28, 777–787.
- Pastor, Lubos, and Robert F. Stambaugh, 2003, Liquidity Risk and Expected Stock Returns, *Journal of Political Economy* 111, 642–685.
- Roussanov, Nikolai, 2009, Composition of Wealth, Conditioning Information, and the Cross-Section of Stock Returns, Working Paper, University of Pennsylvania.
- Santos, Tano, and Pietro Veronesi, 2006, Labor Income and Predictable Stock Returns, *Review of Financial Studies* 19, 1–44.
- Yao, Qiwei, and Howell Tong, 1998, Cross-validatory bandwidth selections for regression estimation based on dependent data, *Journal of Statistical Planning and Inference* 68, 387–415.

Table I: Calculation of Test Statistics

The matrices  $\widehat{\mathcal{H}}_t^G$  and  $\widehat{\mathcal{H}}_t^N$  are as defined in Section B, but with unconditional instead of conditional moments in the cases of the unconditional and fixed IV estimators.  $DF$  denotes degrees of freedom,  $R$  the number of basis assets,  $K$  the number of  $SDF$  parameters,  $L$  the number of fixed instruments, and  $G$  the number of additional  $SDF$  parameters describing the alternative relative to the null  $SDF$  specification.

Test statistic		Unconditional	Fixed IV	Optimal IV
$\tau_T(I)$	$h_{t+1}$ $B_t$ $DF$	$m_{t+1} (\theta_T^G) r_{t+1} - p$ $I_R$ $R - K$	$(m_{t+1} (\theta_T^G) r_{t+1} - p) \otimes w_t$ $I_{LR}$ $LR - K$	$m_{t+1} (\theta_T^G) r_{t+1} - p$ $I_R$ $R$
$\tau_T(B^{Wald})$	$h_{t+1}$ $B_t$ $DF$	$m_{t+1} (\theta_T^N) r_{t+1} - p$ $\widehat{\mathcal{H}}_t^G \widehat{\Sigma}_t^{G-1}$ $G$	$(m_{t+1} (\theta_T^N) r_{t+1} - p) \otimes w_t$ $\widehat{\mathcal{H}}_t^G \widehat{\Sigma}_t^{G-1}$ $G$	$m_{t+1} (\theta_T^N) r_{t+1} - p$ $\widehat{\mathcal{H}}_t^G \widehat{\Sigma}_t^{G-1}$ $G$
$\tau_T(B^{LM})$	$h_{t+1}$ $B_t$ $DF$	$m_{t+1} (\theta_T^N) r_{t+1} - p$ $\widehat{\mathcal{H}}_t^N \widehat{\Sigma}_t^{N-1}$ $G$	$(m_{t+1} (\theta_T^N) r_{t+1} - p) \otimes w_t$ $\widehat{\mathcal{H}}_t^N \widehat{\Sigma}_t^{N-1}$ $G$	$m_{t+1} (\theta_T^N) r_{t+1} - p$ $\widehat{\mathcal{H}}_t^N \widehat{\Sigma}_t^{N-1}$ $G$

Table II: Consumption CAPM, moments conditioned on *cay*

Test asset returns are the excess returns on the four size and B/M portfolios and the gross return on the T-Bill. Standard errors (in parentheses) and  $p$ -values (in brackets) are robust to misspecification of conditional moments, except those shown in italics, which assume correctly specified conditional moments. Conditional moments for opt. IV-local are estimated with local regressions; for opt. IV-sieve they are based on the sieve method.

	const.	$\Delta c_{t+1}$	$\tau(I)$
Uncond.	2.95 (0.74)	-365.35 (135.26)	9.30 [0.03]
Fixed IV	1.00 (0.00)	-0.11 (0.15)	215.12 [0.00]
Opt. IV – Local	0.99 (0.00) <i>(0.00)</i>	0.47 (0.24) <i>(0.34)</i>	67.17 [0.00] <i>[0.00]</i>
Opt. IV – Sieve	1.00 (0.00) <i>(0.00)</i>	0.12 (0.19) <i>(0.12)</i>	113.41 [0.00] <i>[0.00]</i>

Table III: Pricing kernel estimates with moments conditioned on  $cay$

Test assets returns are the excess returns on the four size and B/M portfolios and the gross return on the T-Bill. Standard errors (in parentheses) and  $p$ -values (in brackets) are robust to misspecification of conditional moments, except those shown in italics, which assume correctly specified conditional moments. Conditional moments for opt. IV-local are estimated with local regressions; for opt. IV-sieve they are based on the sieve method.

	const.	$cay_t$	$\Delta c_{t+1}$	$cay_t \times \Delta c_{t+1}$	$\tau(I)$	$\tau(B^{Wald})$	$\tau(B^{LM})$
Uncond.	-3.24 (8.84)	-40.83 (206.91)	626.99 (1437.79)	-70564.09 (99269.77)	0.09 [0.77]	0.59 [0.74]	7.90 [0.02]
Fixed IV	1.00 (0.00)	-0.64 (0.16)	-0.47 (0.30)	105.42 (35.02)	143.91 [0.00]	21.37 [0.00]	51.05 [0.00]
Opt. IV – Local	1.27 (0.29) <i>(0.23)</i>	-9.12 (9.15) <i>(6.96)</i>	-50.00 (49.84) <i>(41.16)</i>	1054.53 (1161.22) <i>(861.37)</i>	47.27 [0.00] <i>[0.00]</i>	1.31 [0.52] <i>[0.42]</i>	1.56 [0.46] <i>[0.42]</i>
Opt. IV – Sieve	1.00 (0.00) <i>(0.00)</i>	-0.06 (0.06) <i>(0.04)</i>	-0.09 (0.27) <i>(0.14)</i>	-2.81 (9.13) <i>(7.21)</i>	89.29 [0.00] <i>[0.00]</i>	5.19 [0.07] <i>[0.00]</i>	4.65 [0.10] <i>[0.00]</i>

Table IV: Pricing kernel estimates with moments conditioned on  $def$

Test assets returns are the excess returns on the four size and B/M portfolios and the gross return on the T-Bill. Standard errors (in parentheses) and  $p$ -values (in brackets) are robust to misspecification of conditional moments, except those shown in italics, which assume correctly specified conditional moments. Conditional moments for opt. IV-local are estimated with local regressions; for opt. IV-sieve they are based on the sieve method.

	const.	$def_t$	$\Delta c_{t+1}$	$def_t \times \Delta c_{t+1}$	$\tau(I)$	$\tau(B^{Wald})$	$\tau(B^{LM})$
Uncond.	4.50 (3.06)	-274.15 (343.00)	-71.89 (381.84)	-11214.69 (39098.00)	6.49 [0.01]	2.62 [0.27]	1.70 [0.43]
Fixed IV	1.05 (0.04)	-5.33 (4.05)	-9.80 (7.25)	945.10 (671.89)	124.17 [0.00]	2.51 [0.29]	38.79 [0.00]
Opt. IV – Local	2.12 (0.49) <i>(0.66)</i>	-30.59 (31.44) <i>(40.27)</i>	-188.65 (78.15) <i>(111.06)</i>	3215.15 (4126.64) <i>(6579.73)</i>	34.98 [0.00] <i>[0.00]</i>	1.14 [0.56] <i>[0.42]</i>	6.06 [0.05] <i>[0.13]</i>
Opt. IV – Sieve	1.01 (0.00) <i>(0.00)</i>	-1.00 (0.38) <i>(0.22)</i>	-1.30 (0.58) <i>(0.40)</i>	117.04 (59.16) <i>(38.10)</i>	52.16 [0.00] <i>[0.00]</i>	10.33 [0.01] <i>[0.00]</i>	9.68 [0.01] <i>[0.00]</i>

Table V: Pricing kernel estimates with moments conditioned on  $yc$

Test asset returns are the excess returns on the four size and B/M portfolios and the gross return on the T-Bill. Standard errors (in parentheses) and  $p$ -values (in brackets) are robust to misspecification of conditional moments, except those shown in italics, which assume correctly specified conditional moments. Conditional moments for opt. IV-local are estimated with local regressions; for opt. IV-sieve they are based on the sieve method.

	const.	$yc_t$	$\Delta c_{t+1}$	$yc_t \times \Delta c_{t+1}$	$\tau(I)$	$\tau(B^{Wald})$	$\tau(B^{LM})$
Uncond.	-5.70 (32.49)	9.33 (35.51)	-140.41 (4454.77)	-214.90 (4922.26)	9.63 [0.00]	0.13 [0.93]	0.14 [0.93]
Fixed IV	0.79 (0.09)	0.24 (0.09)	34.16 (15.23)	-38.31 (16.62)	128.69 [0.00]	7.43 [0.02]	44.72 [0.00]
Opt. IV – Local	0.72 (0.11) <i>(0.15)</i>	0.31 (0.12) <i>(0.17)</i>	53.95 (19.08) <i>(27.06)</i>	-59.64 (21.19) <i>(29.95)</i>	56.31 [0.00] <i>[0.00]</i>	8.46 [0.01] <i>[0.11]</i>	2.26 [0.32] <i>[0.27]</i>
Opt. IV – Sieve	0.99 (0.05) <i>(0.02)</i>	0.01 (0.06) <i>(0.02)</i>	-1.36 (8.59) <i>(3.78)</i>	1.52 (9.45) <i>(4.13)</i>	94.29 [0.00] <i>[0.00]</i>	2.00 [0.37] <i>[0.12]</i>	2.03 [0.36] <i>[0.12]</i>

Table VI: Pricing errors in cross section and time series

The table reports the time-series standard deviation (S.D.) of conditional pricing errors and the cross-sectional root mean squared error (RMSE) of the test assets' unconditional pricing errors. Test asset returns are the excess returns on the four size and B/M portfolios and the gross return on the T-Bill. Conditional moments for opt. IV-local are estimated with local regressions; for opt. IV-sieve they are based on the sieve method.

	Time-series S.D. of conditional pricing errors					Cross-sectional RMSE of uncond. pricing errors
	SmGrw	SmVal	BigGrw	BigVal	T-Bill	
Panel A: <i>SDF</i> with $\Delta c_{t+1}$ scaled by $cay_t$ , moments conditioned on $cay_t$						
Uncond.	0.17	0.21	0.15	0.17	5.41	0.02
Fixed IV	0.02	0.02	0.02	0.02	0.00	0.05
Opt. IV – Local	0.03	0.02	0.03	0.02	0.01	0.04
Opt. IV – Sieve	0.03	0.03	0.03	0.03	0.00	0.05
Panel B: <i>SDF</i> with $\Delta c_{t+1}$ scaled by $def_t$ , moments conditioned on $def_t$						
Uncond.	0.09	0.09	0.05	0.06	1.36	0.02
Fixed IV	0.02	0.02	0.01	0.01	0.00	0.05
Opt. IV – Local	0.01	0.01	0.00	0.01	0.04	0.03
Opt. IV – Sieve	0.03	0.03	0.02	0.02	0.00	0.05
Panel C: <i>SDF</i> with $\Delta c_{t+1}$ scaled by $yc_t$ , moments conditioned on $yc_t$						
Uncond.	0.02	0.00	0.02	0.02	0.38	0.02
Fixed IV	0.00	0.00	0.01	0.02	0.00	0.05
Opt. IV – Local	0.00	0.00	0.01	0.02	0.00	0.05
Opt. IV – Sieve	0.03	0.02	0.02	0.03	0.00	0.05



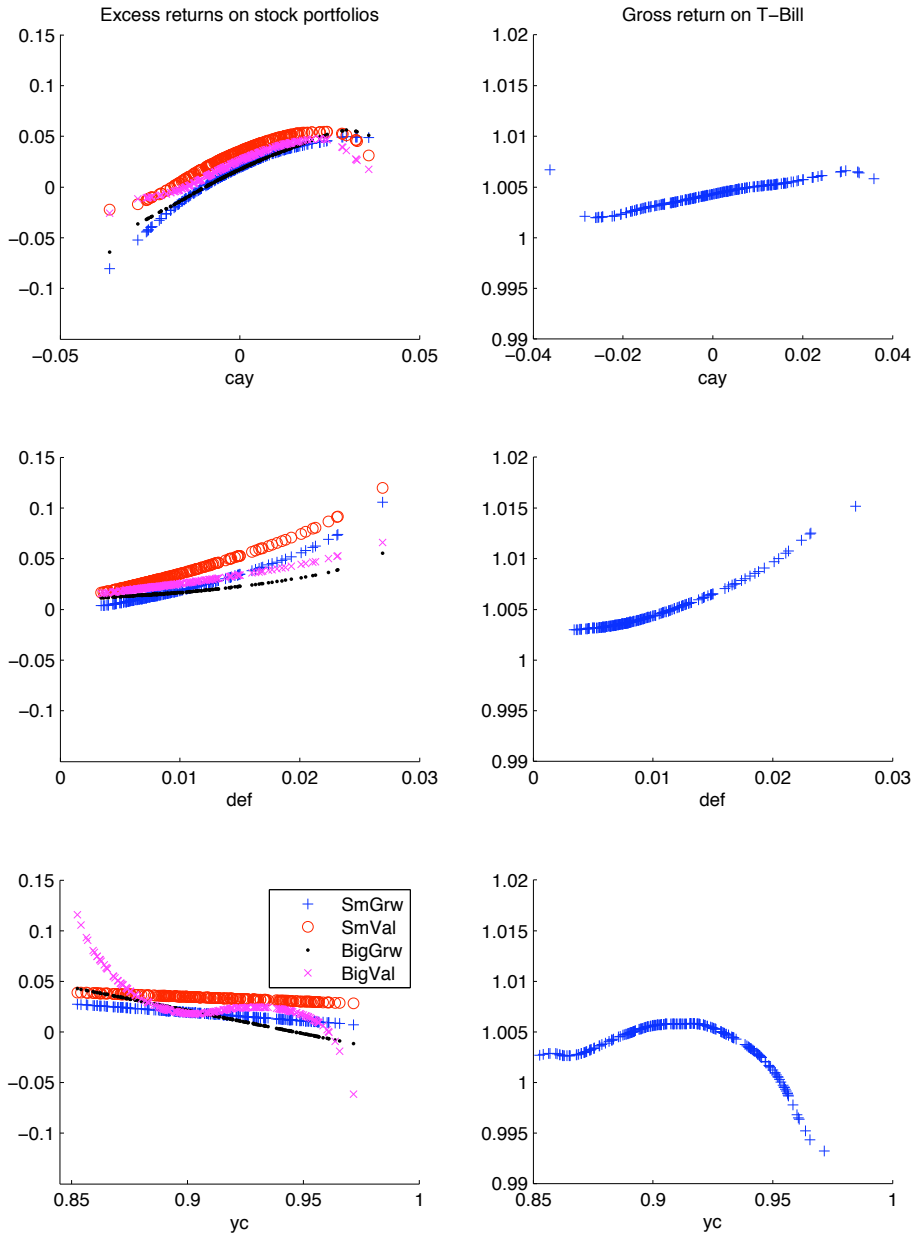


Figure 1: Fitted conditional expected returns from the local regression method

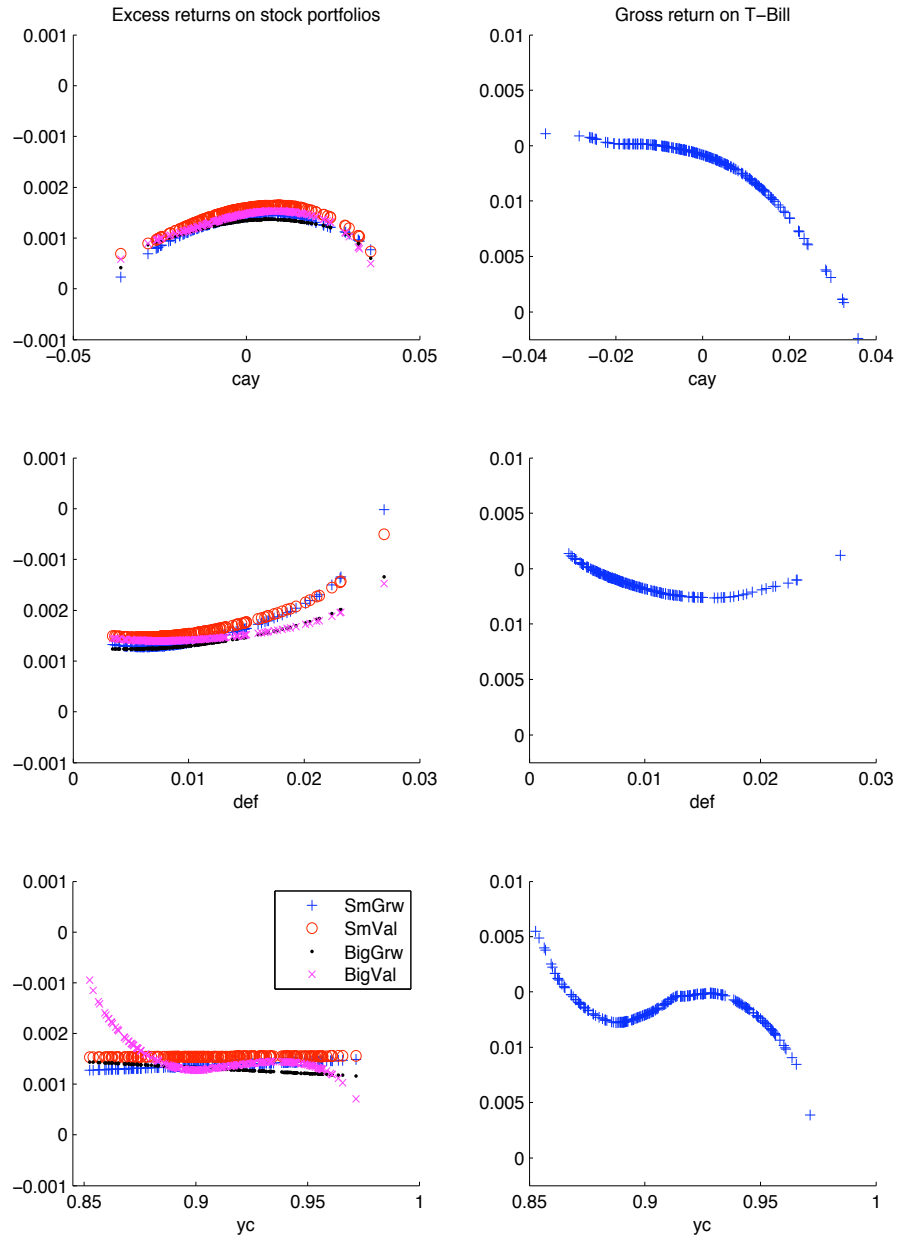


Figure 2: Fitted conditional expected cross-products of return and log consumption growth from the local regression method

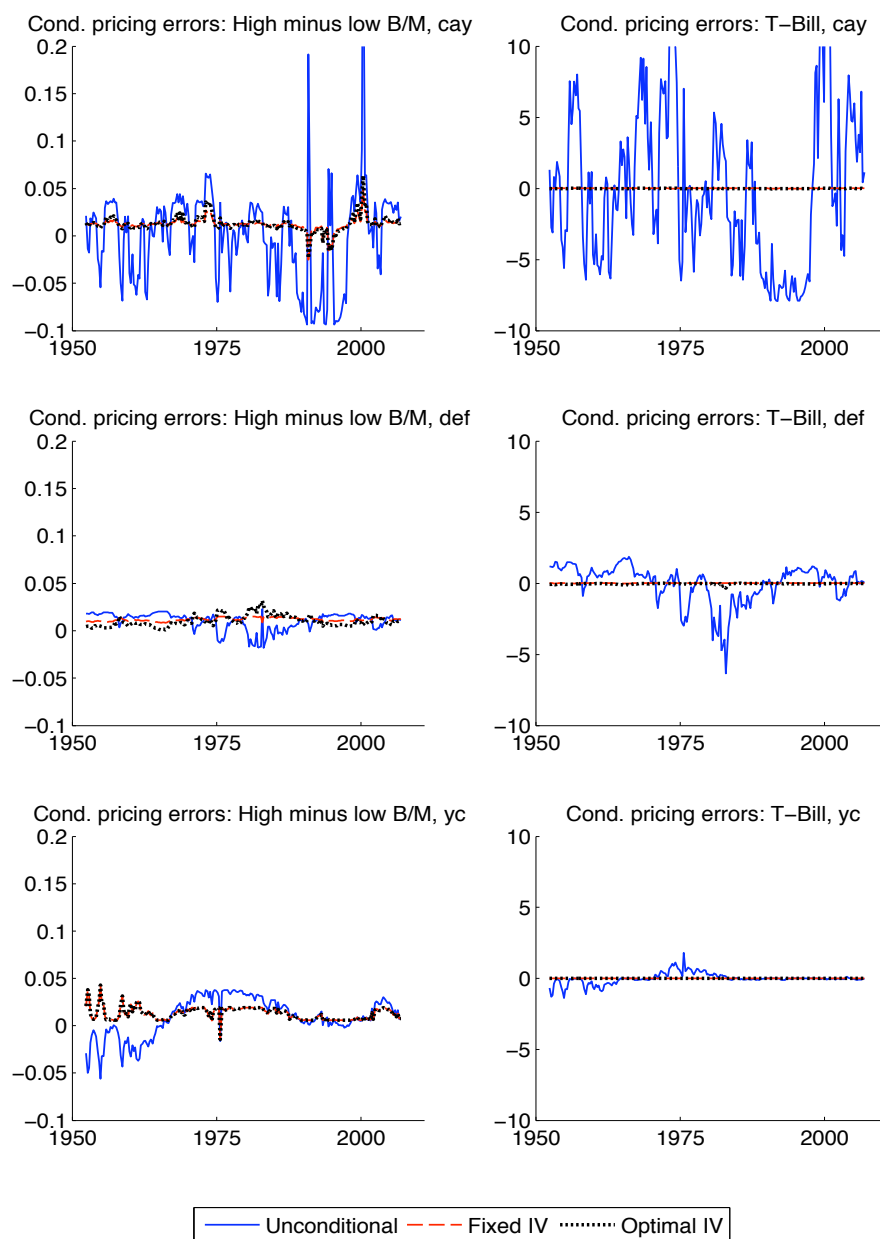


Figure 3: Conditional pricing errors implied by unconditional, fixed IV, and optimal IV-local estimates of pricing kernels with time-varying weights: High minus low book-to-market zero investment portfolio (left) and T-Bill (right) with local regression estimates of moments conditioned on *cay* (top row), *def* (middle row), and *yc* (bottom row)

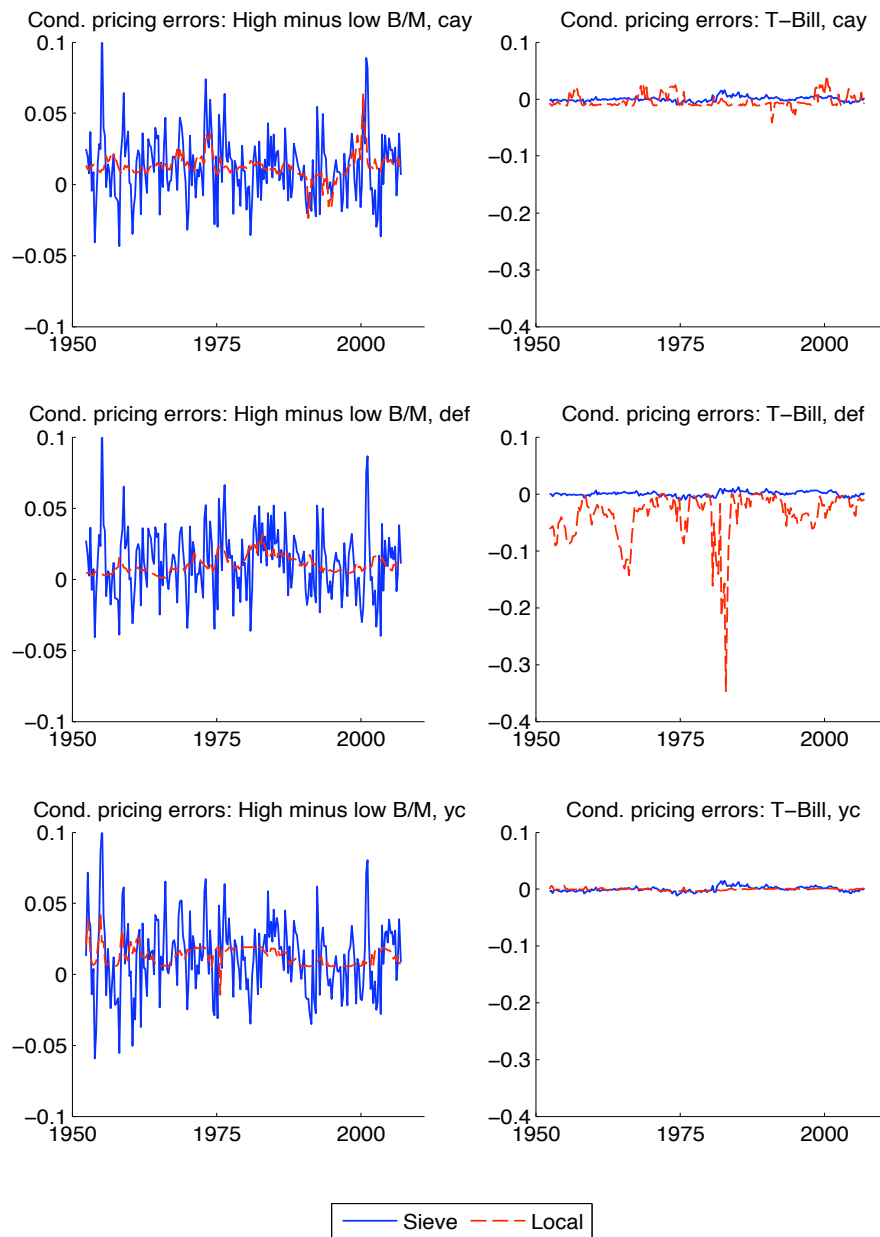


Figure 4: Conditional pricing errors implied by optimal IV-local and optimal IV-sieve estimates of pricing kernels with time-varying weights: High minus low book-to-market zero investment portfolio (left) and T-Bill (right) and moments conditioned on *cay* (top row), *def* (middle row), and *yc* (bottom row)

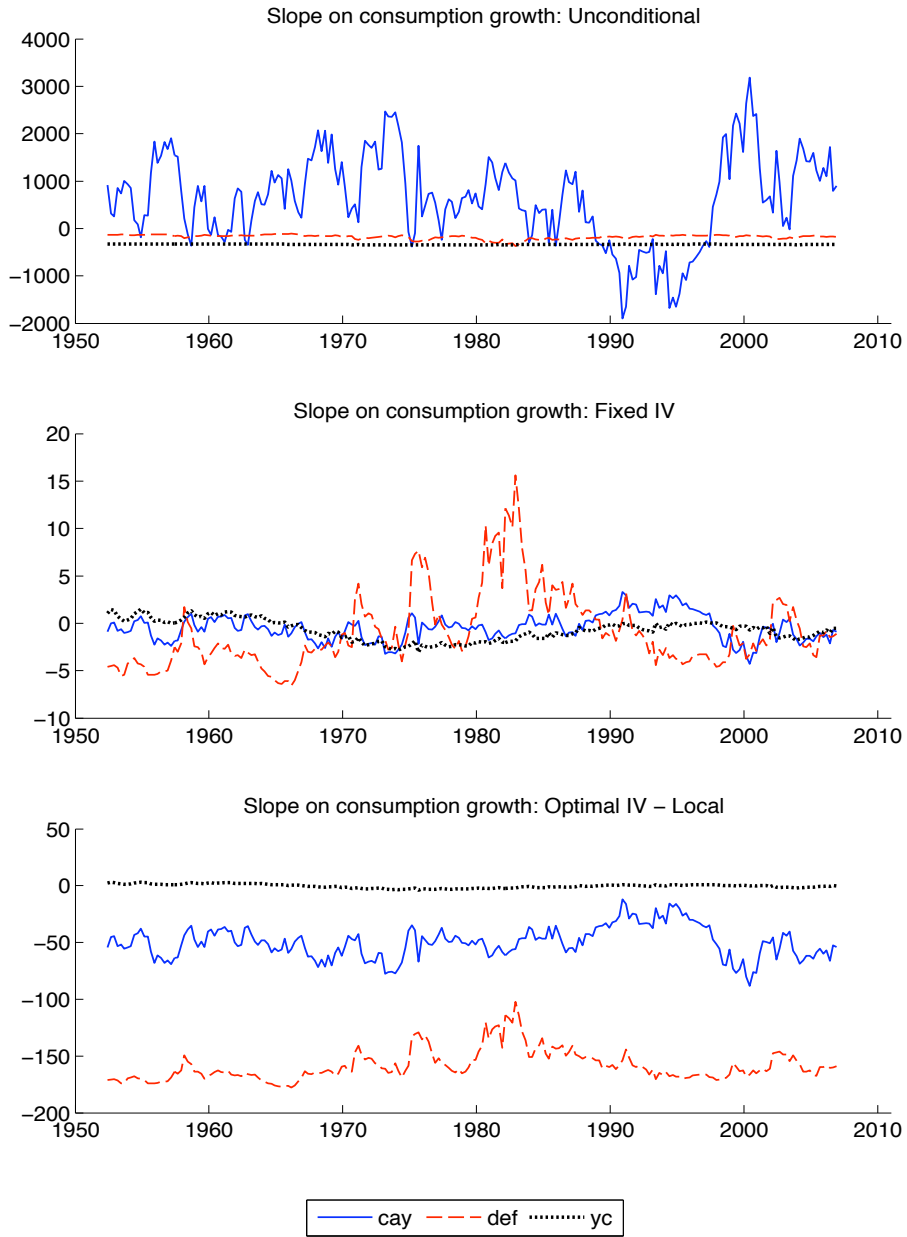


Figure 5: Time-series of estimated SDF weights from with unconditional (top row), fixed IV (middle row), and optimal IV-local estimators (bottom row)

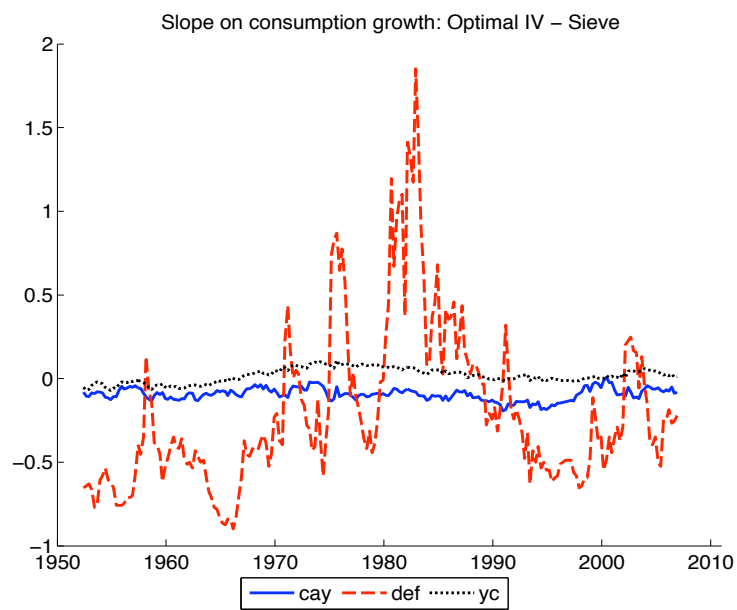


Figure 6: Time-series of optimal IV estimates of SDF weight with conditional moments obtained with the sieve method

Internet Appendix  
for  
Estimation and Evaluation of  
Conditional Asset Pricing Models\*

Stefan Nagel<sup>†</sup>  
*Stanford University and NBER*

Kenneth J. Singleton<sup>‡</sup>  
*Stanford University and NBER*

September 28, 2010

---

\*Citation Format: Nagel, Stefan, and Kenneth J. Singleton, YEAR, Internet Appendix to “Estimation and Evaluation of Conditional Asset Pricing Models,” *Journal of Finance* VOL, pages, <http://www.afajof.org/IA/YEAR.asp>. Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the authors of the article.

<sup>†</sup>Stanford University, Graduate School of Business, 518 Memorial Way, Stanford, CA 94305, e-mail: [Nagel.Stefan@gsb.stanford.edu](mailto:Nagel.Stefan@gsb.stanford.edu), <http://faculty-gsb.stanford.edu/nagel>

<sup>‡</sup>Stanford University, Graduate School of Business, 518 Memorial Way, Stanford, CA 94305, e-mail: [kenneths@stanford.edu](mailto:kenneths@stanford.edu), <http://www.stanford.edu/~kenneths/>

## A The Asymptotic Distribution of $\tau_T(B, A)$

A standard, coordinate by coordinate, mean-value expansion of the sample moment conditions (10) gives

$$\sqrt{T}(\theta_T^A - \theta_0) = - \left[ \frac{1}{T} \sum_t A_t \frac{\partial h_{t+1}(\theta_T^{Am})}{\partial \theta} \right]^{-1} \frac{1}{\sqrt{T}} \sum_t A_t h_{t+1}(\theta_0), \quad (\text{A.1})$$

where  $\theta_T^{Am}$  is a collection of vectors, one for each coordinate of  $A_t h_{t+1}$ , that lie between  $\theta_T^A$  and  $\theta_0$ , almost surely. Similarly, a mean-value expansion of the sample mean of  $B_t h_{t+1}(\theta_T^A)$  gives

$$\frac{1}{\sqrt{T}} \sum_t B_t h_{t+1}(\theta_T^A) = \frac{1}{\sqrt{T}} \sum_t B_t h_{t+1}(\theta_0) + \frac{1}{T} \sum_t B_t \frac{\partial h_{t+1}(\theta_T^{Bm})}{\partial \theta} \times \sqrt{T}(\theta_T^A - \theta_0), \quad (\text{A.2})$$

with  $\theta_T^{Bm}$  interpreted similarly. Substitution of (A.1) into (A.2) leads to

$$\frac{1}{\sqrt{T}} \sum_t B_t h_{t+1}(\theta_T^A) = \frac{1}{\sqrt{T}} \sum_t C_t^A h_{t+1}(\theta_0) + o_p(1), \quad (\text{A.3})$$

where  $C_t^A$  is given by (15). The limiting distribution in (14) follows immediately under the regularity conditions in Hansen (1982) using the fact that  $h_{t+1}(\theta_0)$  follows a martingale difference sequence with conditional covariance matrix  $E[h_{t+1}(\theta_0)h_{t+1}(\theta_0)'] = \Sigma_t$ .

## B Intermediate Steps in Section III

To express the Wald statistic  $\zeta_T^W(A^*)$  as in (27) we proceed as follows. From the intermediate steps in deriving the asymptotic distribution of  $\theta_T^A$  we can express  $(\theta_T^* - \theta_0)$



as

$$\sqrt{T}(\theta_T^* - \theta_0) \stackrel{a}{=} - (E [\Psi_t^{\theta'} \Sigma_t^{\mathcal{G}-1} \Psi_t^\theta])^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^T \Psi_t^{\theta'} \Sigma_t^{\mathcal{G}-1} h_{t+1}(\theta_0). \quad (\text{A.4})$$

Noting that  $\sqrt{T}(\gamma_T^* - \gamma_0) = [0, I_G] \sqrt{T}(\theta_T^* - \theta_0)$ , and using the partitioned matrix formula for inverting  $\Omega_0^*$ , we obtain

$$\sqrt{T}(\gamma_T^* - \gamma_0) \stackrel{a}{=} - \Omega_{\gamma\gamma}^* \frac{1}{\sqrt{T}} \sum_{t=1}^T \mathcal{H}_t^{\mathcal{G}'} \Sigma_t^{\mathcal{G}-1} h_{t+1}(\theta_0). \quad (\text{A.5})$$

The random vector  $\frac{1}{\sqrt{T}} \sum_{t=1}^T \mathcal{H}_t^{\mathcal{G}'} \Sigma_t^{\mathcal{G}-1} h_{t+1}(\theta_0)$  converges in distribution to a normal random vector with mean zero and covariance matrix

$$(\Omega_{\gamma\gamma}^*)^{-1} = \mathcal{K}^{\gamma\gamma} - \mathcal{K}^{\gamma\beta} (\mathcal{K}^{\beta\beta})^{-1} \mathcal{K}^{\beta\gamma}, \quad (\text{A.6})$$

where the last equality follows from the partitioned matrix inversion formula applied to  $\Omega_0^*$ . Therefore, the asymptotic distribution of  $\zeta_T^W(A^*)$  in (27) is  $\chi^2(G)$ .

## C Derivation the Lagrange Multiplier

The relevant Lagrange multipliers come from solving the *GMM* estimation problem subject to the constraint that  $\gamma_0 = 0$ . More precisely, the moment conditions associated with the optimal *GMM* estimator of  $\theta_0$  for the unconstrained  $m_{t+1}^{\mathcal{G}}$  are

$$E \left[ \begin{pmatrix} \Psi_t^{\beta'} \\ \Psi_t^{\gamma'} \end{pmatrix} \Sigma_t^{-1} h_{t+1}(\beta_0, \gamma_0) \right] = 0. \quad (\text{A.7})$$

Under the constraint that  $\gamma_0 = 0$ , (A.7) gives more moment equations ( $K$ ) than unknown parameters ( $K - G = \dim \beta_0$ ). Therefore, the *LM* statistic for testing  $H_0 : \gamma_0 = 0$

is obtained by minimizing a quadratic form in the sample version of the moments (A.7) for joint estimation of  $\beta_0$  and  $\gamma_0$ , subject to the constraint that  $\gamma_T = 0$  (see Eichenbaum, Hansen, and Singleton (1988)). Letting  $h_{t+1}^{\mathcal{N}}(\beta) = h_{t+1}(\beta, 0)$ , the pricing errors under the constraint that  $\gamma = 0$ , the optimal distance matrix in this quadratic form is a consistent estimator of

$$W_0 = E \left( \begin{pmatrix} \Psi_t^{\beta'} \Sigma_t^{\mathcal{N}-1} h_{t+1}^{\mathcal{N}} \\ \Psi_t^{\gamma'} \Sigma_t^{\mathcal{N}-1} h_{t+1}^{\mathcal{N}} \end{pmatrix} \begin{pmatrix} h_{t+1}^{\mathcal{N}'} \Sigma_t^{\mathcal{N}-1} \Psi_t^{\beta}, h_{t+1}^{\mathcal{N}'} \Sigma_t^{\mathcal{N}-1} \Psi_t^{\gamma} \end{pmatrix} \right).$$

The first-order conditions to this minimization problem are

$$\left( \frac{1}{T} \sum_t \mathcal{P}_{t+1} \right) W_T^{-1} \frac{1}{T} \sum_t \begin{pmatrix} \Psi_t^{\beta'} \\ \Psi_t^{\gamma'} \end{pmatrix} \Sigma_t^{\mathcal{N}-1} h_{t+1}^{\mathcal{N}}(\beta_T) = \begin{pmatrix} 0 \\ \lambda_T \end{pmatrix}, \quad (\text{A.8})$$

where  $\lambda_T$  is the  $G \times 1$  vector of Lagrange multipliers associated with the constraint that  $\gamma_T = 0$ ; it is understood that  $\Sigma_t^{\mathcal{N}}$ ,  $\Psi_t^{\gamma}$ , and  $\Psi_t^{\theta}$  have been replaced by consistent estimators of these constructs; and the matrix  $\mathcal{P}$  is given by

$$\mathcal{P}_{t+1} = \begin{bmatrix} \frac{\partial h_{t+1}^{\mathcal{N}}(\beta_T)'}{\partial \beta} \Sigma_t^{\mathcal{N}-1} \Psi_t^{\beta} & \frac{\partial h_{t+1}^{\mathcal{N}}(\beta_T)'}{\partial \beta} \Sigma_t^{\mathcal{N}-1} \Psi_t^{\gamma} \\ \frac{\partial h_{t+1}^{\mathcal{N}}(\beta_T)'}{\partial \gamma} \Sigma_t^{\mathcal{N}-1} \Psi_t^{\beta} & \frac{\partial h_{t+1}^{\mathcal{N}}(\beta_T)'}{\partial \gamma} \Sigma_t^{\mathcal{N}-1} \Psi_t^{\gamma} \end{bmatrix}. \quad (\text{A.9})$$

The lead matrix  $T^{-1} \sum_t \mathcal{P}_{t+1}$  in (A.8) is a consistent estimator of  $W_0$ . Therefore, the first  $K - G$  first-order conditions in (A.8) are

$$\frac{1}{T} \sum_t \Psi_t^{\beta'} \Sigma_t^{\mathcal{N}-1} h_{t+1}^{\mathcal{N}}(\beta_T^{\mathcal{N}}) = 0. \quad (\text{A.10})$$

These are the sample first-order conditions for the optimal *GMM* estimator of the parameters of the *SDF* under the null hypothesis  $\gamma_0 = 0$ ; that is, they are the first-

order conditions when estimation proceeds with the constrained  $SDF$   $m_{t+1}^N$ .<sup>1</sup> We let  $\beta_T^N$  denote this optimal  $GMM$  estimator obtained when the  $SDF$  is taken to be  $m_{t+1}^N(\beta_0)$ .

The last  $G$  first-order conditions in (A.8) yield the Lagrange multipliers

$$\lambda_T = \frac{1}{T} \sum_t \Psi_t^{\gamma'} \Sigma_t^{-1} h_{t+1}^N(\beta_T^N), \quad (\text{A.11})$$

as in (34).

## D An Alternative Representation of the Wald Statistic for Completely Affine $SDF$ s

We want to prove that  $\frac{1}{T} \sum_{t=1}^T \hat{\mathcal{H}}_t^{\mathcal{G}} \hat{\Sigma}_t^{\mathcal{G}-1} p = \frac{1}{T} \sum_{t=1}^T \hat{\mathcal{H}}_t^{\mathcal{G}} \hat{\Sigma}_t^{\mathcal{G}-1} h_{t+1}^N(\beta_T^N)$  for completely affine  $SDF$ s.

We have  $p_R - h_{t+1}^N(\beta_T^N) = r_{t+1} f_{t+1}^{\#N'} \beta_T^N$  and so

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T \left[ \hat{\mathcal{H}}_t^{\mathcal{G}} \hat{\Sigma}_t^{\mathcal{G}-1} \{p - h_{t+1}^N(\beta_T^N)\} \right] \\ &= \frac{1}{T} \sum_{t=1}^T \left[ \left( \hat{\Psi}_t^{\gamma'} - \hat{\mathcal{K}}_T^{\gamma\beta} \left( \hat{\mathcal{K}}_T^{\beta\beta} \right)^{-1} \hat{\Psi}_t^{\beta'} \right) \hat{\Sigma}_t^{\mathcal{G}-1} r_{t+1} f_{t+1}^{\#N'} \beta_T^N \right] \\ &= \frac{1}{T} \sum_{t=1}^T \left[ \hat{\Psi}_t^{\gamma'} \hat{\Sigma}_t^{\mathcal{G}-1} r_{t+1} f_{t+1}^{\#N'} \beta_T^N - \hat{\mathcal{K}}_T^{\gamma\beta} \left( \hat{\mathcal{K}}_T^{\beta\beta} \right)^{-1} \hat{\Psi}_t^{\beta'} \hat{\Sigma}_t^{\mathcal{G}-1} r_{t+1} f_{t+1}^{\#N'} \beta_T^N \right] \\ &= \hat{\mathcal{K}}_T^{\gamma\beta} \beta_T^N - \hat{\mathcal{K}}_T^{\gamma\beta} \left( \hat{\mathcal{K}}_T^{\beta\beta} \right)^{-1} \left( \hat{\mathcal{K}}_T^{\beta\beta} \right) \beta_T^N = 0, \end{aligned}$$

---

<sup>1</sup>This derivation addresses an important question that was left implicit up to this point. In previous sections we first constructed the optimal  $GMM$  estimator  $\theta_T^*$  of the parameters governing  $m_{t+1}(\theta_0)$ , and then proceeded to construct tests based on managed portfolio weights  $B_t$  and the moment conditions  $E[B_t h_{t+1}(\theta_0)] = 0$ . Readers may wonder whether we would have obtained even more efficient estimators than  $\theta_T^*$  by using the moment conditions  $E[A_t^* h_{t+1}(\theta_0)] = 0$  and  $E[B_t h_{t+1}(\theta_0)] = 0$  simultaneously to estimate  $\theta_0$ . By analogous derivations to those above we see that the answer is no. For otherwise  $A^*$  would not have been the optimal set of instruments to begin with.

where we are relying on the robust formulation of  $\hat{\mathcal{K}}_T^{\gamma\beta}$  as discussed in Section III.B.

## E Robust Statistics

The robust version of the asymptotic variance of the *SDF* parameter estimates follows Eq. (11), while the non-robust version replaces  $\partial h_{t+1}(\theta_0)/\partial\theta$  and the realized cross-products of pricing errors in Eq. (11) with their conditional expectations,  $\Psi_t^\theta$  and  $\Sigma_t$ , respectively, which yields the asymptotic variance as in Eq. (21).

Similarly, we compute the *LM* test statistic  $\tau_T(B^{LM})$  in its robust version following the *LM* analog of Eq. (38) with  $\hat{\mathcal{H}}_t^{\mathcal{N}}\hat{\Sigma}_t^{\mathcal{N}-1}h_{t+1}^{\mathcal{N}}(\beta_T)h_{t+1}^{\mathcal{N}}(\beta_T)'\hat{\Sigma}_t^{\mathcal{N}-1}\hat{\mathcal{H}}_t^{\mathcal{N}'}$  in the summation terms in the inverse. In the non-robust version of the *LM* statistic, these terms are reduced to  $\hat{\mathcal{H}}_t^{\mathcal{N}}\hat{\Sigma}_t^{\mathcal{N}-1}\hat{\mathcal{H}}_t^{\mathcal{N}'}$ .

The robust version of the *Wald* statistic is analogous to the *LM* statistic, just with  $\hat{\mathcal{H}}_t^{\mathcal{G}}$  in place of  $\hat{\mathcal{H}}_t^{\mathcal{N}}$ ,  $\hat{\Sigma}_t^{\mathcal{G}}$  in place of  $\hat{\Sigma}_t^{\mathcal{N}}$ , and the pricing error cross-product matrix in the inverse term based on  $h_{t+1}^{\mathcal{G}}(\theta_T)$  instead of  $h_{t+1}^{\mathcal{N}}(\beta_T)$ . We could also compute the non-robust version of the *Wald* statistic analogous to the corresponding version of the *LM* statistic, but in this case it would not be numerically identical to the *Wald* statistic computed in the traditional way as a quadratic form in  $\gamma_T$  as in Eq. (25) (the numerical equivalence of the portfolio representation shown in Section III.B holds only for the robust version). For the *Wald* test we therefore report the non-robust version in its traditional form as a quadratic form in the  $\gamma_T$  estimates with the asymptotic covariance taken from Eq. (21). Of course, under the null hypothesis and local alternatives, the robust and non-robust statistics and the different ways of computing them are all asymptotically equivalent.

## F Small-Sample Properties

We perform Monte Carlo simulations to investigate the small-sample properties of the estimators employed in our empirical analysis. The results we report here should be regarded as a preliminary first step towards understanding the small-sample properties of optimal-instrument estimators in an asset-pricing setting. The behavior of these estimators is likely to depend in various ways on the specification of the hypothesized data-generating process. Factors that are likely to play a role include the amount of time-variation in various conditional moments, the degree of non-linearity in the conditional moment functions, the specification of the *SDF*, and the length of the data sample. A comprehensive analysis of the behavior of the optimal instruments estimators along these dimensions touches on some deep econometric issues that we cannot hope to adequately address within the scope of this appendix.<sup>2</sup>

We pursue the more limited objective of obtaining some first insights into the small sample properties of the optimal IV estimator under a specific null hypothesis that is consistent in many ways with the empirical evidence on time-varying conditional moments that we reported in our paper (NS). Given the poor empirical performance of the *SDF* candidates analyzed in the main paper, we have to choose whether to generate data under a null that would seem reasonable based on theoretical considerations (e.g., with reasonable implied relative risk aversion) or one that matches the empirical data well. Here we choose the latter, which means we pick *SDF* parameters that generate mean returns and time-variation of conditional expected returns close to what is found in the empirical data.

We simulate returns of five assets and these returns are assumed to be consistent

---

<sup>2</sup>In fact, the literature on small-sample properties of *GMM* estimators in asset-pricing applications is sparse to begin with (Tauchen (1986), Hansen, Heaton, and Yaron (1996), Ferson and Siegel (2003)).

with a linearized pricing kernel of the type that we investigate empirically in NS:

$$m_{t+1}^G(\theta_0) = (\beta_1 + \gamma_1 z_t) + (\beta_2 + \gamma_2 z_t) \Delta c_{t+1}. \quad (\text{A.12})$$

Combining the pricing kernel with the pricing restriction, Eq. (1) in NS, and conditioning on the state variable  $z_t$ , we obtain

$$E[r_{t+1}|z_t] = \frac{p_t - (\beta_2 + \gamma_2 z_t) \text{Cov}(f_{t+1}, r_{t+1}|z_t)}{\beta_1 + \gamma_1 z_t + (\beta_2 + \gamma_2 z_t) E[f_{t+1}|z_t]}. \quad (\text{A.13})$$

To generate artificial data on conditional expected returns consistent with this pricing model, we need to model the dynamics of  $z_t$ . Given a process for  $z_t$  we then need to make sure the *SDF* parameters and the dynamics of  $\text{Cov}(f_{t+1}, r_{t+1}|z_t)$  and  $E[f_{t+1}|z_t]$  are consistent with  $E[r_{t+1}|z_t]$  according to Eq. (A.13).

Regarding the dynamics of  $z_t$ , we assume a homoskedastic AR(1) with normally distributed innovations, and we set the AR(1)-parameters equal to the point estimates that we obtain from estimating an AR(1) for the conditioning variable *cay* used in NS.

We assume that the risk factor  $f_{t+1}$  is mean zero with IID normal innovations and variance equal to the variance of consumption growth in our empirical data sample. Conditional correlations between returns and  $f_{t+1}$  for assets 1 to 4 (the simulated equity portfolios) are assumed to follow the quadratic function  $0.30 - 200(z_t - 0.01)^2$ . This delivers conditional expected cross products of returns and  $f_{t+1}$  that are roughly consistent with those that we reported with *cay* as predictor in the empirical analysis. For asset 5 (the simulated Treasury Bill), we assume a correlation of zero.

Given the simulated  $z_t$  and  $f_{t+1}$ , we choose *SDF* parameters  $\beta_2$  and  $\gamma_2$  such that the term  $\beta_2 + \gamma_2 z_t$  (which corresponds approximately to a time-varying relative risk aversion coefficient) has mean 200 and standard deviation 70. These parameter values

allow us to roughly match the mean and standard deviation of conditional expected stock returns from the local linear conditional moment estimates in NS.

We further proceed to pick  $\gamma_1$  so that the standard deviation of the conditional mean return of the conditionally risk free asset; that is,  $1/E[m_{t+1}^G|z_t]$  matches the standard deviation of the conditional mean of the real T-Bill return, where the latter is obtained from the local polynomial conditional moment estimates in NS. We choose  $\beta_1$  so that the mean of  $1/E[m_{t+1}^G|z_t]$  matches the mean real T-Bill return.

Given the  $\text{Cov}(f_{t+1}, r_{t+1}|z_t)$  and  $\text{Var}(f_{t+1}|z_t)$  as specified above, we simulate return innovations from a conditional one-factor factor model. The factor related component is  $\text{Cov}(f_{t+1}, r_{t+1}|z_t)\text{Var}(f_{t+1}|z_t)^{-1}f_{t+1}$ . We then add an IID normal residual for each asset (uncorrelated between assets) to match the unconditional variance of the unexpected return of the four stock portfolios and the T-Bill in the empirical data (i.e., the residuals from the local polynomial regressions estimates in NS). This completes the specification of the joint dynamics of  $z_t$ ,  $f_{t+1}$ , and  $r_{t+1}$ .

We generate 5,000 Monte Carlo samples. In each Monte Carlo sample, we generate 219 observations, the same sample size (in quarters) as our data set in NS. In each Monte Carlo sample, we apply the same types of estimators as in NS: unconditional, fixed IV with instruments  $w_t = (1, z_t, r_t, f_t)$ , and optimal IV with local polynomial estimates of conditional moments. For the local polynomial estimation we perform a data-driven bandwidth selection with cross-validation, as in NS.

Ensuring a global optimum for the optimal IV estimator across all simulations can be a challenge. For example, the numerical non-linear equation solver might run off towards a “solution” with extremely large *SDF* parameters which make  $E[A_t h_{t+1}|z_t]$  close to zero not by making  $h_{t+1}$  small, but instead by blowing up  $E[h_{t+1} h'_{t+1}|z_t]$  (which appears with an inverse in  $A_t$ ) to huge values. A method that we found to work well is to

first construct preliminary fixed IV estimates, using the first few principal components of  $E[\partial h_{t+1}/\partial\theta|z_t]$  (taken from the local polynomial estimation) as instruments, and then using these preliminary estimates as initial values for the non-linear equation solver, supplemented if necessary with an extensive grid search over initial values.

It also helps to impose a common bandwidth  $b_k$  in the estimation of the two conditional moments  $g_k(z_t) = E[(r'_{k,t+1}, \Delta c_{t+1} r'_{k,t+1})|z_t]$  corresponding to asset  $k$ . In some of the Monte-Carlo samples, the local polynomial regressions can produce quite extreme values for the estimates of  $E[r'_{k,t+1}|z_t]$  or  $E[\Delta c_{t+1} r'_{k,t+1}|z_t]$  for outlier observations of  $z_t$ , and this seems to be more of a problem if only one of the two elements of the estimate of  $g_k(z_t) = E[(r'_{k,t+1}, \Delta c_{t+1} r'_{k,t+1})'|z_t]$  is affected (because is estimated with a small bandwidth), while the other is not (because it is estimated with a wide bandwidth). Imposing the same bandwidth ensures that  $E[r'_{k,t+1}|z_t]$  and  $E[\Delta c_{t+1} r'_{k,t+1}|z_t]$  are estimated from the same local neighborhood around  $z_t$ .

Figure 1 presents the Monte Carlo density of the parameter estimates. The estimates from fixed IV and optimal IV estimators are considerably more precise and better centered around the true parameter values than the estimates based on the unconditional estimator. For the  $\beta_1$  and  $\gamma_1$  estimates, the fixed IV estimates seem to be slightly more precise, but for the  $\beta_2$  and  $\gamma_2$  estimates, the fixed IV estimates show considerably higher dispersion and also some bias. For  $\beta_2$ , the RMSE of the fixed IV estimates is about five times as big as with the optimal IV estimator. Overall, the optimal IV estimates look well behaved.

Figure 2 plots the empirical distribution of  $p$ -values from the  $\tau(I)$  test to illustrate the actual size of the test in relation to its nominal size. The test based on the unconditional estimator under-rejects compared with the nominal size of the test. The test based on the fixed IV estimator severely over-rejects. Its actual size is much



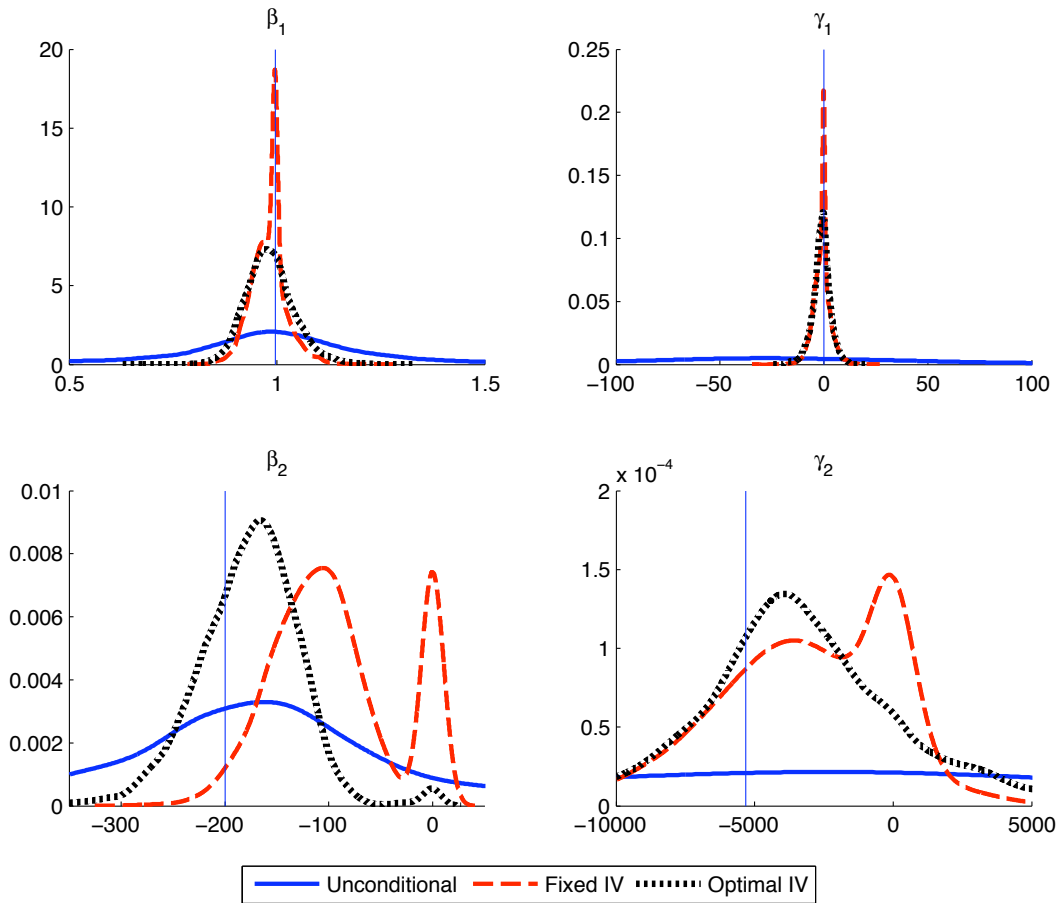


Figure 1: Kernel-smoothed Monte Carlo density of SDF parameter estimates. The vertical line indicates the true parameter value.

higher than the nominal size of the test, particularly for small nominal sizes. This is a consequence of the large number of instruments relative to the sample size (which is also often typical in empirical applications of the fixed IV estimator). If one reduced the number of instruments, the tendency to over-reject would likely be reduced. In the extreme case of only a constant as the “instrument”, the fixed IV estimator becomes the unconditional estimator. The  $\tau(I)$  test based on the optimal IV estimator also exhibits a tendency to over-reject, but considerably less so than the test based on the fixed IV estimator, a likely consequence of the fact that it does not use a large

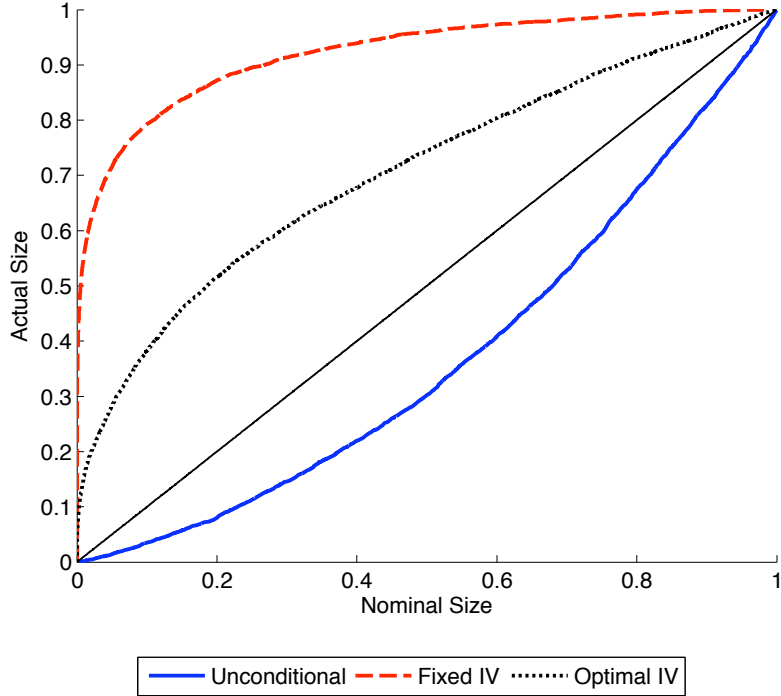


Figure 2:  $p$ -value plots for  $\tau(I)$  test of zero average pricing errors

number of moment conditions in the construction of the test statistic. Nevertheless, an interpretation of empirical results based on the  $\tau(I)$  test statistic should take into account this tendency to over-reject.

Next, we investigate the size and power of the Wald and LM tests of  $H_0 : \gamma_1 = 0, \gamma_2 = 0$ . To generate data under this null hypothesis, we simulate from the  $SDF$   $m_{t+1}^N$  with  $\gamma_1 = 0, \gamma_2 = 0, \beta_2 = -200$ , and  $\beta_1$  chosen such that  $1/E[m_{t+1}^N|z_t]$  matches the mean gross return on Treasury Bills in our sample.

Figure 3 compares actual and nominal sizes of the Wald and LM tests with data generated under the null  $m_{t+1}^N$ . For the Wald statistic, the unconditional estimator produces an under-sized test, while the tests based on the fixed IV and optimal IV estimators tend to over-reject the null. For the LM statistic, all three estimators produce tests that are much closer to the correct size.

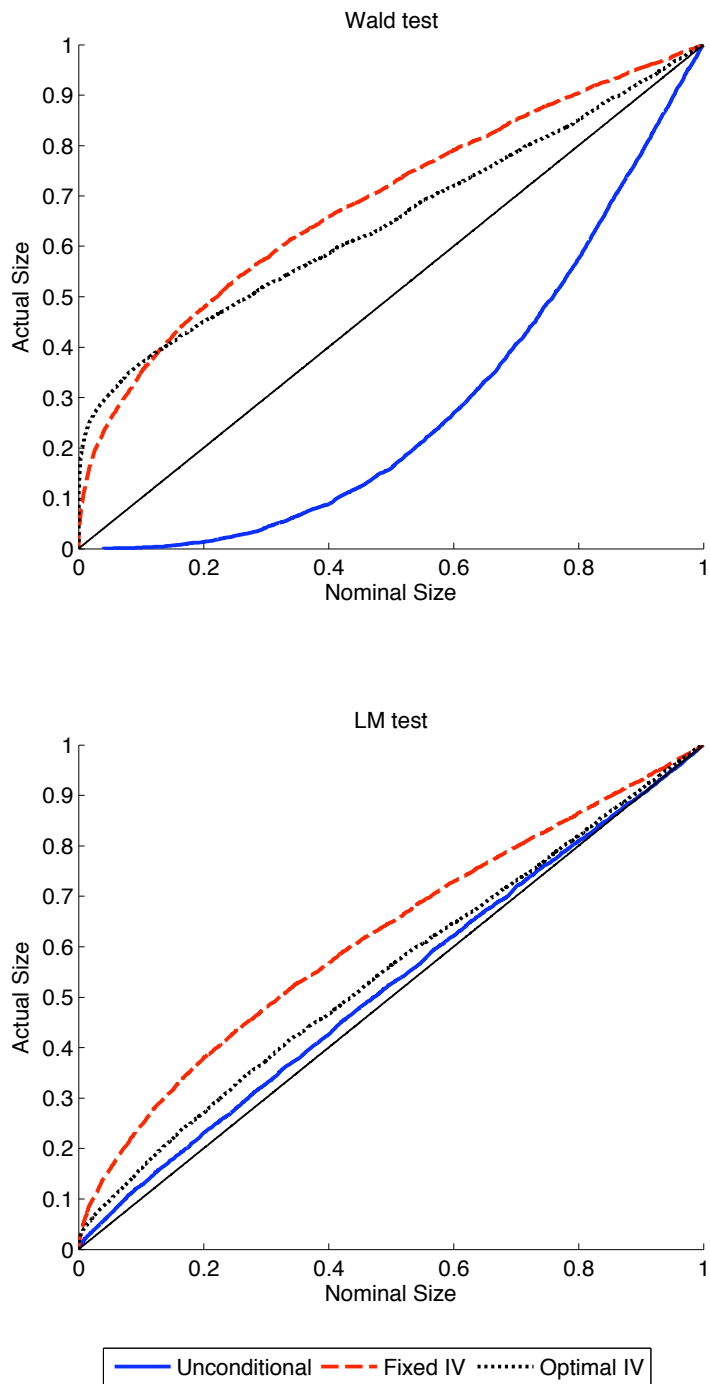


Figure 3:  $p$ -value plots for Wald and LM tests of  $m_{t+1}^N$

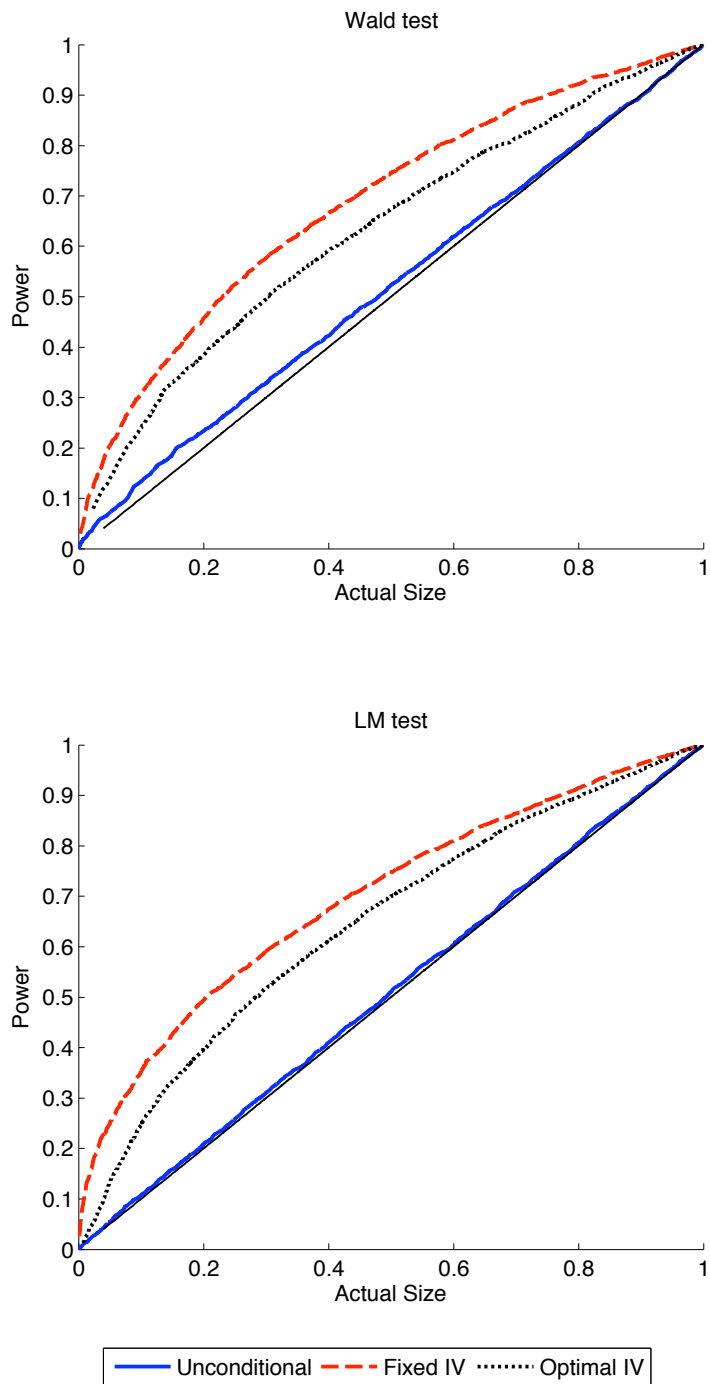


Figure 4: Size-power plots for Wald and LM tests of  $m_{t+1}^N$

Figure 4 shows the results of a simple and preliminary investigation of the power of the Wald and LM tests with different estimators. This analysis is preliminary in the sense that we investigate the power only under one alternative hypothesis, the *SDF*  $m_{t+1}^{\mathcal{G}}$  that we described above. Power depends on the specification of the alternative, and so with different alternatives, results may be different. To take into account the fact that the Wald and LM tests are not always correctly sized, particularly for the Wald test (see Figure 3), we investigate power not as a function of nominal size (which would ignore the size distortions of the test), but as a function of actual size. We do this by plotting the empirical distribution function of  $p$ -values under the  $m_{t+1}^{\mathcal{N}}$  null (as a function of nominal size) against the empirical distribution of  $p$ -values under  $m_{t+1}^{\mathcal{G}}$  alternative (as a function of nominal size). For example, this means that we ask how often the tests rejects under the null at nominal size of 0.05, and we plot this number against the proportion of the simulations under the alternative that lead to rejection at a nominal size of 0.05.

As Figure 4 shows, the Wald and LM tests based on the unconditional estimator essentially have no power in our setting. The tests reject as frequently under the null as they do under the alternative hypothesis. The fixed IV and optimal IV estimators have similar properties and are more powerful than the test based on the unconditional estimator. However, if a size correction is implemented, as in these plots of power against actual size instead of nominal size, they have only moderate power. For example, with nominal size of the LM test set such that actual size is 0.10 (this test rejects 10% of the time under the null), the tests based on fixed and optimal IV estimators reject around 30% of the time under the alternative. Clearly, these results will be sensitive to the distance between  $(\gamma_1, \gamma_2)$  under the null and alternative, as well as the sample size.

Overall, our preliminary analysis suggests that the optimal IV estimator is reason-

ably well behaved in small samples. It shares some of the over-rejection problems of the fixed IV estimator, but we found some indication that the optimal IV estimator may have some advantages over fixed IV estimators that employ a large number of moment conditions. An interesting question that we leave for future research is the extent of the efficiency gains and increased power from using the optimal IV estimator with larger sample sizes or different specifications of the null hypothesis.

## References

- Eichenbaum, M., L. Hansen, and K. Singleton, 1988, A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice Under Uncertainty, *Quarterly Journal of Economics* 103, 51–78.
- Ferson, W., and A. Siegel, 2003, Stochastic Discount Factor Bounds with Conditioning Information, *Review of Financial Studies* 16, 567–595.
- Hansen, Lars P., 1982, Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* 50, 1029–1054.
- Hansen, Lars P., John C. Heaton, and Amir Yaron, 1996, Finite-Sample Properties of Some Alternative GMM Estimators, *Journal of Business and Economic Statistics* 14, 262–280.
- Tauchen, G., 1986, Statistical Properties of GMM Estimates of Structural Parameters Using Financial Market Data, *Journal of Business and Economic Statistics* 4, 397–416.