Estimation and Evaluation of Conditional Asset Pricing Models

Stefan Nagel*
Stanford University and NBER

Kenneth J. Singleton†
Stanford University and NBER

November 3, 2009

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 address two methodological issues that are central to assessing the goodness-of-fit of asset pricing models in which the stochastic discount factor (SDF) is a conditionally affine function of a set of priced risk factors. First, we show that there is an *optimal* GMM estimator for this class of SDFs. That is, there is a choice of instruments that leads to the most efficient estimator within a class that subsumes virtually all of the GMM estimators used to date in assessing the fit of conditionally affine factor models. Second, 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. The form of the optimal choice of test portfolios is derived directly from the (locally) most powerful Wald and Lagrange-multiplier tests of the null against the alternative specification of the SDF.

*Stanford University, Graduate School of Business, 518 Memorial Way, Stanford, CA 94305, e-mail: Nagel_Stefan@gsb.stanford.edu, http://faculty-gsb.stanford.edu/nagel

†Stanford University, Graduate School of Business, 518 Memorial Way, Stanford, CA 94305, e-mail: kenneths@stanford.edu,http://www.stanford.edu/ kenneths/

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 $(\phi_t^0, \phi_t^{f\prime})$ are in the modeler's information set $\mathcal{J}_t$, and θ_0 is an unknown vector of parameters. *SDF*'s 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 *SDF*s, researchers typically posit an R-vector of "test-asset" returns r_{t+1}, construct GMM estimators θ_T of θ_0, and then examine whether the test asset payoffs are correctly priced by the candidate *SDF*; that is, whether $T^{-1} \sum_{t=1}^{T} (m_{t+1}(\theta_T) r_{t+1} - \imath)$ is close to zero, where $\imath$ is an R-vector of ones. Based on these assessments, several candidate SDFs have been found to adequately describe the expected excess 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 (2008), 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[1]

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

[1]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 $E\left[m_{t+1} r_{i,t+1} | \mathcal{A}_t\right] = 1$, where $\mathcal{A}_t$ is agents' information set. Conditioning down to the econometrician's information set $\mathcal{J}_t$ gives this pricing relation.

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 θ_T of θ_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 θ_0. Yet in practice instrument selection for GMM estimation has not been tied to the specific formulation of a 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 θ^*_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 $B_t r_{t+1}$:

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

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 θ_T^* of θ_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

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

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 extended $ICAPM$s examined by Lettau and Ludvigson (2001b) ($z_t = CAY_t$) and Santos and Veronesi (2006) ($z_t =$ the ratio of labor income to total income) where $m_{t+1}^{\mathcal{N}}$ is the pricing kernel induced by constant relative risk averse preferences. Also included is the conditional *CAPM* of 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 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 θ_0^* are the (locally) most powerful chi-square tests against the alternative hypothesis that the pricing kernel is $m^{\mathcal{G}}_{t+1}$. Moreover, these optimal tests can be reinterpreted as tests of the null hypothesis $E[B^*_t(m^{\mathcal{N}}_{t+1}(\beta_0)r_{t+1}-\imath)] = 0$, 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 $m^{\mathcal{N}}_{t+1}$ against the alternative $m^{\mathcal{G}}_{t+1}$. The portfolio weights B^*_t take an economically intuitive form: letting $h_{t+1}(\theta_0) = (m^{\mathcal{G}}_{t+1}(\theta_0)r_{t+1} - \imath)$ 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 $m^{\mathcal{G}}_{t+1}$– that is conditionally orthogonal to its counterpart for the parameters β of the null specification, $E[\partial h_{t+1}(\theta_0)/\partial\beta|\mathcal{J}_t]$. Maximal power is achieved using the optimal portfolio weights B^*_t *and* evaluating m_{t+1} at the efficient GMM estimator θ^*_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 SDFs. The characterization of the optimal choice of managed-portfolio weights B^*_t for maximizing the power of tests of $m^{\mathcal{N}}_{t+1}$ against the alternative $m^{\mathcal{G}}_{t+1}$ 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 non-parametric estimation strategy conditioning on the source z_t of the state-dependence of the SDF weights $\phi^f(z_t, \theta_0)$. The other is a semi-nonparametric strategy based with conditioning on a 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. These findings are explored in more depth by examining the model-implied pricing errors and time series of coefficients of relative risk aversion. While these model seemingly do quite well in fitting the cross-section of average returns of size and book-to-market portfolios when estimation and testing is 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 hide enormous time-variation in *conditional* pricing errors.

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 $\tilde{\phi}_t' = (\phi_t^0, \phi_t^{f\prime}) \in \mathcal{J}_t$ 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,[2]

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

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

[2]This follows from the observation that

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

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).

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 ϕ_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

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

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

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

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

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

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 beta's 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 ϕ_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, $E[m_{t+1}(\theta_0)(r_{t+1} - \imath r_t^f)|\mathcal{J}_t] = 0$. 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.

[3] More generally, the links are between the return on a zero-beta portfolio and the conditional mean of m_{t+1}.

II Efficient GMM Estimation of Affine SDFs

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 θ_0 (and hence $B_t h_{t+1}(\theta_0)$) is unknown. Standard practice is to first construct a GMM estimator θ_T of θ_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 θ_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 θ_0 are satisfied. Accordingly, we begin our discussion of the estimation of θ_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 θ_0 governing the SDF is constructed from the moment condition[4]

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

for some $K \times R$ matrix A_t with entries in $\mathcal{J}_t$. Since (9) constitutes K equations in the K unknowns θ_0, we can define the *GMM* estimator θ_T^A of θ_0, indexed by the modeler's choice of instrument process $\{A_t\}$, as the value of θ that solves

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

[4]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.

Under regularity, the asymptotic covariance matrix of θ_T^A is (Hansen (1982))

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

where[5]

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

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

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

In Appendix A we show that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} B_t h_{t+1}(\theta_T^A) \xrightarrow{\mathcal{D}} N(0, \Gamma_0^A), \quad \Gamma_0^A = E[C_t^A \Sigma_t C_t^{A\prime}], \tag{14}$$

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

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

The form of C_t^A reflects the fact that pre-estimation of θ_0 using the instruments A_t

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

affects the asymptotic distribution of the sample mean (13). It follows that

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

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

where $\stackrel{a}{=}$ 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

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

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 θ_T^A is a consistent estimator of θ_0. The sample counterpart of the left-hand side of (18) is (13), because θ_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.

II.A Optimal *GMM* Estimation of Conditional Factor Models

If we index each estimator θ_T^A by its associated instrument matrix A_t, then we can define the admissible class of *GMM* estimators as[6]

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

Researchers have considerable latitude in selecting the sequence of matrices $\{A_t\}$ to construct a consistent estimator of θ_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 θ_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 $A_t \in \mathcal{A}$.[7]

The most efficient *GMM* estimator is the one that produces the smallest Ω_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). Specifi-

[6]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.

[7]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.

cally, the optimal choice is

$$A_t^* = \Psi_t^{\theta\prime}\Sigma_t^{-1}, \text{ where } \Psi_t^\theta \equiv E\left[\frac{\partial h_{t+1}(\theta_0)}{\partial\theta}\Big|\mathcal{J}_t\right], \tag{20}$$

and the associated asymptotic covariance matrix is

$$\Omega_0^* = \left(E\left[\Psi_t^{\theta\prime}\Sigma_t^{-1}\Psi_t^\theta\right]\right)^{-1}. \tag{21}$$

The first term in the definition of A^*, $\Psi_t^{\theta\prime}$, 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).[8] The post-multiplication by Σ_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,[9] for models with conditionally affine pricing kernels of the form (3), the building blocks of A_t^* take tractable forms. Specifically, letting $\tilde{\phi}(z_t,\theta_0)' = (\phi^{0\prime}(z_t,\theta_0), \phi^{f\prime}(z_t,\theta_0))$ and $\tilde{f}_{t+1}' = (1, f_{t+1}')$, a typical element of the first term in (20) takes the form

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

[8]This step is exactly analogous to the projection of "right-hand-side" regressors onto the predetermined variables in $2SLS$ and $3SLS$ estimation. In linear models, these regressors comprise the partial derivatives of the equation error with respect to θ_0.

[9]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.

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,

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

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 $r_{i,t+1}r_{j,t+1}f_{k,t+1}f_{\ell,t+1}$). Once again, any nonlinearity inherent in the specification of the factor weights $\tilde{\phi}$ does not add complexity to the computation of the optimal instruments.

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.[10] 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,

[10]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.

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 θ_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 dependence of A^* on conditional moments does raise the practical question of whether, in deriving the large-sample distribution of θ_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 Ψ_t^θ and Σ_t. In this case, the asymptotic covariance matrix of θ_T^* is (21).

The second case arises when either non-parametric or semi-non-parametric methods are used to estimate conditional moments. Many of these methods have the property that the quality of the approximations to the true functional forms of Ψ_t^θ and Σ_t improve with sample size. (That is, one employs increasingly flexible specifications of the approximating functions as T increases.) However these approximations often do not converge at a sufficiently fast rate to ensure that $(1/\sqrt{T})\sum_{t=1}^{T}\tilde{A}_t^{*T}h_{t+1}(\theta_0)$ converges in distribution to a normal random variable, where we have used the notation $\tilde{A}_t^{*T}$ to denote the approximation of A_t^* used for a sample size of T. In our subsequent illustrations we use both non-parametric and (what we think of as) semi-non-parametric specifications of Ψ_t^θ and Σ_t and, for our sample size, neither may give completely accurate representations of A_t^*. With this possibility in mind, we report two sets of standard errors: those based on a consistent estimator (21), presuming that our specification $\tilde{A}_t^{*T}$ is equal to A_t^*; and the counterpart based on (11) which treats $\tilde{A}_t^{*T}$ as a generic instrument matrix. The standard errors computed by the latter method are robust to any approximation error inherent in using $\tilde{A}_t^{*T}$ in place of A_t^*.

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

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

where $\Gamma_T^{A^*}$ is a consistent estimator of $\Gamma_0^{A^*} = E[C_t^{A^*} \Sigma_t^{-1} C^{A^*\prime}]$. The robust version of our chi-square statistic evaluates (16) directly at $\tilde{A}_t^{*T}$.

II.B The Wald Test with Maximal Power

Consider again the case where the goal is an evaluation of the improvement in fit of $m_{t+1}^{\mathcal{G}}(\beta_0, \gamma_0)$, as given by (3), relative to the null specification $m_{t+1}^{\mathcal{N}}(\beta_0)$ obtained as the special case with $\gamma_0 = 0$. Suppose that θ_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 θ_T^A is given by (11). Letting $\Omega_{\gamma\gamma}^A$ denote the lower-diagonal $G \times G$ block of Ω_0^A, where G is the dimension of γ_0, it follows under $H_0 : \gamma_0 = 0$ that

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

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 θ_0 affects the power of tests of fit. In order to explore this dependence we focus on the *local* alternative $H_{1T} : m_{t+1}^{\mathcal{G}}(\beta_0, \gamma = \gamma_T^{\mathcal{L}})$, for which the parameter sequence $\gamma_T^{\mathcal{L}}$ converges to the null of $\gamma_0 = 0$ at the rate $\sqrt{T}$: $\gamma_T^{\mathcal{L}} = \delta/\sqrt{T}$, for some nonzero $G \times 1$ vector δ of proportionality constants.[11] Under this local alternative,[12] $\sqrt{T}\left(\gamma_T^A - \gamma_0\right) \xrightarrow{\mathcal{D}} N\left(\delta, \Omega_{\gamma\gamma}^A\right)$. It follows that the asymptotic distribution of $\varsigma_T^W(A)$ is that of a non-central chi-square distribution with G degrees of freedom and non-centrality parameter

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

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 γ_T^A. The more econometrically efficient is the estimator γ_T^A of γ_0, the smaller is this covariance matrix and the higher is the power of the associated test based on $\varsigma_T^W(A)$. 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 $m_{t+1}^{\mathcal{N}}$ against the alternative specification $m_{t+1}^{\mathcal{G}}$.

III Portfolio Selection for Maximal (Local) Power

Though the construction of the $Wald$ statistic $\varsigma_T^W(A^*)$ 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

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

[12]This form of the asymptotic distribution of γ_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).

intimate connection to this issue. Indeed, tests based on $\varsigma_t^W(A^*)$ 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 θ_0 governing $m_{t+1}^{\mathcal{G}}$, the *Wald* statistic $\varsigma_T^W(A^*)$ can be expressed in the asymptotically equivalent form (see Appendix B)

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

where

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

$\mathcal{K}^{\beta\gamma} \equiv E\left[\Psi_t^{\beta\prime}\Sigma_t^{-1}\Psi_t^{\gamma}\right]$, and $\mathcal{H}_t \equiv \Psi_t^{\gamma} - \Psi_t^{\beta}\left(\mathcal{K}^{\beta\beta}\right)^{-1}\mathcal{K}^{\beta\gamma}$. 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

$$E\left[\mathcal{H}_t^{\mathcal{G}\prime}\Sigma_t^{\mathcal{G}-1}h_{t+1}(\theta_0)\right] = 0; \qquad (28)$$

that is, the Wald test evaluates whether the managed portfolio returns $\mathcal{H}_t^{\mathcal{G}\prime}\Sigma_t^{\mathcal{G}-1}r_{t+1}$ are priced by $m_{t+1}^{\mathcal{G}}$. Factoring Σ_t^{-1} as $D_t^{-1/2\prime}D_t^{-1/2}$, the component $D^{-1/2}\mathcal{H}_t^{\mathcal{G}}$ 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 $E[D^{-1/2}\Psi_t^{\beta\prime}\Sigma_t^{\mathcal{G}-1}h_{t+1}(\theta_0)] = 0$ 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

$$m^{\mathcal{G}}_{t+1}(\theta_0) = (\beta_1 + \gamma_1 z_t) + (\beta_2 + \gamma_2 z_t)\,\Delta c_{t+1}. \tag{29}$$

The model in Lettau and Ludvigson (2001b) is the special case with z_t equal to CAY, and Santos and Veronesi (2006) examined a model in which z_t was equal to the ratio of labor income to total income. 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,

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

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

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 $E[r_{t+1} z_t \mid \mathcal{J}_t]$ and $E[\Delta c_{t+1} r_{t+1} z_t \mid \mathcal{J}_t]$ that are orthogonal (in a linear projection sense) to the information contained in $E[r_{t+1} \mid \mathcal{J}_t]$ and $E[\Delta c_{t+1} r_{t+1} \mid \mathcal{J}_t]$.[13]

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 Δc_{t+1} is examined. Equally importantly, it is not the

[13] More precisely, we are projecting the scaled versions of these constructs on each other, where scaling is by the square root of Σ_t^{-1}, as discussed above.

projection of the factors themselves onto $\mathcal{J}_t$ that is relevant for portfolio construction, but rather the return-augmented projections $E[r_{t+1}z_t \mid \mathcal{J}_t]$ and $E[\Delta c_{t+1}r_{t+1}z_t \mid \mathcal{J}_t]$ are used. Among other considerations, this observation leads us to examine the conditional second moment $E[\Delta c_{t+1}r_{t+1} \mid \mathcal{J}_t]$ 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

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

For this example, the relevant expressions related to β_0 are identical to (30) and

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

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

III.A Optimal Test Portfolios as Lagrange Multipliers

An alternative approach to deriving the optimal test portfolios starts with constrained estimates using $m_{t+1} = m^{\mathcal{N}}_{t+1}$, 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 Appendix C we show that the Lagrange multiplier for the constraints $\gamma_T = 0$ can be expressed as

$$\lambda_T = \frac{1}{T}\sum_t \Psi_t^{\gamma\prime}\Sigma_t^{\mathcal{N}-1}h_{t+1}^{\mathcal{N}}(\beta_T) \overset{a}{=} \frac{1}{T}\sum_t \mathcal{H}_t^{\mathcal{N}\prime}\Sigma_t^{\mathcal{N}-1}h_{t+1}^{\mathcal{N}}(\beta_0), \tag{34}$$

where $\mathcal{H}_t^{\mathcal{N}}$ is the matrix $\mathcal{H}_t$ evaluated at the constrained $(\beta_0, \gamma_0 0)$. Therefore, the asymptotic distribution of λ_T is normal with mean zero and covariance matrix $E[\mathcal{H}_t^{\mathcal{N}\prime}\Sigma_t^{\mathcal{N}-1}\mathcal{H}_t^{\mathcal{N}}]$, from which it follows that

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

Summarizing our results,

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

Both tests effectively assess whether the managed portfolio returns $\mathcal{H}_t'\Sigma_t^{-1}r_{t+1}$ are correctly priced by m_{t+1}. The difference is that the (locally) most powerful, managed portfolio weights $\mathcal{H}_t^{\mathcal{G}\prime}\Sigma_t^{\mathcal{G}-1}$ underlying the $Wald$ test are evaluated at θ_0, whereas the weights $\mathcal{H}_t^{\mathcal{N}\prime}\Sigma_t^{\mathcal{N}-1}$ 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.

III.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,[14] 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

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

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 $\theta_T^{\mathcal{G}}$

[14] 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.

for the G-subvector $\gamma_T^{\mathcal{G}}$ gives[15]

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

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

$$\hat{\mathcal{K}}_T^{\gamma\beta}(\theta_T^{\mathcal{G}}) \equiv \frac{1}{T}\sum_{t=1}^{T}\left[\hat{\Psi}_t^{\gamma\prime}\hat{\Sigma}_t^{\mathcal{G}-1}r_{t+1}f_{t+1}^{\#\mathcal{N}\prime}\right], \tag{37}$$

the robust, sample version of $E[\Psi_t^{\gamma\prime}\Sigma_t^{\mathcal{G}-1}\Psi_t^{\beta}]$, 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

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

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

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{\mathcal{H}}_t^{\mathcal{G}}\hat{\Sigma}_t^{\mathcal{G}-1}\imath_R = \frac{1}{T}\sum_{t=1}^{T}\widehat{\mathcal{H}}_t^{\mathcal{G}}\hat{\Sigma}_t^{\mathcal{G}-1}h_{t+1}^{\mathcal{N}}\left(\beta_T^{\mathcal{N}}\right). \tag{39}$$

Thus, we can interpret the sample $Wald$ statistic as checking whether the SDF under H_0 prices the managed portfolios $B_t^{Wald} = \widehat{\mathcal{H}}_t^{\mathcal{G}}\hat{\Sigma}_t^{\mathcal{G}-1}$ evaluated at $\theta_T^{\mathcal{G}}$. Recall from

[15]That is, we solve (10), after substitution of the relevant special case of A^* in (20), for $\gamma_T^{\mathcal{G}}$.

Section III.A that the sample moment entering the LM statistic ς_T^{LM} is[16]

$$\frac{1}{T}\sum_t \Psi_t^{\gamma\prime}\hat{\Sigma}_t^{\mathcal{N}-1}h_{t+1}^{\mathcal{N}}(\beta_T) = \frac{1}{T}\sum_{t=1}^{T}\widehat{\mathcal{H}}_t^{\mathcal{N}}\hat{\Sigma}_t^{\mathcal{N}-1}h_{t+1}^{\mathcal{N}}\left(\beta_T^{\mathcal{N}}\right). \tag{40}$$

This expression is identical to (39), except that the managed portfolio weights $B_t^{LM} = \widehat{\mathcal{H}}_t^{\mathcal{N}}\hat{\Sigma}_t^{\mathcal{N}-1}$ are evaluated under the null at $\beta_T^{\mathcal{N}}$. Similarly the matrices that define the quadratic forms ς_T^W and ς_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 SDFs 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,

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

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),

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

or the labor income-consumption ratio of Santos and Veronesi (2006) (yc_t).[17]

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 Treasury 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,

[17] Jagannathan and Wang (1996) and Santos and Veronesi (2006) use these conditioning variables in β-style representations of excess returns, while we use them as conditioning variables in a consumption-based pricing kernel.

due to the strong commonality in the returns of these portfolios (Fama and French (1993); Lewellen, Nagel, and Shanken (2008)). By construction of B_t^{Wald} and B_t^{LM}, we are asking candidate SDFs 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 real returns in excess of real 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).

IV.A Estimation of Conditional Moments

Implementation of the optimal estimator requires estimates of the conditional moments given by (22) and (23)

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

and

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

where $\partial\tilde{\phi}(z_t,\theta_0)'/\partial\theta_0 = (I_2 \otimes \tilde{z}_t)$ for the affine pricing kernels (41) that we consider here. In our empirical implementation, we work with $\mathrm{Var}\left[h_{t+1}(\theta_0)\,|\mathcal{J}_t\right]$ instead of the uncentered $E\left[h_{t+1}(\theta_0)\,h_{t+1}(\theta_0)'\,|\mathcal{J}_t\right]$. Both are equivalent under the null hypothesis, but the centered $\mathrm{Var}\left[h_{t+1}(\theta_0)\,|\mathcal{J}_t\right]$ should be better behaved under misspecification.

To estimate the moments given in (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 semi-nonparametric estimators.

Nonparametric estimators converge asymptotically, under regularity and as the flexibility of the approximating conditional densities increases with sample size, to the true moments conditional on $\mathcal{J}_t$. The downside is that computational considerations typically dictate that non-parametric 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 $y_{t+1} \equiv \left(r'_{t+1}, \Delta c_{t+1} r'_{t+1}\right)'$ 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 $\hat{g}(z_i)$ of the conditional mean function, a linear regression is estimated locally, with weighted least squares based on the nearest neighbors of z_i (in terms of the distance $|z_j - z_i|$). The weights are determined by the kernel function, the distance $|z_j - z_i|$, and the bandwidth b. The fitted value at z_i yields the conditional moment estimate $\hat{g}(z_i)$.

We use the Epanechnikov kernel function,

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

where $u \equiv |z_j - z_i| / b$. The bandwidth b determines the weighting of the neighborhood observations around each point z_i, and hence the smoothness of the estimated function. We allow a different optimal bandwidth b_k^* for the estimation of each element of $g(z_i)$. To determine b_k^*, we use automatic bandwidth selection by leave-one-out cross-validation, i.e.,

$$b_k^* = \arg\min_{b_k} \frac{1}{T} \sum_{i=1}^{T} (y_{ik} - \hat{g}_{k,-i}(z_i))^2 ,$$

where $\hat{g}_{k,-i}(z_i)$ denotes the local linear regression estimate of the k-th element of $g(z_i)$ with bandwidth b_k that is obtained when observation i is not included in the estimation.[18] As $T \to \infty$, and more and more observations exist in the neighborhood of z_i, the optimal bandwidth shrinks, and the nonparametric regression estimates converge to the true conditional moments.

To estimate $\Omega(z_t) \equiv \mathrm{Var}[(r'_{t+1}, \Delta c_{t+1} r'_{t+1})' | z_t]$ we calculate the residuals $y_{t+1} - \hat{g}(z_t)$ from the "first step" nonparametric regressions, and we use all elements of the cross-product matrix of these residuals as the dependent variables for "second step" nonparametric regressions. We make two modifications compared with the "first stage" 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

[18]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)).

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

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

where $\hat{\Omega}_{-i}(z_t)$ denotes the estimate of $\hat{\Omega}_{-i}(z_t)$ obtained with observation i 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, only the conditional expectation of the T-Bill exhibits substantial 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. Looking across the results for all three conditioning variables, it is apparent that in each case the estimated conditional mean functions are quite similar for all four equity portfolio returns.

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.[19] 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

[19]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 yiels virtually unchanged results.

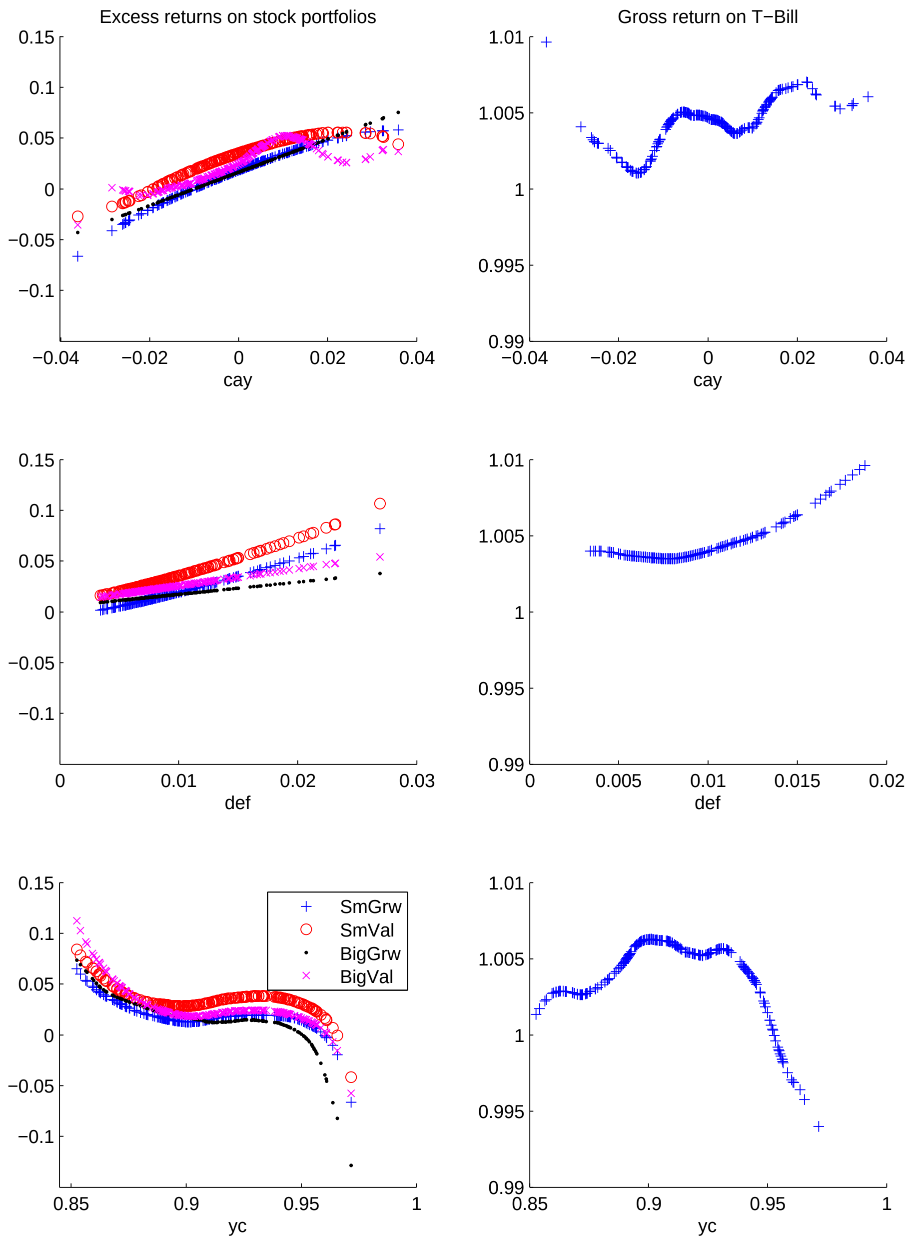

Figure 1: Fitted conditional first moments: conditional expected returns

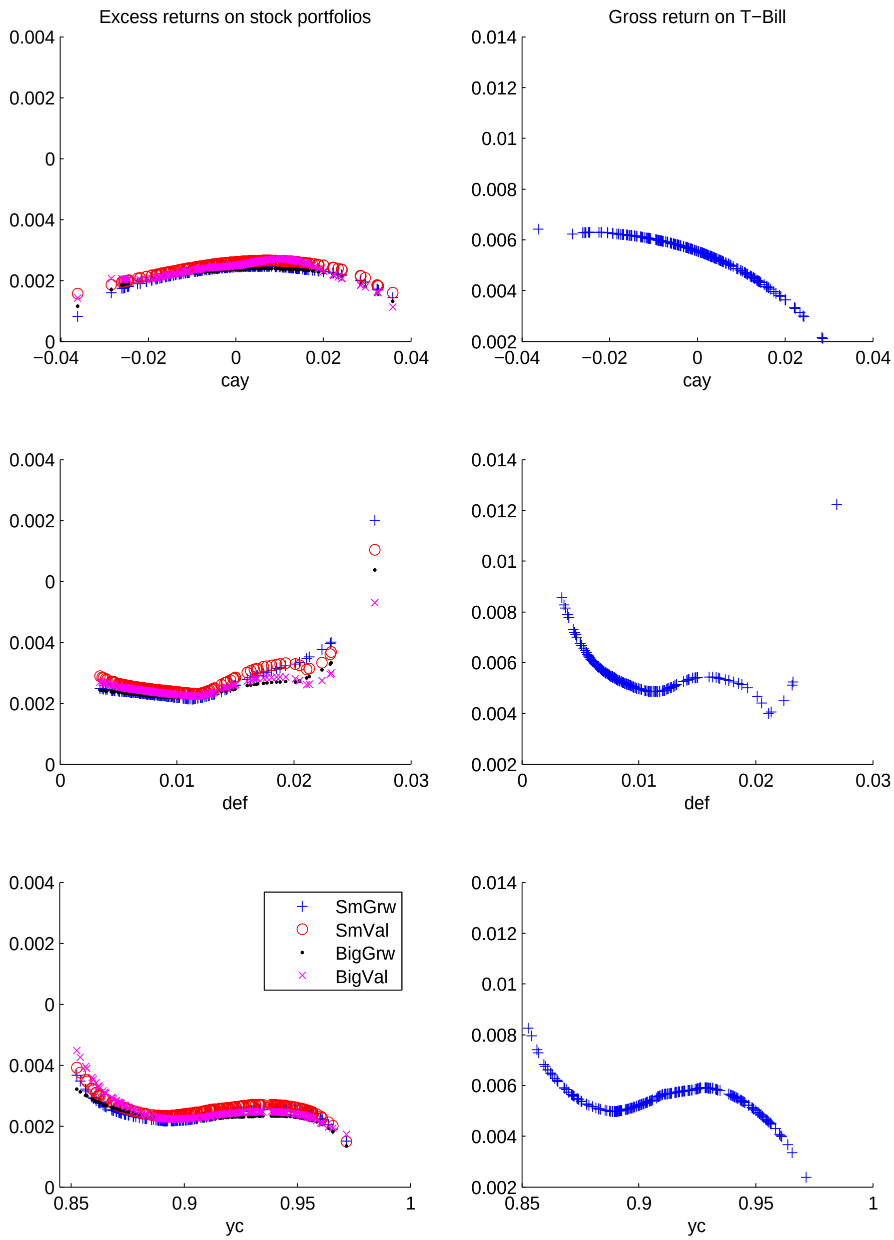

Figure 2: Fitted conditional second moments: conditional expected cross-product of return and log consumption growth

again common to the four stock portfolios.

Overall, the non-parametric regressions pick up considerable time-variation in conditional moments related to cay_t, def_t, and yc_t. 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 essentially infinity for $zt = cay$ and $z_t = yc$, and it is a still high 0.66 for def. This means that the estimated $\Omega(z_t)$ is the constant unconditional sample covariance for cay and yc, and it does not have much time-variation for def. 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 important time-variation in $g(z_t)$, i.e., from predictability of returns and cross-products of returns and consumption growth, not from the higher moments captured by $\Omega(z_t)$.

As an alternative to the fully nonparametric estimates of conditional moments we employ a semi-nonparametric estimator. 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 $x_t \equiv (r_t, \Delta c_t, z_t, z_t^2, 1/z_t)$.[20] We use the sample covariance matrix of these residuals 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 non-parametric case discussed above, as well as a paucity

[20]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.

of evidence for significant conditional heteroskedasticity in quarterly returns and consumption growth.[21]

While this semi-nonparametric method is potentially less flexible in adapting to highly non-linear functional forms than the fully non-parametric method, it allows us to condition on a broader set of instruments that includes $(r_t, \Delta c_t)$. To reduce the possibility of overfitting the conditional moments $E[r_{t+1}|\mathcal{J}_t]$ and $E[r_{t+1}\Delta c_{t+1}|\mathcal{J}_t]$, we use the Akaike Information Criterion (AIC) to select regressors. We calculate the AIC for all specifications that use any possible combination of $(1, r_t, \Delta c_t, z_t, z_t^2, 1/z_t)$ and, for each element of the conditional mean vector, we choose the specification with the minimum AIC. The resulting estimates of $E[r_{t+1}|\mathcal{J}_t]$ and $E[r_{t+1}\Delta c_{t+1}|\mathcal{J}_t]$ look very similar to those obtained with the nonparametric method; in particular, they capture well the linear, parabolic, and "S on its side" patterns displayed in Figures 1 and 2.

We emphasize again that, for valid inference, it is not necessary to assume that these non-parametric and semi-non-parametric estimators $\tilde{A}_t^{*T}$ perfectly match the population counterpart A_t^*. In cases where one is concerned about the accuracy of these approximations in small samples, the statistic $\tau_T(B, \tilde{A}^{*T})$ based on (16) should be used in place of the statistic $\tau_T(B, A^*)$ given by (24).

[21] We experimented with time-varying conditional covariance matrix from a dynamic conditional correlation (DCC) model (Engle (2002)), but the evidence for GARCH effects and time-varying is weak. Moreover, allowing for a time-varying conditional covariance matrix has only negligible effects on our asset-pricing results. Accordingly, we proceed with the simpler specification outlined above.

IV.B Estimators and Test Statistics

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

$$E\left[m_{t+1}\left(\theta_0\right)r_{t+1} - p_t\right] = 0, \tag{44}$$

where the components of p_t are 1 for the case of returns and 0 for the case of excess returns. The second (denoted "fixed IV") is based on the LR moment restrictions,

$$E\left[\left(m_{t+1}\left(\theta_0\right)r_{t+1} - p_t\right) \otimes w_t\right] = 0, \tag{45}$$

where $w_t = (1, r_t', \Delta c_t, z_t)'$ is an $L \times 1$ vector, and z_t equals cay_t, def_t, or yc_t, depending on the asset-pricing model. Our third estimator (denoted "optimal IV – NP" is our optimal *GMM* estimator, based on the K moment restrictions

$$E\left[A_t^*\left(m_{t+1}\left(\theta_0\right)r_{t+1} - p_t\right)\right] = 0, \tag{46}$$

and nonparametrically estimated conditional moments. Finally, we let "optimal IV – SNP" denote the optimal *GMM* estimator based on conditional moments from the semi-nonparametric model.

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 θ_T with both A_t^* and m_{t+1} depending on θ_T and, thus, this calculation is analogous to the continuously-updated GMM estimator.

For each of the choices of GMM estimator θ_T^A we present three test statistics for model evaluation: $\tau_T(I)$, for the null hypothesis that the means of the "pricing errors"

Table I: Test Statistics

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

(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,

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

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.

Finally, the estimators of the asymptotic variances of our estimators and the weighted pricing error measures involve terms like

$$E\left[\Psi_t \Sigma_t^{-1} \frac{\partial h_{t+1}(\theta_0)}{\partial \theta}\right] \tag{48}$$

or

$$E\left[\Psi_t \Sigma_t^{-1} h_{t+1}(\theta_0) h_{t+1}(\theta_0)' \Sigma_t^{-1} \Psi_t'\right], \tag{49}$$

that both reduce to $E[\Psi_t \Sigma_t^{-1} E[\frac{\partial h_{t+1}(\theta_0)}{\partial \theta} | \mathcal{J}_t]] = E[\Psi_t \Sigma_t^{-1} \Psi_t']$ under the assumption that conditional moments are correctly specified. For example, the constructs Ω_0^* and $\mathcal{H}_t^*$ have this assumption built in. In our empirical analysis, we report standard errors and test statistics based on this assumption of correctly specified conditional moments, but we also report standard errors and tests statistics that are robust to misspecification of conditional moments. To compute the robust statistics we work with the realized values of $\partial h_{t+1}(\theta_0) / \partial \theta$ and $h_{t+1}(\theta_0) h_{t+1}(\theta_0)'$ in (48) and (49), without replacing them with estimates of their conditional expectations.

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

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

	const.	Δc_{t+1}	$\tau(I)$
Uncond.	2.95	-365.35	9.30
	(0.74)	(135.26)	[0.03]
Fixed IV	1.00	-0.11	215.12
	(0.00)	(0.15)	[0.00]
Opt. IV – NP	0.99	0.51	76.99
	(0.00)	(0.27)	[0.00]
	(*0.00*)	(*0.38*)	[*0.00*]
Opt. IV – SNP	1.00	0.12	113.41
	(0.00)	(0.19)	[0.00]
	(*0.00*)	(*0.12*)	[*0.00*]

Notes: Standard errors are in parentheses, p-values are in brackets. Standard errors and p-values are robust to misspecification of conditional moments, except those shown in italics, which assume correctly specified conditional moments. Conditional moments for uncond., fixed IV, and opt. IV-NP are estimated non-parametrically; for opt. IV-SNP they are based on the semi-nonparametric model.

the models' restrictions on the conditional distributions of returns, 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 of the relative risk aversion coefficient that is much closer to zero. Again, consistent with this prior literature, the test statistics constructed with all three estimators suggest that CRRA preferences fail to describe the real returns on common stocks and Treasury bills.

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 for the case of the unconditional estimator than

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

	const.	cay_t	Δc_{t+1}	$cay_t \times \Delta c_{t+1}$	$\tau(I)$	$\tau(B^{Wald})$	$\tau(B^{LM})$
Uncond.	-3.24	-40.83	626.99	-70564.09	0.09	0.59	7.90
	(8.84)	(206.91)	(1437.79)	(99269.77)	[0.77]	[0.74]	[0.02]
Fixed IV	1.00	-0.64	-0.47	105.42	143.91	21.37	51.05
	(0.00)	(0.16)	(0.30)	(35.02)	[0.00]	[0.00]	[0.00]
Opt. IV – NP	0.99	0.03	0.45	-13.54	63.50	1.93	1.59
	(0.03)	(0.80)	(4.44)	(110.77)	[0.00]	[0.38]	[0.45]
	(0.03)	*(0.78)*	*(4.91)*	*(94.18)*	*[0.00]*	*[0.56]*	*[0.42]*
Opt. IV – SNP	1.00	-0.06	-0.09	-2.81	89.29	5.19	4.65
	(0.00)	(0.06)	(0.27)	(9.13)	[0.00]	[0.07]	[0.10]
	(0.00)	*(0.04)*	*(0.14)*	*(7.21)*	*[0.00]*	*[0.00]*	*[0.00]*

Notes: 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 are in parentheses, p-values are in brackets. Standard errors and p-values are robust to misspecification of conditional moments, except those shown in italics, which assume correctly specified conditional moments. Conditional moments for uncond., fixed IV, and opt. IV-NP are estimated non-parametrically; for opt. IV-SNP they are based on the semi-nonparametric model.

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.

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 $z_t = cay$, the $\tau_T(B^{Wald})$ and $\tau_T(B^{LM})$ statistics provide some evidence to reject the null that the extension of the basic model with $CRRA$ preferences does not help in pricing stocks and T-Bills for Fixed IV, and, less so, for optimal IV - SNP, but not for optimal IV - NP. There is more support for the rejection of the null hypothesis that $(\gamma_1, \gamma_2) = 0$ for both the model with $z_t = def$ and $z_t = yc$ in Tables IV and V, particularly in the case of the Optimal IV - NP estimator.

However this evidence that conditioning the SDF on def or yc helps 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 $z_t = cay$ 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. For these tests may point to non-rejection of the simpler null model. This is what we find for the Lettau and Ludvigson (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 $CRRA$ model helps in pricing the test assets.

Looking across the three models, the point estimates of the parameters based on the optimal IV - NP and optimal IV - SNP 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 same is largely true of the estimated standard errors. The optimal GMM estimators, particularly those based on the SNP method, often produce

Table IV: Pricing kernel estimates with moments conditioned on def

	const.	def_t	Δc_{t+1}	$def_t \times \Delta c_{t+1}$	$\tau(I)$	$\tau(B^{Wald})$	$\tau(B^{LM})$
Uncond.	4.50	-274.15	-71.89	-11214.69	6.49	2.62	1.70
	(3.06)	(343.00)	(381.84)	(39098.00)	[0.01]	[0.27]	[0.43]
Fixed IV	1.05	-5.33	-9.80	945.10	124.17	2.51	38.79
	(0.04)	(4.05)	(7.25)	(671.89)	[0.00]	[0.29]	[0.00]
Opt. IV – NP	1.01	-0.93	-1.72	71.87	51.40	18.37	11.74
	(0.00)	(0.31)	(0.72)	(36.15)	[0.00]	[0.00]	[0.00]
	(0.01)	*(0.39)*	*(1.06)*	*(59.48)*	*[0.00]*	*[0.00]*	*[0.00]*
Opt. IV – SNP	1.01	-1.00	-1.30	117.04	52.16	10.33	9.68
	(0.00)	(0.38)	(0.58)	(59.16)	[0.00]	[0.01]	[0.01]
	(0.00)	*(0.22)*	*(0.40)*	*(38.10)*	*[0.00]*	*[0.00]*	*[0.00]*

Notes: 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 are in parentheses, p-values are in brackets. Standard errors and p-values are robust to misspecification of conditional moments, except those shown in italics, which assume correctly specified conditional moments. Conditional moments for uncond., fixed IV, and opt. IV-NP are estimated non-parametrically; for opt. IV-SNP they are based on the semi-nonparametric model.

considerably smaller standard errors than the fixed IV estimators, despite the fact that the latter incorporates conditioning information through the use of the instruments w_t. 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.

Comparing the optimal GMM estimators based on the nonparametric and semi-nonparametric 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. In addition, it is apparent that the SNP method often produces lower standard errors than the NP method. This could be an indication that the 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

Table V: Pricing kernel estimates with moments conditioned on yc

	const.	yc_t	Δc_{t+1}	$yc_t \times \Delta c_{t+1}$	$\tau(I)$	$\tau(B^{Wald})$	$\tau(B^{LM})$
Uncond.	-5.70	9.33	-140.41	-214.90	9.63	0.13	0.14
	(32.49)	(35.51)	(4454.77)	(4922.26)	[0.00]	[0.93]	[0.93]
Fixed IV	0.79	0.24	34.16	-38.31	128.69	7.43	44.72
	(0.09)	(0.09)	(15.23)	(16.62)	[0.00]	[0.02]	[0.00]
Opt. IV – NP	0.71	0.32	54.94	-60.65	56.99	7.82	17.22
	(0.11)	(0.13)	(20.13)	(22.36)	[0.00]	[0.02]	[0.00]
	(0.16)	*(0.18)*	*(29.37)*	*(32.48)*	*[0.00]*	*[0.02]*	*[0.00]*
Opt. IV – SNP	0.99	0.01	-1.36	1.52	94.29	2.00	2.03
	(0.05)	(0.06)	(8.59)	(9.45)	[0.00]	[0.37]	[0.36]
	(0.02)	*(0.02)*	*(3.78)*	*(4.13)*	*[0.00]*	*[0.12]*	*[0.12]*

Notes: 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 are in parentheses, p-values are in brackets. Standard errors and p-values are robust to misspecification of conditional moments, except those shown in italics, which assume correctly specified conditional moments. Conditional moments for uncond., fixed IV, and opt. IV-NP are estimated non-parametrically; for opt. IV-SNP they are based on the semi-nonparametric model.

test statistics and those that assume correctly specified conditional moments is, in most cases, quite small, particularly relative to the differences that are evident across the unconditional, fixed IV, and optimal IV estimators. This suggests that our methods of empirically approximating the conditional moments work reasonably well.

V.A Conditional Pricing Errors

The main motivation for moving from simple constant-weight pricing kernels to models where these weights are time-varying 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 (2008) 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 to inspect the conditional pricing errors implied by the estimated pricing kernels.

Figure 3 presents our nonparametric estimates of the conditional pricing errors of the five "primitive" assets for each one of the unconditional, fixed IV, and optimal IV - NP estimators. 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. These conditional pricing errors are much larger in magnitude than those that one would get by naively setting the pricing kernel to a constant, say 0.99. 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 0.15 for the high-low B/M portfolio, and from -1.5 to 1.5 for the T-Bill.

Employing conditional moment restrictions should help alleviate this mismatch between model-implied and actual variation in conditional moments. And indeed, the

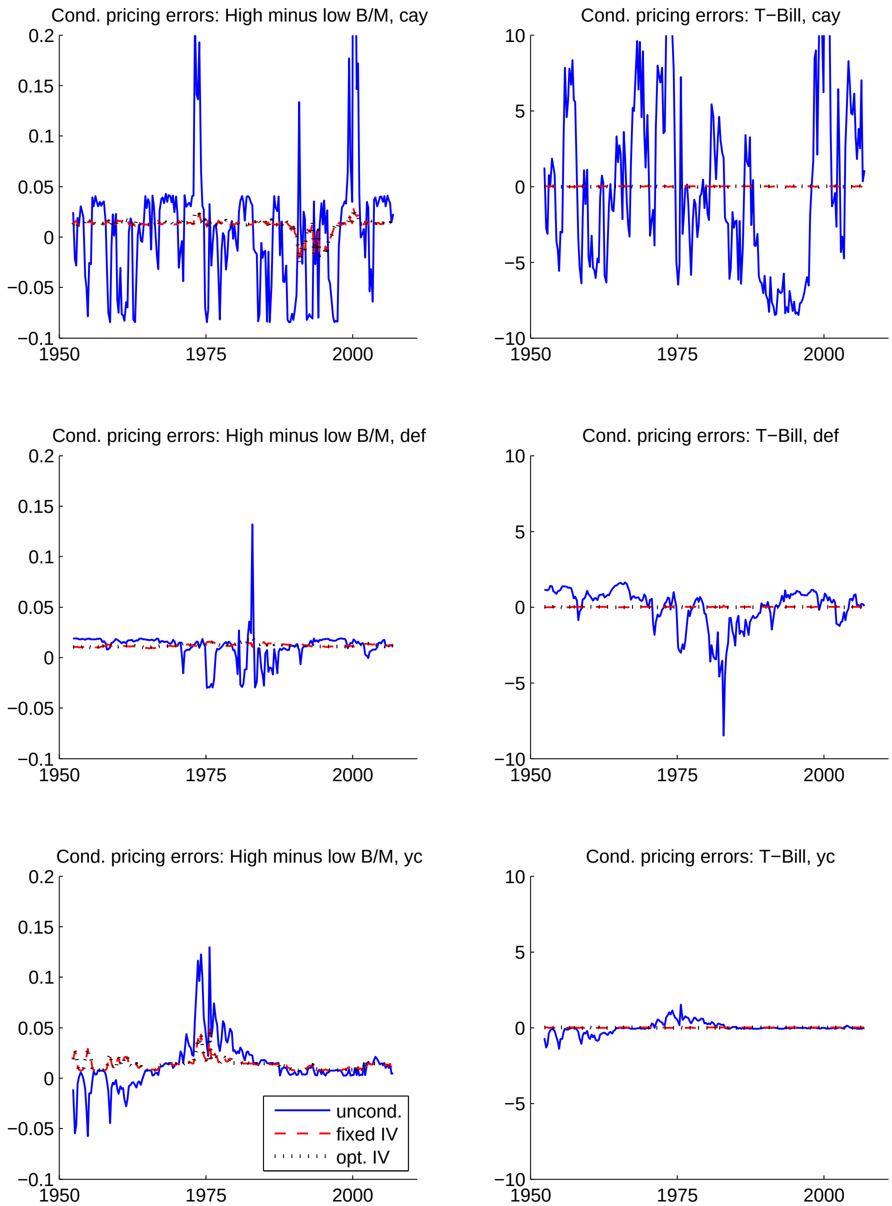

Figure 3: Conditional pricing errors of pricing kernel with time-varying weights: Stock portfolios (left) and T-Bill (right) with nonparametric estimates of moments conditioned on cay (top row), def (middle row), and yc (bottom row)

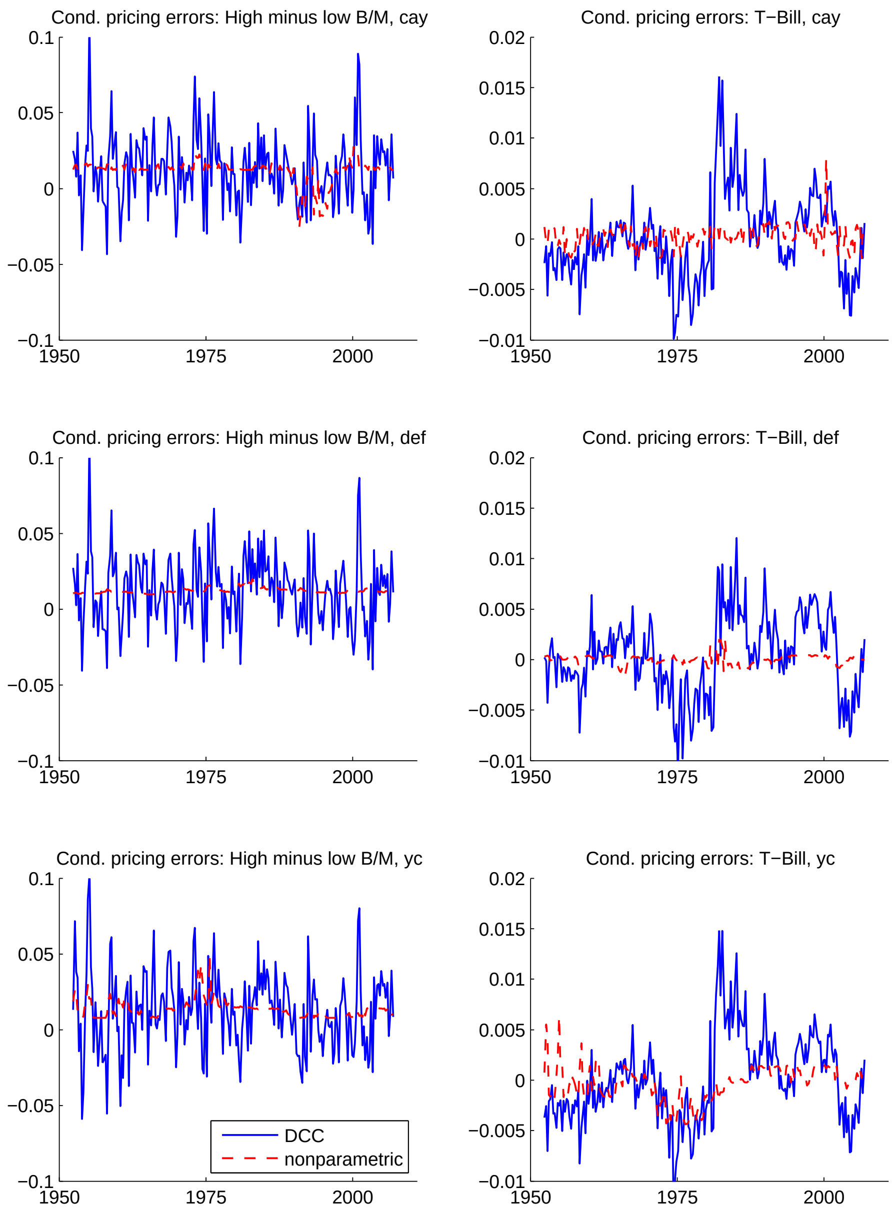

Figure 4: Conditional pricing errors of pricing kernel with time-varying weights: Stock portfolios (left) and T-Bill (right) with semi-nonparametric estimates of moments conditioned on cay (top row), def (middle row), and yc (bottom row)

fixed IV and optimal IV estimates produce conditional pricing errors that are more than one order of magnitude smaller than those based on unconditional estimates for the stock portfolios, and several orders of magnitude smaller for the T-Bill. The IV estimators force the model to pay attention to conditional moments in estimation, and so enforce consistency between the model implied conditional moments and those in the data.

Figure 4 shows the corresponding optimal IV estimates of conditional pricing errors for the semi-nonparametric model. For the sake of comparision, we also include the non-parametric optimal IV estimates from Figure 3. It is important to note, however, that the scale of the axes is completely different from the scale in Figure 3, as the optimal IV conditional pricing errors are of much smaller magnitude than obtained from the unconditional estimator. Both optimal IV methods produce conditional pricing errors that are positively correlated with each other, but the ones from the semi-nonparametric method exhibit more higher-frequency variation. But any differences that exist between the two methods are small relative to the differences that exist between the estimates based on unconditional moment restrictions and the optimal IV ones.

The message from Figures 3 and 4 is also underscored by Table VI, which summarizes the time-series standard deviation of conditional pricing errors, and the cross-sectional standard deviation of unconditional pricing errors. As Panel A shows, the unconditional estimates with $z_t = cay$ imply an enormous standard deviation of conditional pricing errors, particularly for the T-Bill. When conditioning information is introduced in estimation, variation in the conditional pricing errors shrinks, while the cross-sectional standard deviation of unconditional pricing errors increases. Evidently, at the unconditional moment restriction estimates the model achieves a relatively good

Table VI: Pricing error variance

Panel A: Δc_{t+1} scaled by cay_t, moments conditioned on cay_t	Time-series S.D. of conditional pricing errors					Cross-sectional S.D of uncond. pricing errors
	SmGrw	SmVal	BigGrw	BigVal	TBill	
Uncond.	0.20	0.21	0.17	0.18	5.54	0.02
Fixed IV	0.03	0.04	0.03	0.03	0.00	0.05
Opt. IV – NP	0.03	0.04	0.03	0.03	0.00	0.05
Opt. IV – SNP	0.04	0.04	0.03	0.04	0.00	0.05
Panel B: Δc_{t+1} scaled by def_t, moments conditioned on def_t						
	Time-series S.D. of conditional pricing errors					Cross-sectional S.D of uncond. pricing errors
	SmGrw	SmVal	BigGrw	BigVal	TBill	
Uncond.	0.12	0.11	0.08	0.07	1.38	0.02
Fixed IV	0.02	0.04	0.02	0.03	0.01	0.05
Opt. IV – NP	0.02	0.04	0.02	0.03	0.00	0.05
Opt. IV – SNP	0.03	0.04	0.03	0.03	0.00	0.05
Panel C: Δc_{t+1} scaled by yc_t, moments conditioned on yc_t						
	Time-series S.D. of conditional pricing errors					Cross-sectional S.D of uncond. pricing errors
	SmGrw	SmVal	BigGrw	BigVal	TBill	
Uncond.	0.03	0.02	0.04	0.02	0.38	0.02
Fixed IV	0.02	0.04	0.03	0.03	0.00	0.05
Opt. IV – NP	0.02	0.04	0.03	0.03	0.00	0.05
Opt. IV – SNP	0.03	0.04	0.03	0.04	0.00	0.05

Notes: 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 uncond., fixed IV, and opt. IV-NP are estimated non-parametrically; for opt. IV-SNP they are based on the semi-nonparametric model.

fit in the cross-section, 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 $z_t = def$ and $z_t = yc$.

Given that the motivation for models with time-varying pricing kernel weights is to match conditional moments of returns and factors, this inability to even approximately price the test assets conditionally is an important failure of the model. 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.

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

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

Misspecification of $E\left[m_{t+1}|z_t\right]$ has a much bigger effect on $E\left[h_{t+1}|z_t\right]$ 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) goes a long ways towards fixing the conditional mean of the SDF.

V.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 ϕ_t^f. Figure 5 plots the estimates of ϕ_t^f with z_t equal to cay, def, or yc.

The coefficient ϕ_t^f has a close connection to the coefficient of relative risk aversion. Consider a constant-relative risk aversion pricing kernel, $m_{t+1} = \delta_t \exp\left(-\gamma_t \Delta c_{t+1}\right)$, with time-varying relative risk aversion γ_t and time-discount factor δ_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 δ_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, ϕ_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 ϕ_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 ϕ_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 $\phi_t^f \in [-100, 0]$.

Focusing first on the estimates based on unconditional moment restrictions (the top graph in Figure 5), the estimates of ϕ_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 ϕ_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 ϕ_t^f for $z_t = cay$ by several orders of magnitude, but the estimated ϕ_t^f are still

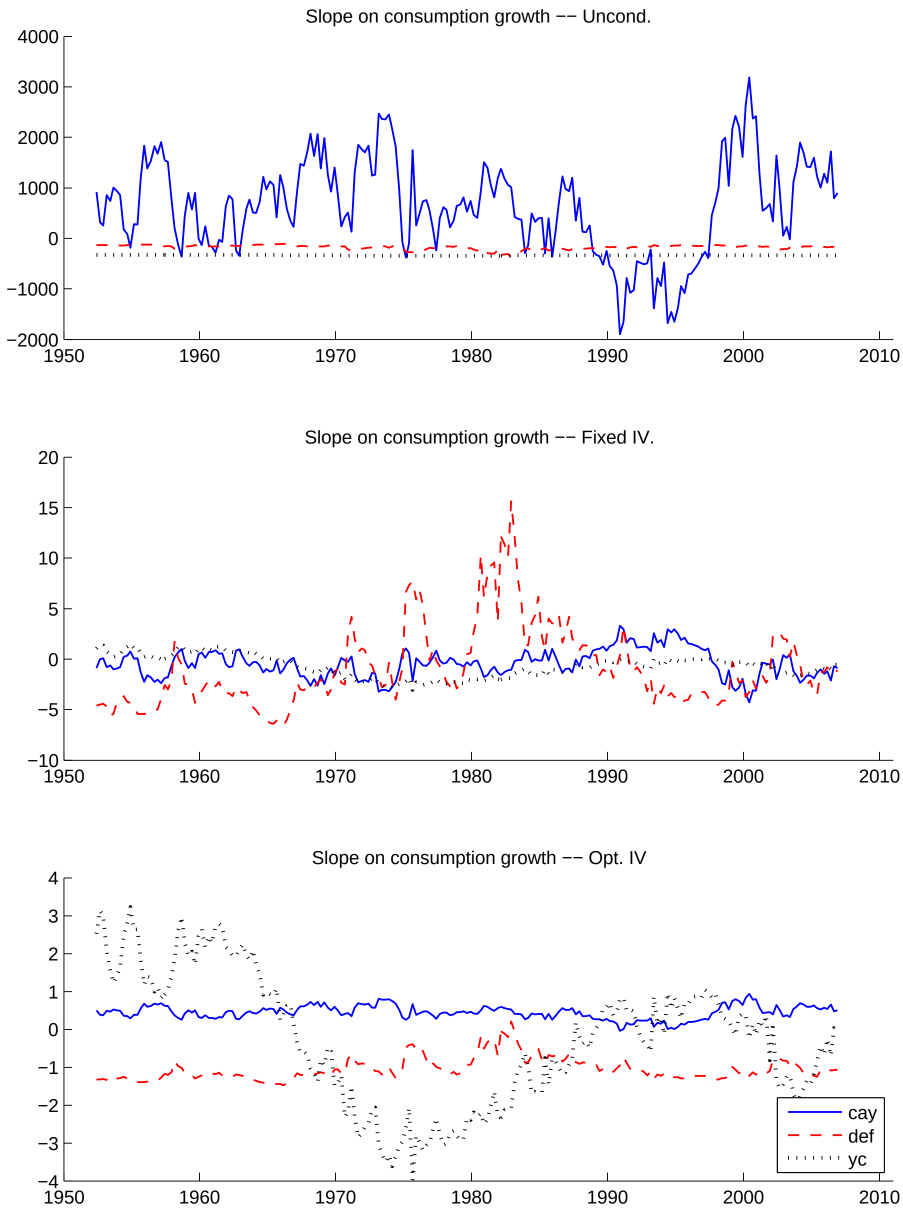

Figure 5: Time-series of estimated SDF weights from with unconditional (top row), fixed IV (middle row), and optimal IV estimators (bottom row)

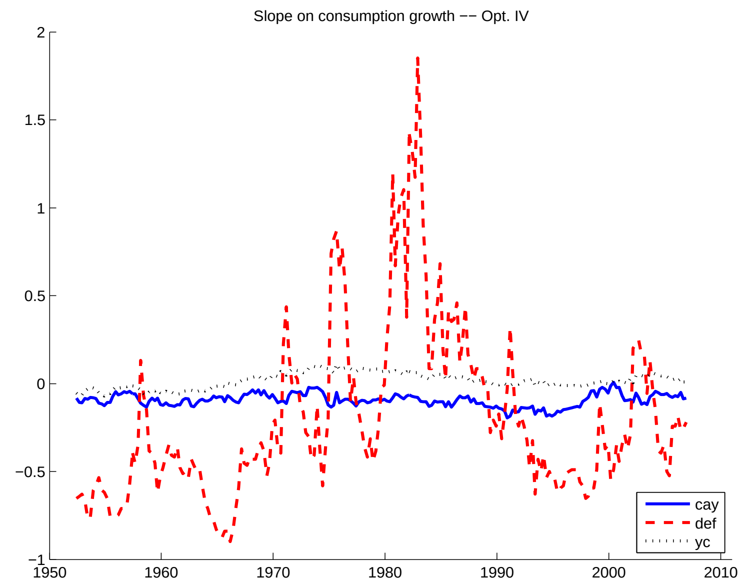

Figure 6: Time-series of optimal IV estimates of SDF weight with conditional moments estimated from semi-nonparametric model

often positive. The corresponding estimates for the model with $z_t = yc$ are also much closer to zero, but are now also sometimes positive. The most volatile ϕ_t^f is obtained with $z_t = def$. 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 (Table IV).

Finally, using the Optimal IV-NP estimator, the estimated ϕ_t^f are now very close to zero for all three choices of z_t. A similar result is shown by Figure 6, which compares the SDF weight on consumption growth implied by the semi-nonparametric optimal IV estimates with the nonparametric ones. In terms of economic magnitudes, the differences between the two methods are small. With both methods, the estimated ϕ_t^f are close to zero. In addition, the SDF $m_{t+1} = \phi_t^0 + \phi_t^f \Delta c_{t+1}$ implied by the Optimal IV estimates (not shown) is always positive, ranging between 0.95 and 1.05, while the SDF implied by the estimates from unconditional moment restrictions frequently take 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 model 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 hide enormous time-variation in *conditional* pricing errors.

Appendices

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

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

$$\sqrt{T}\left(\theta_T^A - \theta_0\right) = -\left[\frac{1}{T}\sum_t A_t \frac{\partial h_{t+1}(\theta_T^{Am})}{\partial \theta}\right]^{-1} \frac{1}{\sqrt{T}}\sum_t A_t h_{t+1}(\theta_0), \tag{50}$$

where θ_T^{Am} is a collection of vectors, one for each coordinate of $A_t h_{t+1}$, that lie between θ_T^A and θ_0, almost surely. Similarly, a mean-value expansion of the sample mean of $B_t h_{t+1}(\theta_T^A)$ gives

$$\frac{1}{\sqrt{T}}\sum_t B_t h_{t+1}(\theta_T^A) = \frac{1}{\sqrt{T}}\sum_t B_t h_{t+1}(\theta_0) + \frac{1}{T}\sum_t B_t \frac{\partial h_{t+1}(\theta_T^{Bm})}{\partial \theta} \times \sqrt{T}\left(\theta_T^A - \theta_0\right), \tag{51}$$

with θ_T^{Bm} interpreted similarly. Substitution of (50) into (51) leads to

$$\frac{1}{\sqrt{T}}\sum_t B_t h_{t+1}(\theta_T^A) = \frac{1}{\sqrt{T}}\sum_t C_t^A h_{t+1}(\theta_0) + o_p\left(1\right), \tag{52}$$

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

$$\sqrt{T}\left(\theta_T^* - \theta_0\right) \overset{a}{=} -\left(E\left[\Psi_t^{\theta\prime}\Sigma_t^{\mathcal{G}-1}\Psi_t^{\theta}\right]\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Psi_t^{\theta\prime}\Sigma_t^{\mathcal{G}-1}h_{t+1}(\theta_0). \tag{53}$$

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 Ω_0^*, we obtain

$$\sqrt{T}(\gamma_T^* - \gamma_0) \overset{a}{=} -\Omega_{\gamma\gamma}^*\frac{1}{\sqrt{T}}\sum_{1}^{T}\mathcal{H}_t^{\mathcal{G}\prime}\Sigma_t^{\mathcal{G}-1}h_{t+1}(\theta_0). \tag{54}$$

The random vector $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathcal{H}_t^{\mathcal{G}\prime}\Sigma_t^{\mathcal{G}-1}h_{t+1}(\theta_0)$ converges in distribution to a normal random vector with mean zero and covariance matrix

$$\left(\Omega_{\gamma\gamma}^*\right)^{-1} = \mathcal{K}^{\gamma\gamma} - \mathcal{K}^{\gamma\beta}\left(\mathcal{K}^{\beta\beta}\right)^{-1}\mathcal{K}^{\beta\gamma}, \tag{55}$$

where the last equality follows from the partitioned matrix inversion formula applied to Ω_0^*. Therefore, the asymptotic distribution of $\varsigma_T^W(A^*)$ 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 θ_0 for the unconstrained $m^{\mathcal{G}}_{t+1}$ are

$$E\left[\begin{pmatrix} \Psi_t^{\beta\prime} \\ \Psi_t^{\gamma\prime} \end{pmatrix} \Sigma_t^{-1} h_{t+1}(\theta_0, \gamma_0)\right] = 0. \tag{56}$$

Under the constraint that $\gamma_0 = 0$, (56) 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 (56) for joint estimation of β_0 and γ_0, subject to the constraint that $\gamma_T = 0$ (see Eichenbaum, Hansen, and Singleton (1988)). Letting $h^{\mathcal{N}}_{t+1}(\beta) = h_{t+1}(\beta, 0)$, the pricing errors under the constraint that $\gamma = 0$, the optimal distance matrix in this quadratic form is a consistent estimator of

$$W_0 = E\left(\begin{pmatrix} \Psi_t^{\beta\prime} \Sigma_t^{\mathcal{N}-1} h^{\mathcal{N}}_{t+1} \\ \Psi_t^{\gamma\prime} \Sigma_t^{\mathcal{N}-1} h^{\mathcal{N}}_{t+1} \end{pmatrix} \begin{pmatrix} h^{\mathcal{N}\prime}_{t+1} \Sigma_t^{\mathcal{N}-1} \Psi_t^{\beta}, \; h^{\mathcal{N}\prime}_{t+1} \Sigma^{\mathcal{N}-1} \Psi_t^{\gamma} \end{pmatrix}\right).$$

The first-order conditions to this minimization problem are

$$\left(\frac{1}{T}\sum_t \mathcal{P}_{t+1}\right) W_T^{-1} \frac{1}{T}\sum_t \begin{pmatrix} \Psi_t^{\beta\prime} \\ \Psi_t^{\gamma\prime} \end{pmatrix} \Sigma_t^{\mathcal{N}-1} h_{t+1}(\theta_T, 0) = \begin{pmatrix} 0 \\ \lambda_T \end{pmatrix}, \tag{57}$$

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

$$\mathcal{P}_{t+1} = \begin{bmatrix} \frac{\partial h_{t+1}(\beta_T,0)'}{\partial\beta} \Sigma_t^{\mathcal{N}-1} \Psi_t^{\beta} & \frac{\partial h_{t+1}(\beta_T,0)'}{\partial\beta} \Sigma_t^{\mathcal{N}-1} \Psi_t^{\gamma} \\ \frac{\partial h_{t+1}(\beta_T,0)'}{\partial\gamma} \Sigma_t^{\mathcal{N}-1} \Psi_t^{\beta} & \frac{\partial h_{t+1}(\beta_T,0)'}{\partial\gamma} \Sigma_t^{\mathcal{N}-1} \Psi_t^{\gamma} \end{bmatrix}. \tag{58}$$

The first $K-G$ rows of the lead matrix $T^{-1}\sum_t \mathcal{P}_{t+1}$ in (57) are the same as the first $K-G$ rows of W_T. Therefore, the first $K-G$ first-order conditions in (57) are

$$\frac{1}{T}\sum_t \Psi_t^{\beta\prime}\Sigma_t^{\mathcal{N}-1}h_{t+1}^{\mathcal{N}}(\beta_T^{\mathcal{N}})=0. \tag{59}$$

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 $m_{t+1}^{\mathcal{N}}$.[22] We let $\beta_T^{\mathcal{N}}$ denote this optimal GMM estimator obtained when the SDF is taken to be $m_{t+1}^{\mathcal{N}}(\beta_0)$.

The Lagrange multiplier is obtained by solving the first-order conditions (57) for λ_T. Partitioning the weighting matrix W_0 conformably with the $K-G$ and G blocks of moment conditions in (56), letting W_0^{ij} denote the ij^{th} block of W_0^{-1}, and

$$F_0^{LM}=E\left[\Psi_t^{\gamma\prime}\Sigma^{-1}\Psi_t^{\theta}\right]W_0^{12}+E\left[\Psi_t^{\gamma\prime}\Sigma_t^{-1}B_t^{\prime}\right]W_0^{22}, \tag{60}$$

λ_T can be expressed as

$$\lambda_T=F_T^{LM}\frac{1}{T}\sum_t \Psi_t^{\gamma\prime}\Sigma_t^{-1}h_{t+1}^{\mathcal{N}}(\beta_T^{\mathcal{N}}), \tag{61}$$

where F_T^{LM} is a consistent estimator of F_0^{LM}. Using the formula for the partitioned inverse of the matrix W_0 it can be verified that $F_0^{LM}=I$ and, therefore, this expression for λ_T simplifies to (34).

[22] This derivation addresses an important question that was left implicit up to this point. In previous sections we first constructed the optimal GMM estimator θ_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 θ_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 θ_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.

D An Alternative Representation of the Wald Statistic for Completely Affine SDFs

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

We have $\imath_R - h_{t+1}^{\mathcal{N}}\left(\beta_T^{\mathcal{N}}\right) = r_{t+1}f_{t+1}^{\#\mathcal{N}\prime}\beta_T^{\mathcal{N}}$ and so

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\left[\widehat{\mathcal{H}}_t^{\mathcal{G}}\hat{\Sigma}_t^{\mathcal{G}-1}\left\{\imath_R - h_{t+1}^{\mathcal{N}}\left(\beta_T^{\mathcal{N}}\right)\right\}\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\left[\left(\hat{\Psi}_t^{\gamma\prime} - \hat{\mathcal{K}}_T^{\gamma\beta}\left(\hat{\mathcal{K}}_T^{\beta\beta}\right)^{-1}\hat{\Psi}_t^{\beta\prime}\right)\hat{\Sigma}_t^{\mathcal{G}-1}r_{t+1}f_{t+1}^{\#\mathcal{N}\prime}\beta_T^{\mathcal{N}}\right] \\
&= \frac{1}{T}\sum_{t=1}^{T}\left[\hat{\Psi}_t^{\gamma\prime}\hat{\Sigma}_t^{\mathcal{G}-1}r_{t+1}f_{t+1}^{\#\mathcal{N}\prime}\beta_T^{\mathcal{N}} - \hat{\mathcal{K}}_T^{\gamma\beta}\left(\hat{\mathcal{K}}_T^{\beta\beta}\right)^{-1}\hat{\Psi}_t^{\beta\prime}\hat{\Sigma}_t^{\mathcal{G}-1}r_{t+1}f_{t+1}^{\#\mathcal{N}\prime}\beta_T^{\mathcal{N}}\right] \\
&= \hat{\mathcal{K}}_T^{\gamma\beta}\beta_T^{\mathcal{N}} - \hat{\mathcal{K}}_T^{\gamma\beta}\left(\hat{\mathcal{K}}_T^{\beta\beta}\right)^{-1}\left(\hat{\mathcal{K}}_T^{\beta\beta}\right)\beta_T^{\mathcal{N}} = 0,
\end{aligned}
$$

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

References

Ait-Sahalia, Y. (1996). Testing Continuous-Time Models of the Spot Interest Rate. *Review of Financial Studies 9*, 385–426.

Altonji, J. and L. Segal (1996). Small Sample Bias in GMM Estimation of Covariance Structures. *Journal of Business and Economic Statistics 14*, 353–366.

Campbell, J. and J. Cochrane (1999). By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. *Journal of Political Economy 107*, 205–251.

Carhart, M. (1997). On Persistence of Mutual Fund Performance. *Journal of Finance 52*, 57–82.

Chamberlain, G. (1987). Asymptotic Efficiency in Estimation with Conditional Moment Restrictions. *Journal of Econometrics 34*, 305–344.

Cochrane, J. (1996). A Cross-Sectional Test of an Investment-Based Asset Pricing Model. *Journal of Political Economy 104*, 572–621.

Daniel, K. and S. Titman (2006). Testing Factor Model Explanations of Market Anamolies. Working Paper, Northwestern University.

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.

Engle, R. (2002). Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models. *Journal of Business and Economic Statistics 17*, 339–350.

Fama, E. and K. French (1992). The Cross-Section of Expected Stock Returns. *Journal of Finance 47*, 427–465.

Fama, E. and K. French (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics 33*, 23–49.

Fama, E. and K. French (1996). Mulitifactor Explanations of Asset Pricing Anomalies. *Journal of Finance 51*, 55–87.

Fan, J. (1992). Design-adaptive Nonparametric Regression. *Journal of the American Statistical Association 87*(420), 998–1004.

Grossman, S. and R. Shiller (1981). The Determinants of the Variability of Stock Market Prices. *American Economic Review 71*, 222–227.

Hansen, L. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica 50*, 1029–1054.

Hansen, L. (1985). A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators. *Journal of Econometrics 30*, 203–238.

Hansen, L., J. Heaton, and M. Ogaki (1988). Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions. *Journal of the American Statistical Association 83*, 863–871.

Hansen, L., J. Heaton, and A. Yaron (1996). Finite-Sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics 14*, 262–280.

Hansen, L. and S. Richard (1987). The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models. *Econometrica 55*, 587–613.

Hansen, L. and K. Singleton (1982). Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models. *Econometrica 50*, 1269–1286.

Hansen, L. and K. Singleton (1996). Efficient Estimation of Linear Asset Pricing Models with Moving Average Errors. *Journal of Business & and Economic Statistics 14*, 53–68.

Hart, J. D. (1994). Automated Kernel Smoothing of Dependent Data by Using Time Series Cross-Validation. *Journal of the Royal Statistical Society, Series B 56*(3), 529–542.

Hodrick, R. and X. Zhang (2001). Evaluating the Specification Errors of Asset Pricing Models. *Journal of Financial Economics 62*, 327–376.

Imbens, G. and R. Spady (2005). the Performance of Empirical Likelihood and its Generalizations. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge University Press: New York.

Jagannathan, R. and Wang (1996). The Conditional CAPM and the Cross-section of Expected Returns. *Journal of Finance 51*, 3–54.

Lettau, M. and S. Ludvigson (2001a). Consumption, Aggregate Wealth, and Expected Stock Returns. *Journal of Finance 56*, 815–849.

Lettau, M. and S. Ludvigson (2001b). Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying. *Journal of Political Economy 109*, 1238–1287.

Lewellen, J. and S. Nagel (2006). The Conditional CAPM Does Not Explain Asset Pricing Anomalies. *Journal of Financial Economics 79*.

Lewellen, J., S. Nagel, and J. Shanken (2008). A Skeptical Appraisal of Asset-Pricing Tests. *forthcoming Journal of Financial Economics*.

Lustig, H. and S. V. Nieuwerburgh (2006). Housing Collateral, Consumption Insurance, and Risk Premia: An Empirical Perspective. *Journal of Finance 60*, 1167–1219.

Newey, W. and R. Smith (2004). Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica 72*, 219–255.

Newey, W. and K. D. West (1987). Hypothesis Testing with Efficient Method of Moment Estimation. *International Economic Review 28*, 777–787.

Pastor, L. and R. Stambaugh (2003). Liquidity Risk and Expected Stock Returns. *Journal of Political Economy 111*, 642–685.

Roussanov, N. (2009). Composition of Wealth, Conditioning Information, and the Cross-Section of Stock Returns. Working Paper.

Santos, T. and P. Veronesi (2006). Labor Income and Predictable Stock Returns. *Review of Financial Studies 19*, 1–44.

Yao, Q. and H. Tong (1998). Cross-validatory bandwidth selections for regression estimation based on dependent data. *Journal of Statistical Planning and Inference 68*(2), 387–415.