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

Stefan Nagel†
Stanford University and NBER

Kenneth J. Singleton‡
Stanford University and NBER

This Internet Appendix presents derivations and proofs for the results used in the main article (NS). We also report findings from simulations of the small sample properties of the estimators used in NS.

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

†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/

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

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

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

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

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

with θ_T^{Bm} interpreted similarly. Substitution of (IA.1) into (IA.2) leads to

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

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

B. Intermediate Steps in Section III

To express the Wald statistic $\varsigma_T^W(A^*)$ 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

$$\sqrt{T}\left(\theta_T^* - \theta_0\right) \overset{a}{=} -\left(E\left[\Psi_t^{\theta\prime}\Sigma_t^{\mathcal{G}-1}\Psi_t^{\theta}\right]\right)^{-1} \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \Psi_t^{\theta\prime}\Sigma_t^{\mathcal{G}-1} h_{t+1}(\theta_0). \qquad \text{(IA.4)}$$

Noting that $\sqrt{T}(\gamma_T^* - \gamma_0) = [0, I_G]\sqrt{T}(\theta_T^* - \theta_0)$, and using the partitioned matrix formula for inverting Ω_0^*, we obtain

$$\sqrt{T}(\gamma_T^* - \gamma_0) \stackrel{a}{=} -\,\Omega_{\gamma\gamma}^* \, \frac{1}{\sqrt{T}} \sum_1^T \mathcal{H}_t^{\mathcal{G}\prime} \Sigma_t^{\mathcal{G}-1} h_{t+1}(\theta_0). \tag{IA.5}$$

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

$$\left(\Omega_{\gamma\gamma}^*\right)^{-1} = \mathcal{K}^{\gamma\gamma} - \mathcal{K}^{\gamma\beta} \left(\mathcal{K}^{\beta\beta}\right)^{-1} \mathcal{K}^{\beta\gamma}, \tag{IA.6}$$

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 $\gamma_0 = 0$. More precisely, the moment conditions associated with the optimal GMM estimator of θ_0 for the unconstrained $m_{t+1}^{\mathcal{G}}$ are

$$E\left[\begin{pmatrix} \Psi_t^{\beta\prime} \\ \Psi_t^{\gamma\prime} \end{pmatrix} \Sigma_t^{-1} h_{t+1}(\beta_0, \gamma_0)\right] = 0. \tag{IA.7}$$

Under the constraint that $\gamma_0 = 0$, (IA.7) gives more moment equations (K) than unknown parameters ($K - G = \dim\beta_0$). Therefore, the LM statistic for testing $H_0 : \gamma_0 = 0$ is obtained by minimizing a quadratic form in the sample version of the moments (IA.7) for joint estimation of β_0 and γ_0, subject to the constraint that $\gamma_T = 0$ (see Eichenbaum, Hansen, and Singleton (1988)). Letting $h_{t+1}^{\mathcal{N}}(\beta) = h_{t+1}(\beta, 0)$, the pricing errors under the constraint that $\gamma = 0$, the optimal distance matrix in this quadratic form is a consistent estimator of

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

The first-order conditions to this minimization problem are

$$\left(\frac{1}{T}\sum_t \mathcal{P}_{t+1}\right) W_T^{-1}\frac{1}{T}\sum_t \left(\begin{array}{c}\Psi_t^{\beta\prime} \\ \Psi_t^{\gamma\prime}\end{array}\right)\Sigma_t^{\mathcal{N}-1}h_{t+1}^{\mathcal{N}}(\beta_T) = \left(\begin{array}{c}0 \\ \lambda_T\end{array}\right), \qquad \text{(IA.8)}$$

where λ_T is the $G\times 1$ vector of Lagrange multipliers associated with the constraint $\gamma_T = 0$, 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

$$\mathcal{P}_{t+1} = \left[\begin{array}{cc}\frac{\partial h_{t+1}^{\mathcal{N}}(\beta_T)'}{\partial\beta}\Sigma_t^{\mathcal{N}-1}\Psi_t^{\beta} & \frac{\partial h_{t+1}^{\mathcal{N}}(\beta_T)'}{\partial\beta}\Sigma_t^{\mathcal{N}-1}\Psi_t^{\gamma} \\ \frac{\partial h_{t+1}^{\mathcal{N}}(\beta_T)'}{\partial\gamma}\Sigma_t^{\mathcal{N}-1}\Psi_t^{\beta} & \frac{\partial h_{t+1}^{\mathcal{N}}(\beta_T)'}{\partial\gamma}\Sigma_t^{\mathcal{N}-1}\Psi_t^{\gamma}\end{array}\right]. \qquad \text{(IA.9)}$$

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

$$\frac{1}{T}\sum_t \Psi_t^{\beta\prime}\Sigma_t^{\mathcal{N}-1}h_{t+1}^{\mathcal{N}}(\beta_T^{\mathcal{N}}) = 0. \qquad \text{(IA.10)}$$

These are the sample first-order conditions for the optimal GMM estimator of the parameters of the SDF under the null hypothesis $\gamma_0 = 0$, that is, they are the first-order conditions when estimation proceeds with the constrained SDF $m_{t+1}^{\mathcal{N}}$.[1] We let $\beta_T^{\mathcal{N}}$

[1] 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 $E[B_t h_{t+1}(\theta_0)] = 0$ simul-

denote this optimal GMM estimator obtained when the SDF is taken to be $m^{\mathcal{N}}_{t+1}(\beta_0)$.

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

$$\lambda_T = \frac{1}{T}\sum_t \Psi_t^{\gamma\prime}\Sigma_t^{-1}h^{\mathcal{N}}_{t+1}(\beta^{\mathcal{N}}_T), \tag{IA.11}$$

as in (34).

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

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

We have $p_R - h^{\mathcal{N}}_{t+1}\left(\beta^{\mathcal{N}}_T\right) = r_{t+1}f^{\#\mathcal{N}\prime}_{t+1}\beta^{\mathcal{N}}_T$ and so

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

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

E. Robust Statistics

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

taneously 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.

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

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

F. Small-Sample Properties

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

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

We simulate returns of five assets and these returns are assumed to be consistent with a linearized pricing kernel of the type that we investigate empirically in NS:

$$m_{t+1}^{\mathcal{G}}(\theta_0) = (\beta_1 + \gamma_1 z_t) + (\beta_2 + \gamma_2 z_t)\,\Delta c_{t+1}. \tag{IA.12}$$

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

$$E\left[r_{t+1}|z_t\right] = \frac{p_t - (\beta_2 + \gamma_2 z_t)\mathrm{Cov}\left(f_{t+1}, r_{t+1}|z_t\right)}{\beta_1 + \gamma_1 z_t + (\beta_2 + \gamma_2 z_t)E\left[f_{t+1}|z_t\right]}. \tag{IA.13}$$

To generate artificial data on conditional expected returns consistent with this pricing

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

model, we need to model the dynamics of z_t. Given a process for z_t we then need to make sure the SDF parameters and the dynamics of $\text{Cov}\left(f_{t+1}, r_{t+1}|z_t\right)$ and $E\left[f_{t+1}|z_t\right]$ are consistent with $E\left[r_{t+1}|z_t\right]$ according to (IA.13).

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

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

Given the simulated z_t and f_{t+1}, we choose SDF parameters β_2 and γ_2 such that the term $\beta_2 + \gamma_2 z_t$ (which corresponds approximately to a time-varying relative risk aversion coefficient) has mean 200 and standard deviation 70. These parameter values allow us to roughly match the mean and standard deviation of conditional expected stock returns from the local linear conditional moment estimates in NS.

We further proceed to pick γ_1 so that the standard deviation of the conditional mean return of the conditionally risk-free asset, that is, $1/E[m^{\mathcal{G}}_{t+1}|z_t]$ matches the standard deviation of the conditional mean of the real T-bill return, where the latter is obtained from the local polynomial conditional moment estimates in NS. We choose β_1 so that the mean of $1/E[m^{\mathcal{G}}_{t+1}|z_t]$ matches the mean real T-bill return.

Given the $\text{Cov}\left(f_{t+1}, r_{t+1}|z_t\right)$ and $\text{Var}\left(f_{t+1}|z_t\right)$ as specified above, we simulate return innovations from a conditional one-factor factor model. The factor related component is

$\text{Cov}\,(f_{t+1}, r_{t+1}|z_t)\text{Var}\,(f_{t+1}|z_t)^{-1} f_{t+1}$. We then add an IID normal residual for each asset (uncorrelated between assets) to match the unconditional variance of the unexpected return of the four stock portfolios and the T-bill in the empirical data (i.e., the residuals from the local polynomial regression estimates in NS). This completes the specification of the joint dynamics of z_t, f_{t+1}, and r_{t+1}.

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

Ensuring a global optimum for the optimal IV estimator across all simulations can be a challenge. For example, the numerical nonlinear equation solver might run off towards a "solution" with extremely large SDF parameters that make $E[A_t h_{t+1}|z_t]$ close to zero not by making h_{t+1} small, but instead by blowing up $E[h_{t+1}h'_{t+1}|z_t]$ (which appears with an inverse in A_t) to huge values. A method that we found to work well is to first construct preliminary fixed IV estimates, using the first few principal components of $E\left[\partial h_{t+1}/\partial\theta|z_t\right]$ (taken from the local polynomial estimation) as instruments, and then using these preliminary estimates as initial values for the nonlinear equation solver, supplemented if necessary with an extensive grid search over initial values.

It also helps to impose a common bandwidth b_k in the estimation of the two conditional moments $g_k\left(z_t\right) = E[(r'_{k,t+1}, \Delta c_{t+1} r'_{k,t+1})|z_t]$ corresponding to asset k. In some of the Monte Carlo samples, the local polynomial regressions can produce quite extreme values for the estimates of $E[r'_{k,t+1}|z_t]$ or $E[\Delta c_{t+1} r'_{k,t+1}|z_t]$ for outlier observations of z_t, and this seems to be more of a problem if only one of the two elements of the

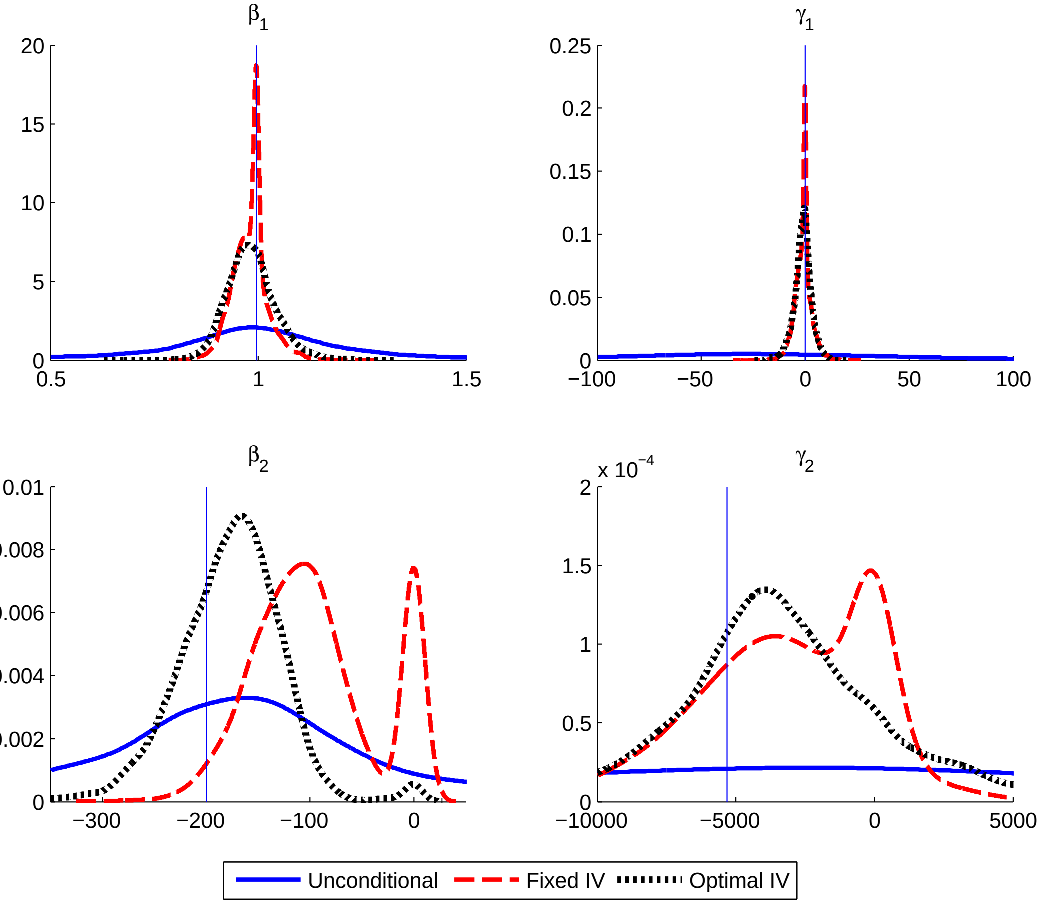

Figure IA.1. Kernel-smoothed Monte Carlo density of SDF parameter estimates. The vertical line indicates the true parameter value.

estimate of $g_k\left(z_t\right) = E[(r'_{k,t+1}, \Delta c_{t+1} r'_{k,t+1})'|z_t]$ is affected (because it is estimated with a narrow bandwidth), while the other is not (because it is estimated with a wide bandwidth). Imposing the same bandwidth ensures that $E[r'_{k,t+1}|z_t]$ and $E[\Delta c_{t+1} r'_{k,t+1}|z_t]$ are estimated from the same local neighborhood around z_t.

Figure IA.1 presents the Monte Carlo density of the parameter estimates. The estimates from fixed IV and optimal IV estimators are considerably more precise and better centered around the true parameter values than the estimates based on the unconditional estimator. For the β_1 and γ_1 estimates, the fixed IV estimates seem to

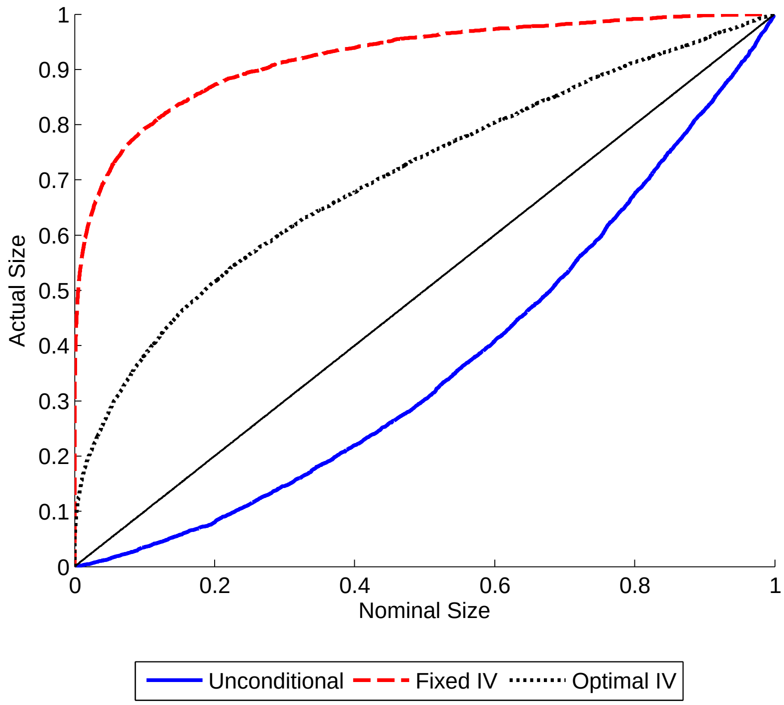

Figure IA.2. p-value plots for $\tau(I)$ test of zero average pricing errors

be slightly more precise, but for the β_2 and γ_2 estimates, the fixed IV estimates show considerably higher dispersion and also some bias. For β_2, the RMSE of the fixed IV estimates is about five times as big as with the optimal IV estimator. Overall, the optimal IV estimates look well behaved.

Figure IA.2 plots the empirical distribution of p-values from the $\tau(I)$ test to illustrate the actual size of the test in relation to its nominal size. The test based on the unconditional estimator underrejects compared with the nominal size of the test. The test based on the fixed IV estimator severely overrejects. Its actual size is much higher than the nominal size of the test, particularly for small nominal sizes. This is a consequence of the large number of instruments relative to the sample size (which is also often typical in empirical applications of the fixed IV estimator). If one reduced the number of instruments, the tendency to overreject would likely be reduced. In the extreme case of only a constant as the "instrument," the fixed IV estimator becomes the unconditional estimator. The $\tau(I)$ test based on the optimal IV estimator also exhibits a tendency to overreject, but considerably less so than the test based on the fixed IV estimator, a likely consequence of the fact that it does not use a large number of moment conditions in the construction of the test statistic. Nevertheless, an interpretation of empirical results based on the $\tau(I)$ test statistic should take into account this tendency to overreject.

Next, we investigate the size and power of the Wald and LM tests of $H_0 : \gamma_1 = 0, \gamma_2 = 0$. To generate data under this null hypothesis, we simulate from the SDF $m_{t+1}^{\mathcal{N}}$ with $\gamma_1 = 0$, $\gamma_2 = 0$, $\beta_2 = -200$, and β_1 chosen such that $1/E[m_{t+1}^{\mathcal{N}}|z_t]$ matches the mean gross return on Treasury bills in our sample.

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

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

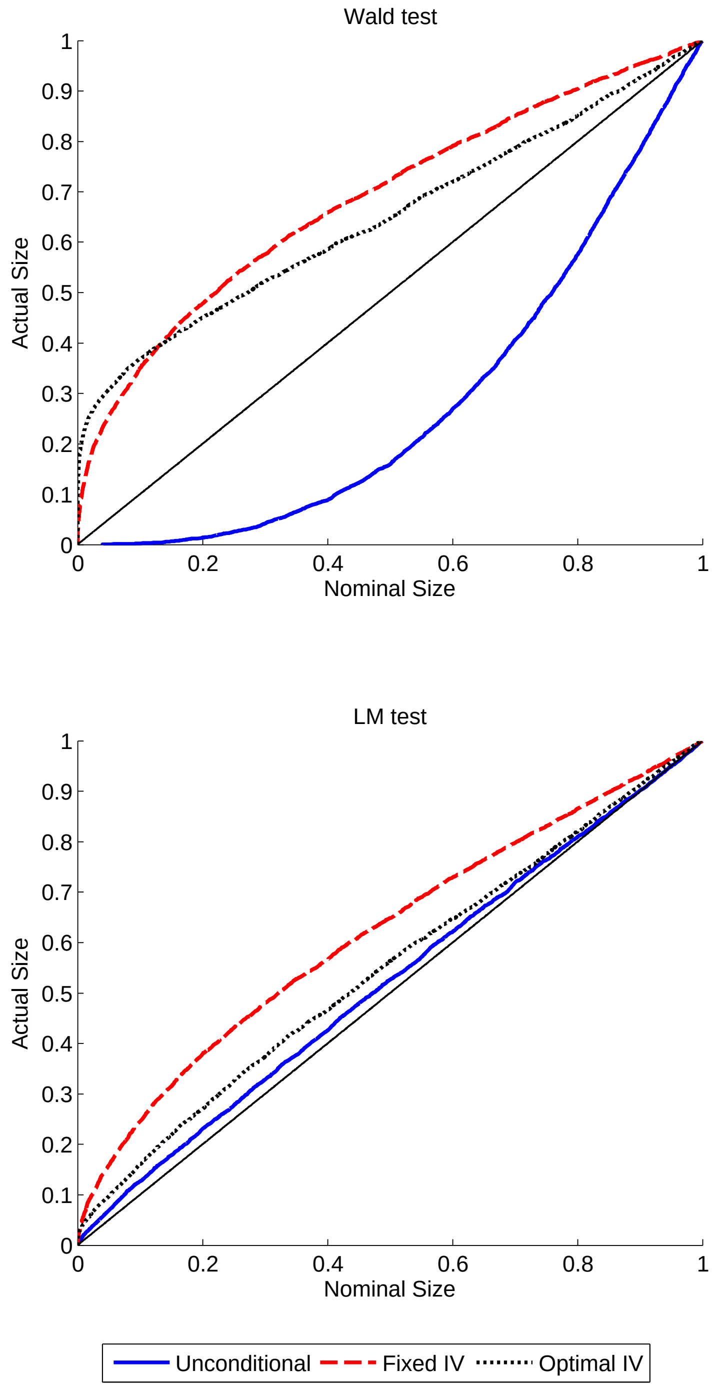

Figure IA.3. p**-value plots for Wald and LM tests of** $m_{t+1}^{\mathcal{N}}$

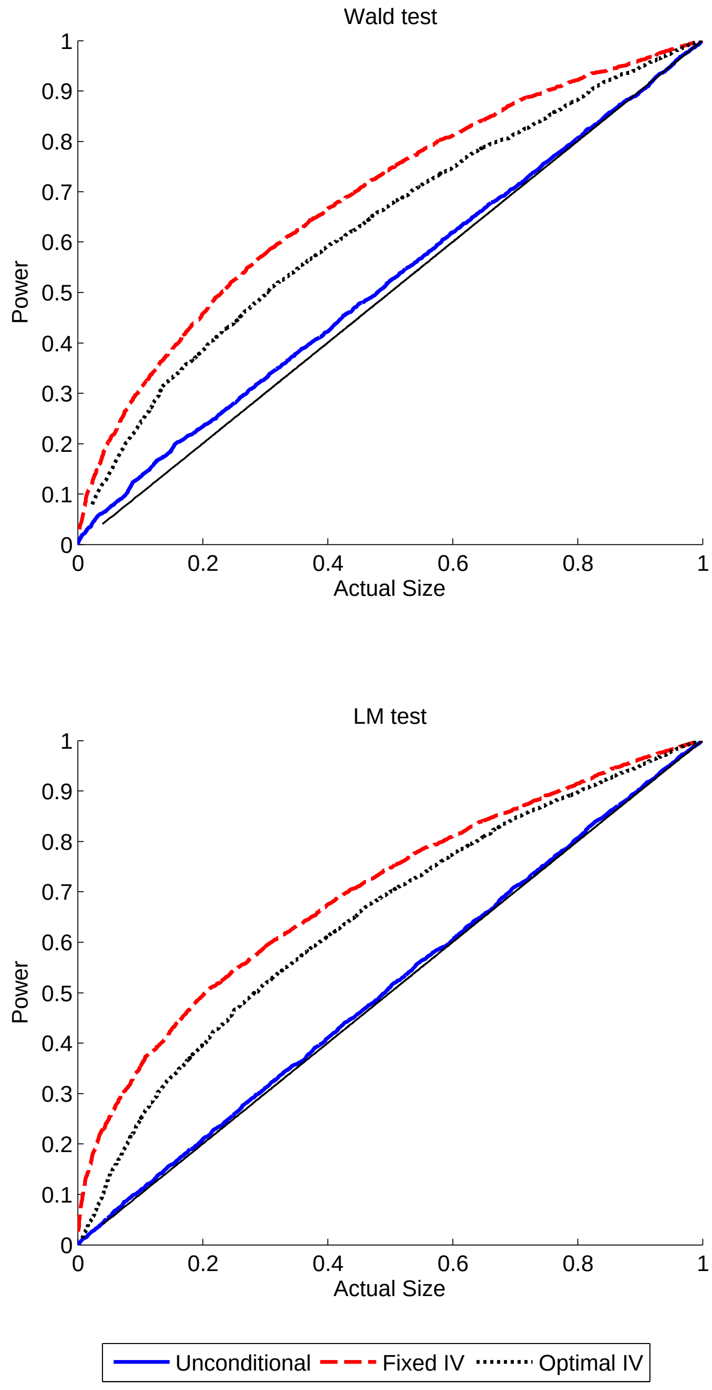

Figure IA.4. Size-power plots for Wald and LM tests of $m_{t+1}^{\mathcal{N}}$

null (as a function of nominal size) against the empirical distribution of p-values under the $m_{t+1}^{\mathcal{G}}$ alternative (as a function of nominal size). For example, this means that we ask how often the tests rejects under the null at a nominal size of 0.05, and we plot this number against the proportion of the simulations under the alternative that lead to rejection at a nominal size of 0.05.

As Figure IA.4 shows, the Wald and LM tests based on the unconditional estimator essentially have no power in our setting. The tests reject as frequently under the null as they do under the alternative hypothesis. The fixed IV and optimal IV estimators have similar properties and are more powerful than the test based on the unconditional estimator. However, if a size correction is implemented, as in these plots of power against actual size instead of nominal size, they have only moderate power. For example, with the nominal size of the LM test set such that actual size is 0.10 (this test rejects 10% of the time under the null), the tests based on fixed and optimal IV estimators reject around 30% of the time under the alternative. Clearly, these results will be sensitive to the distance between (γ_1, γ_2) under the null and the alternative, as well as the sample size.

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

References

Eichenbaum, M., L. Hansen, and K. Singleton, 1988, A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice Under Uncertainty, *Quarterly Journal of Economics* 103, 51–78.

Ferson, W., and A. Siegel, 2003, Stochastic Discount Factor Bounds with Conditioning Information, *Review of Financial Studies* 16, 567–595.

Hansen, Lars P., 1982, Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* 50, 1029–1054.

Hansen, Lars P., John C. Heaton, and Amir Yaron, 1996, Finite-Sample Properties of Some Alternative GMM Estimators, *Journal of Business and Economic Statistics* 14, 262–280.

Tauchen, G., 1986, Statistical Properties of GMM Estimates of Structural Parameters Using Financial Market Data, *Journal of Business and Economic Statistics* 4, 397–416.