THE CROSS-SECTION OF FOREIGN CURRENCY RISK PREMIA AND CONSUMPTION GROWTH RISK:
A COMMENT

Craig Burnside

The Cross-Section of Foreign Currency Risk Premia and Consumption Growth Risk: A Comment
Craig Burnside
NBER Working Paper No. 13129
May 2007
JEL No. F31,G12

ABSTRACT

Lustig and Verdelhan (2007) argue that the excess returns to borrowing US dollars and lending in foreign currency "compensate US investors for taking on more US consumption growth risk," yet these excess returns are all approximately uncorrelated with the consumption risk factors they study. Hence, their model cannot explain the cross-sectional variation of the returns. Their positive assessment results from allowing for a large constant in the model, and from ignoring sampling uncertainty in estimated betas used as explanatory variables in cross-sectional regressions that determine estimated consumption risk premia.

Craig Burnside
Department of Economics
Duke University
213 Social Sciences Building
Durham, NC 27708-0097
and NBER
burnside@econ.duke.edu

Lustig and Verdelhan (2007) claim that aggregate consumption growth risk explains the excess returns to borrowing U.S. dollars to finance lending in other currencies. They reach this conclusion after estimating a simple consumption-based asset pricing model using data on the returns of portfolios of short-term foreign-currency denominated money market securities sorted according to their interest differential with the U.S. They argue that the R^2 statistic corresponding to their benchmark estimates means that their model explains about 87 percent of the cross-sectional variation in expected returns.

To the contrary, I argue that their model explains very little of the cross-sectional variation in the expected returns of their portfolios. My reasoning is straightforward. Any risk-based explanation relies on significant spread in the covariances between the risk factors and the returns, yet, to a close approximation, the returns in Lustig and Verdelhan's data set are all uncorrelated with their consumption-based risk factors.

How could Lustig and Verdelhan reach the opposite conclusion based on the same data set? First, Lustig and Verdelhan misinterpret their own evidence by including in their model's *predicted* expected return a component that, in fact, should be interpreted as part of the model's pricing *error*. Were they to properly take this term into account their R^2 measure of fit would drop from 0.87 to no more than 0.34. Correctly measuring the pricing error—but using their approach to inference—would lead them to reject their model at well below the 5 percent level of significance.

The second problem centers on Lustig and Verdelhan's implementation of a two-pass procedure in estimating their model. The first pass is a time series regression of each portfolio's return on the risk factors. This regression determines the betas. The second pass is a cross-sectional regression of average portfolio returns on these betas. This regression determines the lambdas, or factor risk premia. Although it is common in the finance literature to do so, Lustig and Verdelhan do not focus on standard errors for the factor risk premia that correct for the betas being generated regressors in the second pass. Were they to do so, they would conclude that none of the factor risk premia are statistically significant. The very end of their paper presents GMM standard errors that correct for estimation of the betas, but these are inappropriately calculated and lead to incorrect inference.[1] I find that as long as

---

[1]Lustig and Verdelhan's corrected standard errors are displayed in their final table, not in the tables of benchmark results. The main discussion in their paper is based on uncorrected standard errors. They acknowledge that standard errors are higher if one uses the Shanken (1992) correction, or if one uses a bootstrap procedure, but they argue that the factor risk premia remain significant if, instead, a GMM procedure is used to estimate the model. For reasons I discuss below, their GMM procedure does not

1

sampling uncertainty in the factor betas is correctly taken into account, the estimates of the factor risk premia for Lustig and Verdelhan's model are only statistically significant in cases where the R^2 is negative.

Taken as a whole, the evidence does not favor Lustig and Verdelhan's story. In section 1 I review their model, data and methodological approach. In section 2, I present the first-pass estimates of the betas that underlie their estimates of the factor risk premia and demonstrate that there is little evidence of significant covariance between any of the portfolio returns and the risk factors. In section 3, I discuss the second-pass estimates of the factor risk premia and the correct interpretation of the pricing errors. In section 4, I recalculate standard errors for the factor risk premia taking the estimation of the betas into account. I discuss robustness of my negative findings in section 5. Section 6 concludes.

1 Model, Data, Estimation and Inference

1.1 Modeling the Stochastic Discount Factor

Lustig and Verdelhan work with Yogo's (2006) model, in which the representative household's lifetime utility at time t, U_t, is recursively represented as

$$U_t = \left\{ (1-\beta)u(C_t, D_t)^{1-1/\sigma} + \beta \left[E_t(U_{t+1}^{1-\gamma}) \right]^{(1-1/\sigma)/(1-\gamma)} \right\}^{1/(1-1/\sigma)}. \tag{1}$$

Here C_t represents the household's consumption of nondurable goods, D_t is the household's durable consumption, $0 < \beta < 1$ is the subjective discount factor, $\sigma > 0$ is the intertemporal elasticity of substitution and $\gamma > 0$ determines risk aversion. The instantaneous utility function is

$$u(C, D) = \left[(1-\alpha)C^{1-1/\rho} + \alpha D^{1-1/\rho} \right]^{1/(1-1/\rho)}. \tag{2}$$

Given this representation of preferences, the intertemporal marginal rate of substitution between $t-1$ and t is

$$M_t = \left\{ \beta \left(\frac{C_t}{C_{t-1}} \right)^{-1/\sigma} \left[\frac{v(D_t/C_t)}{v(D_{t-1}/C_{t-1})} \right]^{1/\rho - 1/\sigma} R_{Wt}^{1-1/\kappa} \right\}^{\kappa}, \tag{3}$$

where $\kappa = (1-\gamma)/(1-1/\sigma)$, R_{Wt} is the gross aggregate return to wealth and

$$v(D/C) = \left[1 - \alpha + \alpha(D/C)^{1-1/\rho} \right]^{1/(1-1/\rho)}. \tag{4}$$

---

produce standard errors appropriate to their two-pass estimation approach.

Lustig and Verdelhan work primarily with a log-linear approximation to (3). Let $c_t = \ln(C_t)$, $d_t = \ln(D_t)$ and $r_{Wt} = \ln(R_{Wt})$. Also define $\mu_c = E(\Delta c_t)$, $\mu_d = E(\Delta d_t)$ and $\mu_r = E(r_{Wt})$. A first-order Taylor series expansion of M_t in Δc_t, d_t/c_t, d_{t-1}/c_{t-1}, and r_{Wt} around the means of these variables is equivalent to

$$m_t = \xi[1 - b_c(\Delta c_t - \mu_c) - b_d(\Delta d_t - \mu_d) - b_r(r_{Wt} - \mu_r)], \tag{5}$$

where $\xi = \{\beta \exp[-\mu_c/\sigma + (1 - 1/\kappa)\mu_r]\}^\kappa$, $b_c = \kappa[1/\sigma + (1/\rho - 1/\sigma)\alpha]$, $b_d = \kappa\alpha(1/\sigma - 1/\rho)$ and $b_r = 1 - \kappa$. The approximation is valid in the neighborhood of $\rho = 1$.

1.2 Data

Lustig and Verdelhan form eight portfolios of long-positions in foreign currency with the U.S. dollar as the home currency. The real excess return to a long position in a foreign currency is

$$r_t^e = \frac{(1 + i_{t-1}^*)S_t/S_{t-1} - (1 + i_{t-1})}{1 + \pi_t}, \tag{6}$$

where i_{t-1}^* and i_{t-1} are, respectively, the interest rates on nominally riskless foreign currency denominated and U.S. dollar denominated securities held from date $t-1$ to date t, S_t is the exchange rate in U.S. dollars per unit of foreign currency, and π_t is the U.S. inflation rate between dates $t-1$ and t.

At each point in time, $t-1$, Lustig and Verdelhan sort individual foreign currencies into portfolios according to their interest differentials with the U.S., $i_{t-1}^* - i_{t-1}$, ordered from lowest to highest. The real excess return to portfolio i in period t, denoted R_{it}^e, is the simple average of the returns to the currencies that were placed in the ith portfolio at time $t-1$. As the interest differentials fluctuate over time, their ordering and the currency composition of the portfolios can change.

Lustig and Verdelhan's measure of C_t is the national income accounts measure of real per household consumption of nondurables and services excluding housing, clothing and shoes. Their measure of D_t assumes that the flow of consumption services from durables is proportional to the per household real stock of durable goods from the National Income and Product Accounts. Finally, their measure of the return on aggregate wealth is the value weighted return of the U.S. stock market, from the Kenneth French's database.

1.3   Estimation and Inference

Lustig and Verdelhan estimate the model by exploiting the null hypothesis that the approximated stochastic discount factor (SDF), m_t, prices the $n \times 1$ vector of portfolio excess returns, R_t^e. The pricing equation is

$$E(R_t^e m_t) = 0. \tag{7}$$

I rewrite (5) generically as

$$m_t = \xi[1 - (f_t - \mu)' b], \tag{8}$$

where f_t is a $k \times 1$ vector of risk factors, $\mu = E(f_t)$, b is a $k \times 1$ vector of coefficients, and ξ is a scalar representing the mean of the SDF.

It follows from (8) and (7) that

$$E(R_t^e) = \text{cov}(R_t^e, f_t')b = \underbrace{\text{cov}(R_t^e, f_t') \text{var}(f_t)^{-1}}_{\beta} \underbrace{\text{var}(f_t)b}_{\lambda}. \tag{9}$$

where β is a $n \times k$ matrix of factor betas, and λ is a $k \times 1$ vector of factor risk premia.

Lustig and Verdelhan estimate β and λ using a two-pass procedure. The first pass is a time series regression of each portfolio's excess return on the vector of risk factors:

$$R_{it}^e = a_i + f_t' \beta_i + \epsilon_{it}, \quad t = 1, \ldots, T, \text{ for each } i = 1, \ldots, n. \tag{10}$$

Here β_i' represents the ith row in β. The system of equations represented by (10) can be estimated equation-by-equation using OLS, or as a system using GLS, GMM, or maximum likelihood. Lustig and Verdelhan use the OLS approach. Given (9), the second pass is a cross-sectional regression of average portfolio returns on the estimated betas:

$$\bar{R}_i^e = \hat{\beta}_i' \lambda + \alpha_i, \qquad i = 1, \ldots, n, \tag{11}$$

where $\bar{R}_i^e = \frac{1}{T} \sum_{t=1}^{T} R_{it}^e$, $\hat{\beta}_i$ is the OLS estimate of β_i obtained in the first stage, and α_i is a pricing error. Let the OLS estimator of λ be $\hat{\lambda} = (\hat{\beta}' \hat{\beta})^{-1} \hat{\beta}' \bar{R}^e$, where $\bar{R}^e$ is an $n \times 1$ vector formed from the individual mean returns. The model's *predicted mean returns* are $\hat{\beta} \hat{\lambda}$ and the *pricing errors* are the residuals, $\hat{\alpha} = \bar{R}^e - \hat{\beta} \hat{\lambda}$.

The model's fit is assessed using the following statistic:

$$R^2 = 1 - \frac{(\bar{R}^e - \hat{\beta} \hat{\lambda})'(\bar{R}^e - \hat{\beta} \hat{\lambda})}{(\bar{R}^e - \ddot{R}^e)'(\bar{R}^e - \ddot{R}^e)}, \tag{12}$$

4

where $\ddot{R}^e = \frac{1}{n}\sum_{i=1}^{n}\bar{R}_i^e$ is the cross-sectional average of the mean returns in the data.

The model is tested on the basis of the estimated pricing errors using the statistic $C = T\hat{\alpha}'\Omega_{\hat{\alpha}}^{-1}\hat{\alpha}$, where $\Omega_{\hat{\alpha}}$ is the asymptotic covariance matrix of $\sqrt{T}\hat{\alpha}$ and the inverse is generalized. Cochrane (2005) discusses how to form $\Omega_{\hat{\alpha}}$ and shows that C is asymptotically χ^2_{n-k}.

It is common to include a constant in the second-pass regression as follows:

$$\bar{R}_i^e = \gamma + \hat{\beta}_i'\lambda + u_i, \qquad i = 1, \ldots, n. \tag{13}$$

The constant γ can be interpreted as the model's pricing error for the risk free rate. The statistical argument for running the regression without the constant is that we know with certainty that the excess return to a risk free asset is zero. One argument for including the constant is the notion that the risk free rate is imperfectly measured as the real return on T-bills.

Including the constant in the regression does not bias the estimates of λ, since, if the model is true, plim $\hat{\gamma} = 0$. Nonetheless, correctly interpreted, the model's *predicted mean returns* emerging from (13) should still be $\hat{\beta}\hat{\lambda}$ and the *pricing errors* should be the residuals *plus* the constant, $\hat{\alpha} = \hat{\gamma} + \hat{u} = \bar{R}^e - \hat{\beta}\hat{\lambda}$. Thus, the R^2 statistic should still be formed using (12) and the test of the pricing errors should be based on $\hat{\alpha} = \hat{\gamma} + \hat{u}$. At the very least, the economic and statistical significance of $\hat{\gamma}$ should be considered before a model with a constant is deemed reasonable.

One shortcoming of the R^2 statistic is that it is not bounded between 0 and 1 unless a constant is included in the second-pass regression and the predicted returns include the constant. In sample, the R^2 statistic defined in (12) can be negative. Nonetheless, if the null hypothesis is true, the probability limit of the R^2 statistic is 1, whether or not a constant is included in the second-pass regression.

Lustig and Verdelhan estimate their model exactly as described above using the second-pass regression that includes the constant. They do not, however, present results from the first pass of the procedure. This is key to my discussion in section 2. They measure predicted returns inclusive of the constant ($\hat{\gamma} + \hat{\beta}\hat{\lambda}$). They also exclude the constant from the pricing errors, which they measure as $\hat{u}$, and do not discuss the economic significance of the constant, $\hat{\gamma}$. This is central to my discussion in section 3.

2 First-Pass Estimates of the Betas

Table 1 presents first-pass estimates of the betas obtained by running the least squares regressions described by (10). Standard errors are computed using either the standard system OLS formulas, or a GMM-based VARHAC procedure.[2] The individual standard errors are similar across the two procedures. None of the betas reported in Table 1are individually statistically significant. Lustig and Verdelhan present factor betas in their Table 6, but these are individual factor betas, not the betas for their multifactor model. Nonetheless, there is little significance in these individual factor betas as well.

Joint tests for the significance of the betas lead to similar conclusions. If, for example, we test whether all the consumption betas are zero, the p-value for the χ_8^2 test statistic is 0.81 (using the VARHAC covariance matrix it is 0.80). For durables growth, the equivalent p-value is 0.62 (using the VARHAC covariance matrix it is 0.67). For the market return, the equivalent p-value is 0.37 (using the VARHAC covariance matrix it is 0.51). In fact, if we test whether all the betas are jointly zero, the χ_{24}^2 statistic has a p-value of 0.64, although with VARHAC standard errors the p-value is very small.

One might argue that it is the covariance of m_t and R_t^e that is crucial, not, per se, the covariance of f_t and R_t^e. That is, by forming linear combinations of the factors, one might induce significant cross-sectional spread in $\text{cov}(m_t, R_{it}^e)$ across i. One way to capture such spread is to measure the SDF's betas. Using (8), (7) can be rewritten as

$$E(R_t^e) = -\text{cov}(R_t^e, m_t)/E(m_t). \tag{14}$$

With the normalization $\xi = 1$, (8) implies that $E(m_t) = 1$, so we can rewrite the expression for $E(R_t^e)$ as

$$E(R_t^e) = -\frac{\text{cov}(R_t^e, m_t)}{\text{var}(m_t)} \text{var}(m_t) = \beta_m \sigma_m^2. \tag{15}$$

To measure m_t, using (8), we need values for the elements of b, so here I use Lustig and Verdelhan's GMM estimates of the elements of the b vector: $b_c = 37.0$, $b_d = 74.7$ and $b_r = 4.65$. A regression of R_{it}^e on m_t gives an estimate of β_{mi}, the SDF beta of the ith portfolio return. As Table 2(a) indicates, I find that none of the estimated SDF betas is significantly different from zero at the 5 percent level, when tested individually. The p-value

---

[2]The GMM-based standard errors I present in this paper are mainly computed using a variant of the VARHAC procedure described by den Haan and Levin (2000). I discuss the procedure in the appendix. Since Lustig and Verdelhan base some of their results on HAC standard errors, I present these when direct comparisons to their results are needed.

for a test that $\beta_{mi} = 0$ for all i is 0.44 (using the VARHAC covariance matrix it is 0.17). Table 2(b) repeats the exercise using the b vector corresponding to the calibrated model discussed in section I.E of Lustig and Verdelhan's paper: $b_c = 6.74$, $b_d = 23.3$ and $b_r = 0.31$. Here, only the return of portfolio #7 is significantly correlated with m_t. However, the p-value for a test that $\beta_{mi} = 0$ for all i is 0.50 (using the VARHAC covariance matrix it is 0.33). Table 2(c) repeats the exercise using the b vector corresponding to Lustig and Verdelhan's two pass estimates of λ: $b_c = -21.0$, $b_d = 130$ and $b_r = 4.46$. Again, the return of portfolio #7 is significantly correlated with m_t, but the p-value for a test that $\beta_{mi} = 0$ for all i is 0.47 (using the VARHAC covariance matrix it is 0.31). These results suggest that there is little significant spread in the constructed SDF betas.

Linear factor models rely, fundamentally, on there being significant spread in the covariance between the risk factors and the returns. The lack of statistical significance in the factor and SDF betas casts doubt on the hypothesis that Lustig and Verdelhan's model explains the cross-sectional variation in the expected returns. One might also be skeptical about the reliability of standard errors for factor risk premia computed treating the betas as known. I return to this issue in section 4.

3 Re-Interpreting the Pricing Errors and the Constant in the Second Pass

Given that there is little evidence of correlation between the risk factors and the returns, how did Lustig and Verdelhan reach the conclusion that the factors price the returns? They did so by focusing mainly on the second-pass estimates of λ. Furthermore, they estimated the second-pass regression using the representation that includes the constant, (13), but did not test the significance of the constant. Their results are reproduced in Table 3(a). Standard errors, like theirs, computed assuming that the first-pass betas are known are shown in the 'OLS' column. The factor risk premia for consumption and durables are both highly statistically significant. If a constant is included in the model's predicted expected return, the R^2 of the model is 0.87 and the p-value for the test for significance of the pricing errors is 0.483. These results are the basis of Lustig and Verdelhan's positive assessment of the model.

As I argued in section 1, however, there is good reason to reject the model on the basis of these same results. The theoretical model does not include a constant, and predicts that

7

the expected returns depend only on the covariance between the factors and the returns. So, whenever a constant is included in the second-pass regression it is important to consider its economic and statistical significance. In fact, as Table 3(a) indicates, the constant is big, implying a -3 percent per annum pricing error for the risk free rate. Furthermore, when the betas are treated as known, the constant is highly statistically significant, with a t-statistic of 3.4. Measurement error in the estimated betas, and resulting downward bias in the estimated factor risk premia, can explain a positive pricing error for the risk free rate. So can a liquidity premium in T-bills. But a large *negative* pricing error for the risk free rate is bad news for the model.

Excluding the constant from the model's predicted expected returns, the R^2 of the model is -2.6. In other words, the constant is crucial to the model's fit. This is confirmed by tests for the significance of the pricing errors inclusive of the constant. The p-value on the test falls to 0.001.

Of course, the model can be estimated excluding the constant. Results for this case are presented in Table 3(b). In this case the factor risk premia are much smaller, none of them are statistically significant, the R^2 is only 0.34, and the model is rejected based on the test of the pricing errors at below the 1 percent level (the 'OLS Covariance Matrix' case provides the p-value when the betas are treated as known).

A scatter plot of expected returns against factor betas would provide insight into the role of the constant, but constructing the scatter plot is not possible for a three factor model. I consider, instead, a scatter plot of expected returns against SDF betas. Figure 1 shows a scatter plot of $\bar{R}_i^e$ against $\hat{\beta}_{mi}$, $i = 1, \ldots, n$, with m_t being constructed using the b vector corresponding to the calibrated model discussed in section I.E of Lustig and Verdelhan's paper: $b_c = 6.74$, $b_d = 23.3$ and $b_r = 0.31$. Equation (15) implies that a scatter plot of $E(R_{it}^e)$ against the SDF betas, β_{mi}, should lie on a line through the origin with slope σ_m^2. The constructed m_t series has sample variance $\hat{\sigma}_m^2 = 0.29$, so I indicate an estimated version of this theoretical line, $\beta_m \hat{\sigma}_m^2$, in bold black in Figure 1. This line correctly prices the risk free asset, since the intercept is zero. It also correctly prices SDF risk because an SDF mimicking portfolio has a beta of 1, and a risk premium σ_m^2. The theoretical line, however, does not correctly price Lustig and Verdelhan's portfolio returns because the values of $(\hat{\beta}_{mi}, \bar{R}_i^e)$, indicated by the small circles in Figure 1, do not fit closely around the black line. Better fit can be obtained by running a regression of $\bar{R}_i^e$ on a constant and $\hat{\beta}_{mi}$. The

regression line (indicated in bold grey) has an intercept of $\hat{\gamma}_m = -3.0$ percent and a slope of $\hat{\lambda}_m = 1.3$. This means that although the regression line fits the scatter plot reasonably well, it misprices the risk free rate by -3 percent. It also misprices SDF risk because an SDF mimicking portfolio has a beta of 1, and an implied expected return of $\hat{\gamma}_m + \hat{\lambda}_m = -1.7$, rather than the theoretically predicted expected return, $\hat{\sigma}_m^2 = 0.29$. When Lustig and Verdelhan report high R^2 statistics for the calibrated model, it is because they measure fit with respect to the grey line, not the black line. They do not discuss the mispricing of the risk free rate or an SDF mimicking portfolio.

Figure 2 shows a scatter plot of the elements of $\bar{R}_i^e$ against the $\hat{\beta}_{mi}$, with m_t being constructed using the b corresponding to Lustig and Verdelhan's two pass estimates of λ: $b_c = -21.0$, $b_d = 130$ and $b_r = 4.46$. Once again, the scatter plot shows that the values of $(\hat{\beta}_{mi}, \bar{R}_i^e)$ do not fit the line $\beta_m \hat{\sigma}_m^2$ (indicated in bold black) with $\hat{\sigma}_m^2 \approx 5.8$. The regression line fit to the scatter (indicated in bold grey) has $\hat{\gamma}_m = -2.9$ percent and slope $\hat{\lambda}_m = \hat{\sigma}_m^2 = 5.8$, by construction. The regression line fits the scatter plot very well, but it misprices the risk free rate and SDF risk by the same amount: -2.9 percent. Again, Lustig and Verdelhan report high R^2 statistics for the estimated model because they measure fit with respect to the grey line and do not discuss the implied mispricing of the risk free rate or an SDF mimicking portfolio. A further problem with this version of the model is the implied value of $\hat{\sigma}_m$ is 2.41. Given that the mean of m_t is 1, by assumption, this is unrealistically large. Indeed, in sample, the constructed m_t is negative in 38 percent of the observations. Furthermore, the implied structural parameters are theoretically implausible. Lustig and Verdelhan set $\rho = 0.79$, so the other implied parameter values are $\alpha = 1.14$, $\gamma = 113$ and $\sigma = -0.032$.[3] The values of α and σ are theoretically inadmissible, the former implying that the marginal utility of nondurables is negative, and the latter implying a negative intertemporal elasticity of substitution.

Thus, on the basis of the Lustig and Verdelhan's own benchmark methodology, there is little evidence in favor of their model. Without the constant in the second pass regression, the estimated model fails to explain the expected returns. Additionally, the estimated SDF is economically implausible. In the next section I consider a further problem, which is that their benchmark inferences do not take into account that the betas are estimated.

---

[3]Lustig and Verdelhan report $\sigma = 0.21$. I explain this discrepancy in the appendix.

4 Correcting Inference for Estimated Betas

As Cochrane (2005) points out, the fact that the betas are estimated in the first pass matters for inference about the factor risk premia, and this is true even asymptotically. There are two standard ways to deal with this problem. One is to use the correction of the standard errors suggested by Shanken (1992). The other is to compute the standard errors using a GMM-based estimation procedure that replicates the point estimates. The latter procedure is more general, but as I show here, the two procedures deliver quite similar results.

By construction, neither procedure affects the point estimates of the factor risk premia, only the standard errors. The Shanken standard errors are a special case of the GMM standard errors when the ϵ_{it}, in (10), are i.i.d. and homoskedastic. The GMM procedure is described in Cochrane (2005). When the constant is included in the model the moment restrictions are $E(R_{it}^e - a_i - \beta_i' f_t) = 0$, $E[(R_{it}^e - a_i - \beta_i' f_t) f_t'] = 0$, and $E(R_{it}^e - \gamma - \beta_i' \lambda) = 0$ for $i = 1, \ldots, n$. When the constant is excluded from the model, the last set of moment restrictions is replaced by $E(R_{it}^e - \beta_i' \lambda) = 0$ for $i = 1, \ldots, n$. In both cases, an identity matrix is used to weight the moment conditions.

The Shanken and GMM-corrected standard errors for the model with the constant [Table 3(a)] are roughly two to three times larger than the OLS standard errors that ignore estimation of the betas. Why is the Shanken correction so big? Let $\theta = (\ \gamma \quad \lambda' \ )'$, $X = (\ \iota \quad \beta\ )$, where ι is an $n \times 1$ unit vector, $\Sigma = E(\epsilon_t \epsilon_t')$ and $\Sigma_f = E[(f_t - \mu)(f_t - \mu)']$, and let $\tilde{\Sigma}_f$ be a matrix with a leading column and row of zeros, and Σ_f in the lower right corner. When the betas are treated as known the covariance matrix of $\sqrt{T}(\hat{\theta} - \theta)$ is

$$\Omega_{\hat{\theta}} = (X'X)^{-1} X' \Sigma X (X'X)^{-1} + \tilde{\Sigma}_f. \tag{16}$$

With the Shanken correction the covariance matrix is

$$\Omega_{\hat{\theta}} = (1 + \lambda' \Sigma_f^{-1} \lambda)(X'X)^{-1} X' \Sigma X (X'X)^{-1} + \tilde{\Sigma}_f. \tag{17}$$

For the single factor CAPM model, using annual returns of the Fama-French 25 portfolios sorted on size and book-to-market value over the period 1953–2002, the Shanken-correction term is just $1 + \lambda^2/\sigma_f^2 = 1.035$. In Lustig and Verdelhan's case, $1 + \lambda' \Sigma_f^{-1} \lambda = 6.79$. Although the individual λs in Lustig and Verdelhan's model are of the same order of magnitude as for the CAPM, the consumption factors have much smaller variance than the market return. This blows up the size of the Shanken correction substantially. The term

$(X'X)^{-1}X'\Sigma X(X'X)^{-1}$ also has a relatively larger share in $\Omega_{\hat{\theta}}$ in Lustig and Verdelhan's case than for the CAPM.

Using either the Shanken or GMM standard errors, none of the estimated factor risk premia in Table 3(a) are statistically significant at the 5 percent level, though the λ corresponding to durables growth is significant at the 10 percent level. There is so much sampling uncertainty in the model with the constant than when the significance of the pricing errors is tested, the model cannot be rejected whether or not the constant is included in the predicted expected returns. But this is not a success for the model. It is simply a consequence of the fact that the model's predicted expected returns are estimated with an enormous degree of uncertainty. The scatter plot of expected returns against SDF betas is informative about why taking into account estimation of the betas matters so much for the test. Figure 3 is a duplicate of Figure 2, but adds two standard error bars around the $(\hat{\beta}_{mi}, \bar{R}_i^e)$. The error bars illustrate the enormous degree of uncertainty about not only the betas, but also the expected returns.

Lustig and Verdelhan (2007) present uncorrected standard errors on the basis that "Jagannathan and Wang (1998) show that the [uncorrected procedure] does not necessarily overstate the precision of the standard errors if conditional heteroskedasticity is present." Indeed, Jagannathan and Wang work out the GMM-based asymptotic theory for inference under more general conditions than Shanken, and under these more general conditions the direction of the bias in the standard errors is unclear. However, as we see in Table 3(a), the GMM-based standard errors are of roughly the same magnitude as the Shanken-corrected standard errors. So while, in principle, the uncorrected procedure need not produce downwardly biased standard errors, it would appear to do so in this case.

The degree of precision in the model without the constant [Table 3(b)] is greater, presumably because the $\beta\lambda$ term has to fit the mean of $\bar{R}^e$ without the use of a constant, and this puts a much tighter restriction on the statistically admissible set of λs. Nonetheless, the Shanken and GMM-corrected standard errors are larger than the OLS standard errors by a factor of about 1.4. None of the factor risk premia is statistically significant for the model without the constant, no matter how standard errors are computed. The pricing errors are sufficiently big that the model is rejected at the 6 percent level using the Shanken covariance matrix, but only at the 17 percent level with GMM covariance matrix.

Given these results I conclude that the model is rejected (or borderline rejected) absent

the constant and that the factor risk premia are statistically insignificant once we take into account sampling uncertainty in the betas. In the next section I explore the robustness of these findings to direct GMM estimation of the model.

5   Robustness

GMM Estimation of the Model   Lustig and Verdelhan claim robustness of their uncorrected standard errors by appealing to a GMM procedure that estimates the model directly. This procedure, which is described in more detail in Cochrane (2005), estimates the model, (8), by GMM using the moment conditions:

$$E\{R_t^e[1 - (f_t - \mu)' b]\} = 0 \tag{18}$$
$$E(f_t - \mu) = 0 \tag{19}$$

Estimates of $\lambda = \Sigma_f b$ can be obtained from the procedure by adding moment conditions that identify the unique elements of Σ_f.

There are two problems with Lustig and Verdelhan using the standard errors for $\hat{\lambda}$ from this GMM procedure to benchmark the standard errors from the two-pass procedure. First, and most importantly, the GMM procedure does not include a constant in the vector λ, so it does not produce comparable standard errors to the ones appropriate to the two-pass procedure *with* a constant in the second pass. Rather, the standard errors are more appropriately compared to the standard errors from the two-pass procedure *without* a constant. Second, the GMM procedure produces the same λ estimates as the two-pass procedure only when an identity matrix is used as the weighting matrix. This is typically the case in the first stage of GMM. Lustig and Verdelhan report results from a second stage of GMM that uses a different weighting matrix. Therefore the standard errors of the corresponding $\hat{\lambda}$s are not valid estimates of the standard errors of the $\hat{\lambda}$s produced by the two-pass procedure.

Of course, estimates of λ based on the first and second stages of the GMM procedure are of independent interest and shed additional light on the model. In Table 4(a) I present the results from the first stage of GMM. Here, none of the b coefficients is individually significant, nor are any of the λs. The R^2 of the model is 0.34. As explained in the appendix, the point estimates of the λs and the R^2 measure of fit are, by construction, the same as those produced by the two-pass procedure without the constant [see Table 3(b)].

Turning to the second stage of GMM [Table 4(b)] I reproduce Lustig and Verdelhan's

results. I present HAC and VARHAC standard errors for comparability to their Table 14. Most importantly, the second stage R^2 of the model is negative (-0.66). This is very bad news, because it indicates that a constant would do a better job explaining the cross-sectional distribution of the returns than the model does. Also, the model only marginally passes the test of the overidentifying restrictions at the 7 percent level. So, while λ is statistically significant for consumption and durables, and b is significant for durables, the model has very poor fit.

The model *with* the constant can also be estimated using GMM by adding an arbitrary constant to the moment condition, (18), replacing it with

$$E\{R_t^e[1 - (f_t - \mu)' b] - \gamma\} = 0. \tag{20}$$

Table 5(a) presents results from the first stage of GMM. As explained in the appendix, the first stage of GMM reproduces the point estimates from the two-pass procedure with the constant [see Table 3(a)], by construction. None of the b coefficients are individually statistically significant, nor are any of the λs. The R^2 of the model is 0.87 but, as in Table 3, if the constant is excluded from the predicted expected return the R^2 is sharply negative.

Turning to second stage GMM estimates [Table 5(b)], the picture is much the same. None of the bs and λs are statistically significant. Consistent with Table 3(a), the model cannot be rejected on the basis of the pricing errors, but this is not a success for the model, it is simply a reflection of the imprecision in the estimates.

I conclude that if we focus our attention on the model without the constant, and on estimates obtained using the two-pass procedure (or, equivalently, the first stage of GMM), it is a robust finding that the R^2 fit of the model is only 0.34 and none of the factor bs and λs are statistically significant. If we base our parameter estimates on the second stage of the GMM procedure, a subset of the factor bs and λs becomes statistically significant, but only at the cost that the fit of the model becomes negative. For the model with the constant, it is a robust finding that none of the model parameters are statistically significant.

Choice of Sample   Lustig and Verdelhan present results for two sample periods, 1953–2002 and 1971–2002. Does choosing the shorter sample (1971–2002) affect my negative conclusions? The answer is mainly no.[4]

---

[4]Full results for the 1971–2002 sample are available upon request.

Consider, first, the factor betas. For the shorter sample, the durables beta of the 4th portfolio is statistically significant at the 5 percent level. This is the only exception to the lack of significance I demonstrated for the full sample in Table 1.

Consider, next, the second pass estimates of the model that ignore sampling uncertainty in the betas. Once again, the fit of the model depends on the presence of the constant. If the constant is included in the model, and the predicted expected returns include the constant, the R^2 is 0.64. However, the constant remains large, negative and statistically significant. Once the constant is dropped from the model, the R^2 falls to 0.38, and the test of the pricing errors rejects the model.

When sampling uncertainty in the betas is taken into account using the Shanken correction or GMM-based standard errors, the factor risk premia from the two-pass procedure with the constant become statistically insignificant, as for the full sample. The same is true for the model without the constant, except for the market return and GMM-based standard errors.

If the model without the constant is estimated by GMM, the bs are statistically insignificant (except in the second stage of GMM for r_W). The λs are statistically insignificant in the first stage of GMM, but are all significant in the second stage of GMM. However, the R^2 at the second stage is 0.06, indicating that, in any case, almost none of the spread in the expected returns is explained by the risk factors.

6   Conclusion

Lustig and Verdelhan's consumption-based model does not explain the cross-sectional variation in the expected returns of their portfolios. A risk-based story requires that at least some of the returns be correlated with the risk factors. As the first-pass regressions reported in section 2 demonstrate, however, Lustig and Verdelhan's risk factors are very close to being uncorrelated with the returns. A symptom of this is that there is no statistically significant spread in the factor betas. Given these facts, there is little evidence to support Lustig and Verdelhan's hypothesis.

They draw the opposite conclusion based on favorable measures of fit, statistically significant risk premia, and tests of the pricing errors based on second-pass regressions. The R^2 they report is over-stated because it relies on the inclusion of a constant in the model. This constant does not belong in the model under the null, yet is significantly different from

zero using their primary approach to inference. The tests of the pricing errors that they present also overstate the success of the model for the same reason because they include the constant. Exclusion of the constant implies an R^2 that is never greater than 0.34, regardless of the estimation procedure used.

Lustig and Verdelhan also largely ignore estimation of the betas when conducting inference about factor risk premia. Once inference takes into account estimation of the betas, I find that the estimated factor risk premia for their model are usually statistically insignificant. In the one case where the factor risk prices are significant [two-stage GMM, shown in Table 4(b)] the model has very poor fit, with an R^2 of -0.66. Thus, the second-pass regressions and GMM-based estimates simply confirm what is clear from the first-pass estimates: a model based primarily on aggregate U.S. consumption and durables growth cannot explain observed currency risk premia.

References

Cochrane, John H. (2005) *Asset Pricing*, Revised edition. Princeton: Princeton University Press.

den Haan, Wouter J. and Andrew T. Levin (2000) "Robust Covariance Matrix Estimation with Data-Dependent VAR Prewhitening Order," NBER Technical Working Paper No. 255.

Hansen, Lars P. (1982) "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica* 50, 1029–54.

Jagannathan, Ravi and Zhenyu Wang (1998) "An Asymptotic Theory for Estimating Beta-Pricing Models Using Cross-Sectional Regressions," *Journal of Finance* 53, 1285–1309.

Lustig, Hanno and Adrien Verdelhan (2007) "The Cross-Section of Foreign Currency Risk Premia and Consumption Growth Risk," *American Economic Review* 97, 89–117.

Shanken, Jay (1992) "On the Estimation of Beta-Pricing Models," *The Review of Financial Studies* 5, 1–33.

Yogo, Motohiro (2006) "A Consumption-Based Explanation of Expected Stock Returns," *Journal of Finance* 61, 539–80.

Portfolio (i)	$\bar{R}_i^e$	β_i'		
		Δc	Δd	r_W
# 1 (smallest $i^* - i$)	-2.336	0.201	0.028	-0.068
	(0.897)	(0.852)	(0.612)	(0.055)
	[0.906]	[0.643]	[0.568]	(0.049)
# 2	-0.873	0.740	0.091	-0.034
	(0.937)	(0.889)	(0.638)	(0.058)
	[0.947]	[0.620]	[0.470]	[0.062]
# 3	-0.747	-0.639	0.962	0.019
	(0.935)	(0.882)	(0.633)	(0.057)
	[1.361]	[1.026]	[0.814]	[0.058]
# 4	0.329	-0.546	0.982	-0.089
	(1.190)	(1.095)	(0.786)	(0.071)
	[1.202]	[1.075]	[0.749]	[0.069]
# 5	-0.151	0.180	0.485	0.009
	(1.053)	(1.006)	(0.722)	(0.065)
	[1.063]	[0.754]	[0.714]	[0.065]
# 6	-0.213	-0.755	1.079	0.023
	(1.148)	(1.089)	(0.781)	(0.071)
	[1.160]	[0.958]	[0.833]	[0.068]
# 7	2.988	0.036	1.234	-0.027
	(1.144)	(1.044)	(0.749)	(0.068)
	[1.155]	[0.797]	[0.730]	[0.063]
# 8 (largest $i^* - i$)	2.031	-1.342	1.426	0.079
	(1.756)	(1.674)	(1.201)	(0.108)
	[2.586]	[1.646]	[1.225]	[0.114]

Notes: Annual data, 1953–2002. The regression equation is $R_{it}^e = a_i + f_t'\beta_i + \epsilon_{it}$, where R_{it}^e is the excess return of portfolio i at time t, $f_t = (\ \Delta c_t \quad \Delta d_t \quad r_{Wt}\ )'$, Δc is real per household consumption (nondurables & services) growth, Δd is real per household durable consumption growth, and r_W is the value weighted US stock market return. The portfolios are equally-weighted groups of short-term foreign-currency denominated money market securities sorted according to their interest differential with the US ($i^* - i$). The table reports β_i' and the sample mean of each portfolio return, $\bar{R}_i^e$. OLS standard errors are in parentheses. GMM-VARHAC standard errors are in square brackets.

TABLE 2: Estimates of the SDF Betas for Specific Values of b

b vector		(a)	(b)	(c)
b_c		37.0	6.74	−21.0
b_d		74.7	23.3	129.9
b_r		4.65	0.31	4.46

Portfolio (i)	$\bar{R}_i^e$	β_{im}		
		(a)	(b)	(c)
# 1 (smallest $i^* - i$)	−2.336	−0.030	0.639	0.060
	(0.897)	(0.452)	(1.665)	(0.373)
	[0.906]	[0.658]	[2.594]	[0.544]
# 2	−0.873	0.437	1.925	0.335
	(0.937)	(0.468)	(1.720)	(0.387)
	[0.947]	[0.285]	[1.064]	[0.241]
# 3	−0.747	0.513	2.229	0.549
	(0.935)	(0.465)	(1.709)	(0.381)
	[1.361]	[0.526]	[2.618]	[0.542]
# 4	0.329	0.308	2.692	0.539
	(1.190)	(0.597)	(2.178)	(0.489)
	[1.202]	[1.031]	[2.289]	[0.702]
# 5	−0.151	0.560	2.158	0.467
	(1.053)	(0.524)	(1.932)	(0.433)
	[1.063]	[0.553]	[2.031]	[0.470]
# 6	−0.213	0.556	2.422	0.603
	(1.148)	(0.573)	(2.106)	(0.470)
	[1.160]	[0.684]	[3.233]	[0.705]
# 7	2.988	1.038	4.714	0.999
	(1.144)	(0.557)	(2.019)	(0.454)
	[1.155]	[0.568]	[2.043]	[0.465]
# 8 (largest $i^* - i$)	2.031	0.653	2.403	0.694
	(1.756)	(0.879)	(3.245)	(0.723)
	[2.586]	[0.844]	[3.481]	[0.697]

Notes: Annual data, 1953–2002. The regression equation is $R_{it}^e = a_i + m_t \beta_{im} + \epsilon_{it}$, where R_{it}^e is the excess return of portfolio i at time t, $m_t = 1 - (f_t - \bar{f})'b$, $f_t = (\; \Delta c_t \quad \Delta d_t \quad r_{Wt} \;)'$, Δc is real per household consumption (nondurables & services) growth, Δd is real per household durable consumption growth, r_W is the value weighted US stock market return, $\bar{f}$ is the sample mean of f_t and b takes on the value indicated for each case. The portfolios are equally-weighted groups of short-term foreign-currency denominated money market securities sorted according to their interest differential with the US ($i^* - i$). The table reports β_{im} and the sample mean of each portfolio return, $\bar{R}_i^e$. OLS standard errors are in parentheses. GMM-VARHAC standard errors are in square brackets.

TABLE 3: SECOND-PASS ESTIMATES OF THE FACTOR RISK PREMIA

	(a) Model With a Constant				(b) Model Without a Constant			

Factor Risk Premia

| | | | Standard Error of $\hat{\lambda}$ | | | | Standard Error of $\hat{\lambda}$ | |
Factor	$\hat{\lambda}$	OLS	Shanken	GMM	$\hat{\lambda}$	OLS	Shanken	GMM
Constant (γ)	-2.94	(0.86)	[2.23]	{2.66}				
Δc	2.19	(0.83)	[2.11]	{2.48}	0.59	(0.73)	[1.01]	{1.17}
Δd	4.70	(0.97)	[2.42]	{2.41}	1.10	(1.02)	[1.40]	{1.69}
r_W	3.33	(7.59)	[18.8]	{23.1}	11.7	(7.40)	[10.1]	{10.6}

R^2 Measures of Fit

Predicted $ER^e = \hat{\gamma} + \hat{\beta}\hat{\lambda}$	0.87	
Predicted $ER^e = \hat{\beta}\hat{\lambda}$	-2.62	0.34

Tests of the Pricing Errors

| | Covariance Matrix | | | Covariance Matrix | | |
	OLS	Shanken	GMM	OLS	Shanken	GMM
$\hat{u} = \bar{R}^e - \hat{\gamma} - \hat{\beta}\hat{\lambda}$	0.483	0.972	0.994			
$\hat{\alpha} = \bar{R}^e - \hat{\beta}\hat{\lambda}$	0.001	0.685	0.891	0.001	0.031	0.173

Notes: Part (a) reports results from running the cross-sectional regression $\bar{R}_i^e = \gamma + \hat{\beta}_i'\lambda + u_i$.where $\bar{R}_i^e$ is the mean excess return of portfolio i and $\hat{\beta}_i$ is the vector of factor betas of portfolio i estimated in the first pass regression. Part (b) reports results from the a cross-sectional regression without the constant: $\bar{R}_i^e = \hat{\beta}_i'\lambda + \alpha_i$. For the factor risk premia ($\hat{\lambda}$) OLS standard errors are in parentheses, Shanken standard errors are in square brackets, and GMM-VARHAC standard errors are in braces. For the tests of the pricing errors I compute the test statistic for each of the three methods of computing the covariance matrix of $\hat{u}$ or $\hat{\alpha}$, and report the p-value associated with the test-statistic. The results in part (a) match Lustig and Verdelhan's exactly except for (i) the p-value on the test of the pricing error (OLS case) and (ii) the Shanken standard errors. I explain these differences in the appendix.

TABLE 4: GMM ESTIMATES OF THE MODEL WITH NO CONSTANT

| Factor | (a) 1st Stage of GMM | | (b) 2nd Stage of GMM | |
	$\hat{b}$	$\hat{\lambda}$	$\hat{b}$	$\hat{\lambda}$
Δc	−22.0	0.59	37.0	2.37
	(63.6)	(1.18)	(45.4)	(1.05)
			[44.8]	[0.99]
Δd	45.5	1.10	74.7	3.48
	(51.0)	(1.72)	(33.5)	(1.27)
			[32.0]	[1.07]
r_W	5.16	11.7	4.65	10.2
	(3.02)	(9.02)	(2.82)	(8.10)
			[2.56]	[7.08]
R^2		0.34		−0.66
Test of the Pricing Errors			0.068	

Notes: Part (a) reports estimates of b and λ obtained in the 1st stage of GMM, by exploiting the moment restrictions $E\{R_t^e[1 - (f_t - \mu)' b]\} = 0$, $E(f_t - \mu) = 0$ and $E[(f_t - \mu)(f_t - \mu)' - \Sigma_f] = 0$, where $f_t = (\ \Delta c_t \quad \Delta d_t \quad r_{Wt}\ )'$, Δc is real per household consumption (nondurables & services) growth, Δd is real per household durable consumption growth, r_W is the value weighted US stock market return. Part (b) reports estimates obtained in the 2nd stage of GMM. GMM-VARHAC standard errors are in parentheses. GMM-HAC standard errors are in square brackets in part (b) for direct comparison to Lustig and Verdelhan's Table 14. For the test of the pricing errors I report the p-value associated with the test-statistic. The appendix provides details of the weighting matrices at each stage, and explains the equivalence of the GMM approach to the two-pass method. It also explains why the test of the pricing errors is the same at both stages of GMM. The point estimates in part (b) match Lustig and Verdelhan's exactly but the HAC standard errors do not. I explain this difference in the appendix.

TABLE 5: GMM ESTIMATES OF THE MODEL WITH THE CONSTANT

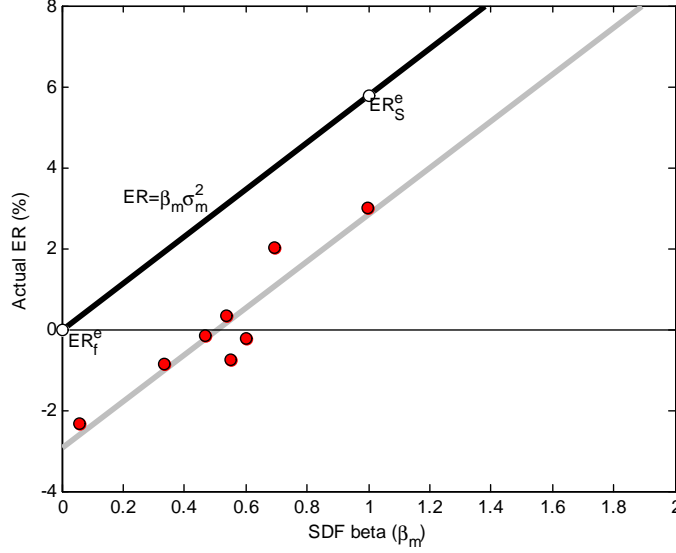| Factor	(a) 1st Stage of GMM		(b) 2nd Stage of GMM	
	$\hat{b}$	$\hat{\lambda}$	$\hat{b}$	$\hat{\lambda}$
Constant (γ)		−2.94		−3.02
		(2.92)		(2.52)
Δc	−21.0	2.19	27.2	2.73
	(88.6)	(2.09)	(74.8)	(2.01)
Δd	129.9	4.70	108.3	4.85
	(109.5)	(3.57)	(108.8)	(3.67)
r_W	4.46	3.33	2.53	0.78
	(5.10)	(13.3)	(4.66)	(11.4)
R^2				
Predicted $ER^e = \hat{\gamma} + \hat{\beta}\hat{\lambda}$		0.87		0.81
Predicted $ER^e = \hat{\beta}\hat{\lambda}$		−2.62		−2.69
Test of the Pricing Errors		0.703		

Notes: Part (a) reports estimates of b and λ obtained in the 1st stage of GMM, by exploiting the moment restrictions $E\{R_t^e[1 - (f_t - \mu)'b] - \gamma\} = 0$, $E(f_t - \mu) = 0$ and $E[(f_t - \mu)(f_t - \mu)' - \Sigma_f] = 0$, where $f_t = (\ \Delta c_t \quad \Delta d_t \quad r_{Wt}\ )'$, Δc is real per household consumption (nondurables & services) growth, Δd is real per household durable consumption growth, r_W is the value weighted US stock market return. Part (b) reports estimates obtained in the 2nd stage of GMM. GMM-VARHAC standard errors are in parentheses. For the test of the pricing errors I report the p-value associated with the test-statistic.

FIGURE 1

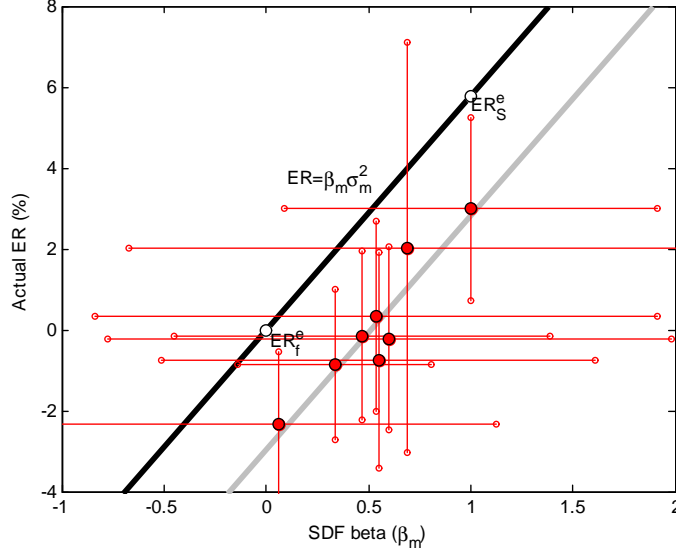SDF Betas and Expected Returns for the Calibrated Model
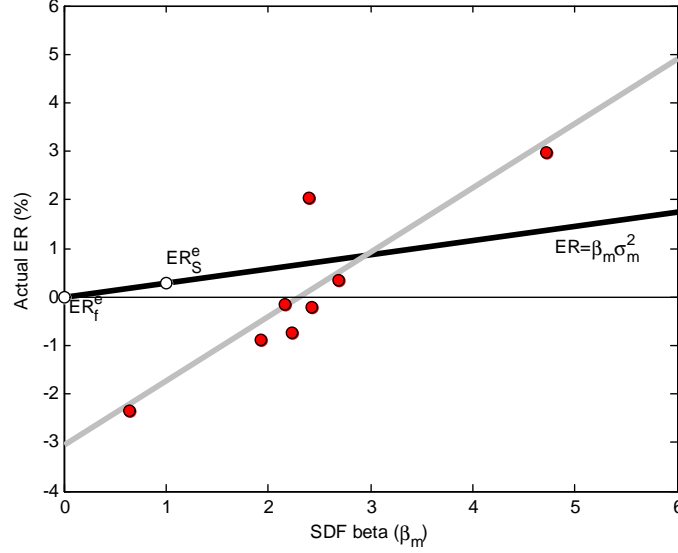
Note: Filled circles represent (SDF beta, mean excess return) pairs for Lustig and Verdelhan's eight portfolios. The "SDF beta", $\hat{\beta}_m$, for each portfolio is slope coefficient from a regression of the portfolio excess return, R_{it}^e, on the SDF, $m_t = 1 - \left(f_t - \bar{f} \right)' b$, with $f_t = (\begin{array}{ccc} \Delta c_t & \Delta d_t & r_{Wt} \end{array})'$, Δc is real per household consumption (nondurables & services) growth, Δd is real per household durable consumption growth, r_W is the value weighted US stock market return, $\bar{f}$ is the sample mean of f_t and b corresponding to the calibrated model with $b_c = 6.74$, $b_d = 23.3$ and $b_r = 0.31$. "Actual ER" is the sample mean of the portfolio return, $\bar{R}_i^e$. The black line corresponds to $\beta_m \hat{\sigma}_m^2$ where $\hat{\sigma}_m^2$ is the variance of the constructed SDF. The grey line is the estimated regression line $\bar{R}_i^e = \hat{\gamma}_m + \hat{\lambda}_m \hat{\beta}_m$. The empty circle marked ER_f^e signifies that a risk free asset has a zero beta, and a zero excess return. The empty circle marked ER_S^e signifies that an SDF mimicking portfolio has a beta of 1 and expected excess return of σ_m^2.

FIGURE 2

SDF Betas and Expected Returns for the Benchmark Estimated Model

Note: Filled circles represent (SDF beta, mean excess return) pairs for Lustig and Verdelhan's eight portfolios. The "SDF beta", $\hat{\beta}_m$, for each portfolio is slope coefficient from a regression of the portfolio excess return, R^e_{it}, on the SDF, $m_t = 1 - \left(f_t - \bar{f}\right)' b$, with $f_t = \left(\begin{array}{ccc} \Delta c_t & \Delta d_t & r_{Wt} \end{array}\right)'$, Δc is real per household consumption (nondurables & services) growth, Δd is real per household durable consumption growth, r_W is the value weighted US stock market return, $\bar{f}$ is the sample mean of f_t and b corresponding to Lustig and Verdelhan's two pass estimates of λ: $b_c = -21.0$, $b_d = 130$ and $b_r = 4.46$. "Actual ER" is the sample mean of the portfolio return, $\bar{R}^e_i$. The black line corresponds to $\beta_m \hat{\sigma}^2_m$ where $\hat{\sigma}^2_m$ is the variance of the constructed SDF. The grey line is the estimated regression line $\bar{R}^e_i = \hat{\gamma}_m + \hat{\lambda}_m \hat{\beta}_m$. The empty circle marked ER^e_f signifies that a risk free asset has a zero beta, and a zero excess return. The empty circle marked ER^e_S signifies that an SDF mimicking portfolio has a beta of 1 and expected excess return of σ^2_m.

FIGURE 3

SDF Betas, Expected Returns and Two Standard Error Bars
for the Benchmark Estimated Model

Note: Filled circles represent (SDF beta, mean excess return) pairs for Lustig and Verdelhan's eight portfolios. The "SDF beta", $\hat{\beta}_m$, for each portfolio is slope coefficient from a regression of the portfolio excess return, R^e_{it}, on the SDF, $m_t = 1 - \left(f_t - \bar{f}\right)' b$, with $f_t = \left(\begin{array}{ccc} \Delta c_t & \Delta d_t & r_{Wt} \end{array} \right)'$, Δc is real per household consumption (nondurables & services) growth, Δd is real per household durable consumption growth, r_W is the value weighted US stock market return, $\bar{f}$ is the sample mean of f_t and b corresponding to Lustig and Verdelhan's two pass estimates of λ: $b_c = -21.0$, $b_d = 130$ and $b_r = 4.46$. The horizontal lines at each circle are two standard error bands around $\hat{\beta}_m$. "Actual ER" is the sample mean of the portfolio return, $\bar{R}^e_i$. The vertical lines are two standard error bands around $\bar{R}^e_i$. The black line corresponds to $\beta_m \hat{\sigma}^2_m$ where $\hat{\sigma}^2_m$ is the variance of the constructed SDF. The grey line is the estimated regression line $\bar{R}^e_i = \hat{\gamma}_m + \hat{\lambda}_m \hat{\beta}_m$. The empty circle marked ER^e_f signifies that a risk free asset has a zero beta, and a zero excess return. The empty circle marked ER^e_S signifies that an SDF mimicking portfolio has a beta of 1 and expected excess return of σ^2_m.

7 Appendix

7.1 Standard Errors for the Two-Pass Procedure

Lustig and Verdelhan compute standard errors under the assumption that the betas are known. I first, consider this case, and then consider the case where the betas are treated as generated regressors. The derivations here are reproduced from or based on Cochrane (2005) and Shanken (1992).

7.1.1 Betas are Known

Equation (10) can be rewritten as

$$R_t^e = a + \beta f_t + \epsilon_t$$

where a is an $n \times 1$ vector formed from the individual a_i, and ϵ_t is an $n \times 1$ vector formed from the individual ϵ_{it}. Traditionally the factors and errors are assumed to be i.i.d. over time, with $\mathrm{var}(f_t) = \Sigma_f$ and $\mathrm{var}(\epsilon_t) = \Sigma$.[5] Taking averages over time:

$$\bar{R}^e = a + \beta \bar{f} + \bar{\epsilon}, \tag{21}$$

where $\bar{R}^e$, $\bar{f}$ and $\bar{\epsilon}$ are the sample means of R_t^e, f_t and ϵ_t.

Without a Constant   When the betas are known and the second stage excludes a constant $\hat{\lambda} = (\beta'\beta)^{-1}\beta'\bar{R}^e$. This implies that

$$\hat{\alpha} = \bar{R}^e - \beta\hat{\lambda} = \left[I - \beta(\beta'\beta)^{-1}\beta'\right]\bar{R}^e = M_\beta \bar{R}^e.$$

Given (9), this implies that

$$\mathrm{plim}\,\hat{\alpha} = M_\beta E(\bar{R}^e) = M_\beta \beta \lambda = 0.$$

Also, the asymptotic covariance matrix of $\sqrt{T}\hat{\alpha}$ is

$$\Omega_{\hat{\alpha}} = M_\beta \Omega_{\bar{R}} M_\beta$$

where $\Omega_{\bar{R}}$ is the asymptotic covariance matrix of $\sqrt{T}(\bar{R}^e - ER^e)$. Given (21) and the assumptions made above:

$$\Omega_{\bar{R}} = \beta\Sigma_f\beta' + \Sigma$$

hence

$$\Omega_{\hat{\alpha}} = M_\beta \left(\beta\Sigma_f\beta' + \Sigma\right) M_\beta = M_\beta \Sigma M_\beta.$$

Since $\Omega_{\hat{\alpha}}$ has rank $n - k$, $C = T\hat{\alpha}'\Omega_{\hat{\alpha}}^{-1}\hat{\alpha}$ must be computed using a generalized inverse, and C is distributed χ^2 with $n - k$ degrees of freedom. Also, the asymptotic covariance matrix of $\sqrt{T}(\hat{\lambda} - \lambda)$ is

$$
\begin{aligned}
\Omega_{\hat{\lambda}} &= (\beta'\beta)^{-1}\beta'\Omega_{\bar{R}}\beta(\beta'\beta)^{-1} \\
&= \Sigma_f + (\beta'\beta)^{-1}\beta'\Sigma\beta(\beta'\beta)^{-1}.
\end{aligned}
$$

---

[5]Lustig and Verdelhan work within this framework, but these assumptions can be generalized.

With a Constant   When a constant is included in the second stage we have

$$\hat{\theta} = (X'X)^{-1}X'\bar{R}^e$$

where $\theta = (\ \gamma\ \ \lambda'\ )'$, $X = (\ \iota\ \ \beta\ )$ and ι is an $n \times 1$ vector of ones. Therefore,

$$\hat{u} = \bar{R}^e - X\hat{\theta} = \left[I - X(X'X)^{-1}X'\right]\bar{R}^e = M_X\bar{R}^e.$$

Given (9) we can write $E(R^e) = \gamma + \beta\lambda = X\theta$ where $\gamma = 0$. So

$$E(\hat{u}) = M_X E(\bar{R}^e) = M_X X'\theta = 0.$$

Also, the asymptotic covariance matrix of $\sqrt{T}\hat{u}$ is

$$\Omega_{\hat{u}} = M_X \Omega_{\bar{R}} M_X.$$

The term $\beta'\Sigma_f\beta$ can be written as $X\tilde{\Sigma}_f X'$ with

$$\tilde{\Sigma}_f = \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_f \end{pmatrix}.$$

Therefore we can rewrite $\Omega_{\bar{R}}$ as $X\tilde{\Sigma}_f X' + \Sigma$ so that:

$$\Omega_{\hat{u}} = M_X(X\tilde{\Sigma}_f X' + \Sigma)M_X = M_X\Sigma M_X.$$

Since $\Omega_{\hat{u}}$ has rank $n - k - 1$, $C = T\hat{u}'\Omega_{\hat{u}}^{-1}\hat{u}$ must be computed using a generalized inverse, and C is distributed χ^2 with $n - k - 1$ degrees of freedom. Also, the asymptotic covariance matrix of $\sqrt{T}(\hat{\theta} - \theta)$ is

$$\begin{aligned} \Omega_{\hat{\theta}} &= (X'X)^{-1}X'\Omega_{\bar{R}}X'(X'X)^{-1} \\ &= (X'X)^{-1}X'(X\tilde{\Sigma}_f X' + \Sigma)X'(X'X)^{-1} \\ &= \tilde{\Sigma}_f + (X'X)^{-1}X'\Sigma X'(X'X)^{-1}. \end{aligned}$$

As suggested in the text, the constant should really be considered part of the pricing error. As such, its significance could be tested alone, as it is the first element of $\hat{\theta}$. Alternatively one might also consider a reformulated χ^2 test based on

$$\hat{\alpha} = \bar{R}^e - \beta\hat{\lambda} = \hat{u} + \hat{\gamma}.$$

Letting

$$P = \begin{pmatrix} 0 & 0 \\ 0 & I_k \end{pmatrix}$$

we have

$$\hat{\alpha} = \bar{R}^e - XP\hat{\theta} = \left[I - XP(X'X)^{-1}X'\right]\bar{R}^e = H\bar{R}^e.$$

Therefore the asymptotic covariance matrix of $\sqrt{T}\hat{\alpha}$ is

$$\Omega_{\hat{\alpha}} = H(X\tilde{\Sigma}_f X' + \Sigma)H' = H\Sigma H'.$$

As in the other cases, this means that a test statistic can be formed as $C = T\hat{\alpha}'\Omega_{\hat{\alpha}}^{-1}\hat{\alpha}$. It will be distributed χ^2_{n-k} since $\Omega_{\hat{\alpha}}$ is of rank $n - k$ and must be computed using a generalized inverse.

7.1.2 Shanken Corrections (Betas are Estimated)

When the betas are unknown the first stage estimates, $\hat{\beta}_i$, are given by

$$\hat{\beta}_i = (\hat{f}'\hat{f})^{-1}\hat{f}'R_i^e$$

where R_i^e is a $T \times 1$ vector with elements R_{it}^e and $\hat{f}$ is a $T \times k$ matrix with rows equal to $(f_t - \bar{f})'$. Given the model, $R_i^e = a_i + f\beta_i + \epsilon_i$, where f is an $T \times k$ matrix with rows equal to f_t' and ϵ_i is a $T \times 1$ vector with elements ϵ_{it}. Hence

$$
\begin{aligned}
\hat{\beta}_i &= (\hat{f}'\hat{f})^{-1}\hat{f}'\left(a_i + f\beta_i + \epsilon_i\right) \\
&= \beta_i + (\hat{f}'\hat{f})^{-1}\hat{f}'\epsilon_i.
\end{aligned}
$$

Assuming that f_t and ϵ_t are independent, the asymptotic covariance between $\sqrt{T}(\hat{\beta}_i - \beta_i)$ and $\sqrt{T}(\hat{\beta}_j - \beta_j)$ is given by $\sigma_{ij}\Sigma_f^{-1}$ where σ_{ij} is the covariance between ϵ_{it} and ϵ_{jt}. If $\hat{\beta}$ is rearranged into a $nk \times 1$ *stacked vector*,

$$\hat{\beta}_v = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_n \end{pmatrix},$$

the asymptotic covariance matrix of $\sqrt{T}(\hat{\beta}_v - \beta_v)$ is $\Sigma \otimes \Sigma_f^{-1}$.

Without a Constant When the second stage excludes a constant $\hat{\lambda} = \hat{A}\bar{R}^e$, where $\hat{A} = (\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}'$. To work out the asymptotics we proceed as follows. Define

$$\bar{\lambda} = \lambda + \bar{f} - \mu. \tag{22}$$

The model implies that $ER^e = a + \beta\mu = \beta\lambda$. Hence we can write $a = \beta(\lambda - \mu)$. Substituting this result into (21) we get

$$\bar{R}^e = \beta(\lambda - \mu + \bar{f}) + \bar{\epsilon}.$$

Using (22) we have

$$
\begin{aligned}
\bar{R}^e &= \beta\bar{\lambda} + \bar{\epsilon} \\
&= \hat{\beta}\bar{\lambda} + \bar{\epsilon} - (\hat{\beta} - \beta)\bar{\lambda}. \tag{23}
\end{aligned}
$$

Premultiplying (23) by $\hat{A}$ we get

$$\hat{\lambda} = \bar{\lambda} + \hat{A}[\bar{\epsilon} - (\hat{\beta} - \beta)\bar{\lambda}],$$

so that

$$\hat{\lambda} - \bar{\lambda} = \hat{A}[\bar{\epsilon} - (\hat{\beta} - \beta)\bar{\lambda}]. \tag{24}$$

Now

$$\hat{\lambda} - \lambda = (\hat{\lambda} - \bar{\lambda}) + (\bar{\lambda} - \lambda)$$
$$= \hat{A}[\bar{\epsilon} - (\hat{\beta} - \beta)\bar{\lambda}] + (\bar{f} - \lambda).$$

The $\bar{f} - \lambda$ term in uncorrelated with the $\bar{\epsilon} - (\hat{\beta} - \beta)\bar{\lambda}$ following arguments in Shanken's (1992) Lemma 1. Also we can rewrite the term in brackets as $\bar{\epsilon} - (I_n \otimes \bar{\lambda}')(\hat{\beta}_v - \beta_v)$. Since plim $\bar{\lambda} = \lambda$, and plim $\hat{A} = A = (\beta'\beta)^{-1}\beta'$ this means that the asymptotic variance of $\sqrt{T}(\hat{\lambda} - \lambda)$ is

$$\Omega_{\hat{\lambda}} = A[\Sigma + (I_n \otimes \lambda')\left(\Sigma \otimes \Sigma_f^{-1}\right)(I_n \otimes \lambda)]A' + \Sigma_f.$$

Using the rules for Kronecker products this reduces to

$$\Omega_{\hat{\lambda}} = (1 + \lambda'\Sigma_f^{-1}\lambda)A\Sigma A' + \Sigma_f.$$

The pricing errors are

$$\hat{\alpha} = \bar{R}^e - \hat{\beta}\hat{\lambda} = \left[I - \hat{\beta}(\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}'\right]\bar{R}^e = M_{\hat{\beta}}\bar{R}^e$$
$$= M_{\hat{\beta}}\left[\hat{\beta}\bar{\lambda} + \bar{\epsilon} - (\hat{\beta} - \beta)\bar{\lambda}\right]$$
$$= M_{\hat{\beta}}\left[\bar{\epsilon} - (\hat{\beta} - \beta)\bar{\lambda}\right]$$

Hence the asymptotic covariance matrix of $\sqrt{T}\hat{\alpha}$ is

$$\Omega_{\hat{\alpha}} = (1 + \lambda'\Sigma_f^{-1}\lambda)M_\beta\Sigma M_\beta.$$

Since $\Omega_{\hat{\alpha}}$ has rank $n - k$, $C = T\hat{\alpha}'\Omega_{\hat{\alpha}}^{-1}\hat{\alpha}$ must be computed using a generalized inverse, and C is distributed χ^2 with $n - k$ degrees of freedom.

With a Constant When a constant is included in the second stage, but the betas are unknown, we have

$$\hat{\theta} = (\hat{X}'\hat{X})^{-1}\hat{X}'\bar{R}^e$$

where $\theta = (\ \gamma \quad \lambda' \ )'$, $\hat{X} = (\ \iota \quad \hat{\beta}' \ )$ and ι_n is an $n \times 1$ vector of ones. If $\hat{X}$ is rearranged into a $n(k+1) \times 1$ *stacked vector*,

$$\hat{X}_v = \begin{pmatrix} 1 \\ \hat{\beta}_1 \\ 1 \\ \hat{\beta}_2 \\ \vdots \\ 1 \\ \hat{\beta}_n \end{pmatrix},$$

the asymptotic covariance matrix of $\sqrt{T}(\hat{X}_v - X_v)$ is $\Sigma \otimes \Xi$ where

$$\Xi = \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_f^{-1} \end{pmatrix}.$$

We have $\hat{\theta} = \hat{A}\bar{R}^e$, where $\hat{A} = (\hat{X}'\hat{X})^{-1}\hat{X}'$. The model implies that $ER^e = a + \beta\mu = \gamma + \beta\lambda = \beta\lambda$ (since, under the null, $\gamma = 0$). Hence we can write $a = \beta(\lambda - \mu)$. Substituting this result into (21) we get

$$\bar{R}^e = \beta(\lambda - \mu + \bar{f}) + \bar{\epsilon}.$$

Defining $\bar{\theta} \equiv (\ 0 \quad \bar{\lambda}'\ )'$ we can then write this as

$$\begin{aligned}\bar{R}^e &= X\bar{\theta} + \bar{\epsilon} \\ &= \hat{X}\bar{\theta} + \bar{\epsilon} - (\hat{X} - X)\bar{\theta}.\end{aligned} \tag{25}$$

Premultiplying (25) by $\hat{A}$ we get

$$\hat{\theta} = \bar{\theta} + \hat{A}[\bar{\epsilon} - (\hat{X} - X)\bar{\theta}],$$

so that

$$\hat{\theta} - \bar{\theta} = \hat{A}[\bar{\epsilon} - (\hat{X} - X)\bar{\theta}].$$

Now

$$\begin{aligned}\hat{\theta} - \theta &= (\hat{\theta} - \bar{\theta}) + (\bar{\theta} - \theta) \\ &= \hat{A}[\bar{\epsilon} - (\hat{X} - X)\bar{\theta}] + \begin{pmatrix} 0 \\ \bar{f} - \lambda \end{pmatrix}.\end{aligned}$$

The $\bar{f} - \lambda$ term in uncorrelated with the $\bar{\epsilon} - (\hat{X} - X)\bar{\theta}$ term following arguments in Shanken's (1992) Lemma 1. And we can rewrite the term in brackets as $\bar{\epsilon} - (I_n \otimes \bar{\theta}')(\hat{X}_v - X_v)$. Since $\text{plim}\,\bar{\theta} = \theta$ and $\text{plim}\,\hat{A} = A = (X'X)^{-1}X'$ this means that the asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta)$ is

$$\Omega_{\hat{\theta}} = A[\Sigma + (I_n \otimes \theta')(\Sigma \otimes \Xi)(I_n \otimes \theta)]A' + \tilde{\Sigma}_f.$$

Using the rules for Kronecker products this reduces to

$$\Omega_{\hat{\theta}} = (1 + \theta'\Xi\theta)A\Sigma A' + \tilde{\Sigma}_f,$$

but because of the form of Ξ it can also be written as

$$\Omega_{\hat{\theta}} = (1 + \lambda'\Sigma_f^{-1}\lambda)A\Sigma A' + \tilde{\Sigma}_f.$$

The pricing errors are

$$\begin{aligned}\hat{u} &= \bar{R}^e - \hat{X}\hat{\theta} = \left[I - \hat{X}(\hat{X}'\hat{X})^{-1}\hat{X}'\right]\bar{R}^e = M_{\hat{X}}\bar{R}^e \\ &= M_{\hat{X}}\left[\hat{X}\bar{\theta} + \bar{\epsilon} - (\hat{X} - X)\bar{\theta}\right] \\ &= M_{\hat{X}}\left[\bar{\epsilon} - (\hat{X} - X)\bar{\theta}\right]\end{aligned}$$

Hence the asymptotic covariance matrix of $\sqrt{T}\hat{u}$ is

$$\Omega_{\hat{u}} = (1 + \lambda'\Sigma_f^{-1}\lambda)M_X\Sigma M_X.$$

Since $\Omega_{\hat{u}}$ has rank $n - k - 1$, $C = T\hat{u}'\Omega_{\hat{u}}^{-1}\hat{u}$ must be computed using a generalized inverse, and C is distributed χ^2 with $n - k - 1$ degrees of freedom.

As suggested in the text, the constant should really be considered part of the pricing error. As such, its significance could be tested alone, as it is the first element of $\hat{\theta}$. Alternatively one might also consider a reformulated χ^2 test based on

$$\hat{\alpha} = \bar{R}^e - \hat{\beta}\hat{\lambda} = \hat{u} + \hat{\gamma}.$$

Letting

$$P = \begin{pmatrix} 0 & 0 \\ 0 & I_k \end{pmatrix}$$

we have

$$\begin{aligned}
\hat{\alpha} &= \bar{R}^e - \hat{X}P\hat{\theta} = \left[I - \hat{X}P(\hat{X}'\hat{X})^{-1}\hat{X}' \right] \bar{R}^e = \hat{H}\bar{R}^e \\
&= \hat{H}\left[\hat{X}\bar{\theta} + \bar{\epsilon} - (\hat{X} - X)\bar{\theta} \right] \\
&= \hat{H}\left[\bar{\epsilon} - (\hat{X} - X)\bar{\theta} \right].
\end{aligned}$$

Therefore the asymptotic covariance matrix of $\sqrt{T}\hat{\alpha}$ is

$$\Omega_{\hat{\alpha}} = (1 + \lambda'\Sigma_f^{-1}\lambda)H\Sigma H'.$$

As in the other cases, this means that a test statistic can be formed as $C = T\hat{\alpha}'\Omega_{\hat{\alpha}}^{-1}\hat{\alpha}$. It will be distributed χ^2_{n-k} since $\Omega_{\hat{\alpha}}$ is of rank $n - k$ and the covariance matrix must be inverted using a generalized inverse.

7.1.3 GMM Standard Errors (Betas are Estimated)

Without a Constant The model is estimated by exploiting the moment restrictions $E(R^e_{it} - a_i - \beta'_i f_t) = 0$, $E[(R^e_{it} - a_i - \beta'_i f_t)f'_t] = 0$, and $E(R^e_{it} - \beta'_i \lambda) = 0$, $i = 1, \ldots, n$. Let $\tilde{f}_t = (\ 1\ \ f'_t\ )'$, $\tilde{\beta}_i = (\ a_i\ \ \beta'_i\ )'$ and

$$\theta = \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \vdots \\ \tilde{\beta}_n \\ \lambda \end{pmatrix}.$$

Define the $n(k + 2) \times 1$ vector

$$u_t(\theta) = \begin{pmatrix} \tilde{f}_t(R^e_{1t} - \tilde{f}'_t\tilde{\beta}_1) \\ \tilde{f}_t(R^e_{2t} - \tilde{f}'_t\tilde{\beta}_2) \\ \cdots \\ \tilde{f}_t(R^e_{nt} - \tilde{f}'_t\tilde{\beta}_n) \\ R^e_t - \beta\lambda \end{pmatrix},$$

the $n(k+2) \times 1$ vector $g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} u_t(\theta)$, and the $[n(k+1)+k] \times [n(k+2) \times 1]$ matrix

$$a_T = \begin{pmatrix} I_{n(k+1)} & 0 \\ 0 & \hat{\beta}'_{OLS} \end{pmatrix}.$$

The GMM estimator that sets $a_T g_T = 0$ reproduces the two-pass estimates of a, β, and λ. Define

$$d_T = \frac{\partial g_T(\theta)}{\partial \theta'} = \begin{pmatrix} -I_n \otimes M_{\tilde{f}} & 0_{n(k+1)\times k} \\ -I_n \otimes \begin{pmatrix} 0 & \hat{\lambda}' \end{pmatrix} & -\hat{\beta}_{OLS} \end{pmatrix}$$

where $M_{\tilde{f}} = \frac{1}{T}\sum_{t=1}^{T} \tilde{f}_t \tilde{f}'_t$.

Let $a = \operatorname{plim} a_T$ and $d = \operatorname{plim} d_T$. The covariance matrix of $\sqrt{T}(\hat{\theta} - \theta)$ is

$$V_\theta = (ad)^{-1} a S a' \left[(ad)^{-1}\right]'$$

and the covariance matrix of $\sqrt{T} g_T(\hat{\theta})$ is

$$V_g = [I - d(ad)^{-1}a] S [I - d(ad)^{-1}a]'$$

where S is the asymptotic covariance matrix of $\sqrt{T} g_T(\theta)$. These results follow from the facts that $\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} (ad)^{-1}a\sqrt{T} g_T(\theta)$ and $\sqrt{T} g_T(\hat{\theta}) \xrightarrow{d} [I - d(ad)^{-1}a]\sqrt{T} g_T(\theta)$. The test statistic for the pricing errors is just $T g_T(\hat{\theta})' V_g^{-1} g_T(\hat{\theta})$, where the inverse is generalized. Since $S = \sum_{j=-\infty}^{\infty} E(u_t u'_{t-j})$, I use a variant of a VARHAC estimator for S: due to limited sample size I only allow lags of an error to enter into the VAR equation for that error.

With a Constant The model is estimated by exploiting the moment restrictions $E(R^e_{it} - a_i - \beta'_i f_t) = 0$, $E[(R^e_{it} - a_i - \beta'_i f_t)f'_t] = 0$, and $E(R^e_{it} - \gamma - \beta'_i \lambda) = 0$, $i = 1, \ldots, n$. Now let

$$\theta = \begin{pmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \\ \vdots \\ \tilde{\beta}_n \\ \gamma \\ \lambda \end{pmatrix}.$$

Define the $n(k+2) \times 1$ vector

$$u_t(\theta) = \begin{pmatrix} \tilde{f}_t(R^e_{1t} - \tilde{f}'_t \tilde{\beta}_1) \\ \tilde{f}_t(R^e_{2t} - \tilde{f}'_t \tilde{\beta}_2) \\ \cdots \\ \tilde{f}_t(R^e_{nt} - \tilde{f}'_t \tilde{\beta}_n) \\ R^e_t - \gamma - \beta\lambda \end{pmatrix},$$

the $n(k+2) \times 1$ vector $g_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} u_t(\theta)$, and the $(n+1)(k+1) \times [n(k+2) \times 1]$ matrix

$$a_T = \begin{pmatrix} I_{n(k+1)} & 0 \\ 0 & \hat{X}' \end{pmatrix}.$$

where $\hat{X} = (\ \iota_{n\times 1}\ \ \hat{\beta}_{OLS}\ )$. The GMM estimator that sets $a_T g_T = 0$ reproduces the two-pass estimates of a, β, γ and λ. Define

$$d_T = \frac{\partial g_T(\theta)}{\partial \theta'} = \begin{pmatrix} -I_n \otimes M_{\tilde{f}} & 0_{n(k+1)\times(k+1)} \\ -I_n \otimes \begin{pmatrix} 0 & \hat{\lambda}' \end{pmatrix} & -\hat{X} \end{pmatrix}.$$

Let $a = \operatorname{plim} a_T$ and $d = \operatorname{plim} d_T$. The covariance matrix of $\sqrt{T}(\hat{\theta} - \theta)$ is

$$V_\theta = (ad)^{-1} a S a' \left[(ad)^{-1}\right]'$$

and the covariance matrix of $\sqrt{T} g_T(\hat{\theta})$ is

$$V_g = [I - d(ad)^{-1}a] S [I - d(ad)^{-1}a]'$$

where S is the asymptotic covariance matrix of $\sqrt{T} g_T(\theta)$. These results follow from the facts that $\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} (ad)^{-1} a \sqrt{T} g_T(\theta)$ and $\sqrt{T} g_T(\hat{\theta}) \xrightarrow{d} [I - d(ad)^{-1}a] \sqrt{T} g_T(\theta)$. The test statistic for the pricing errors is just $T g_T(\hat{\theta})' V_g^{-1} g_T(\hat{\theta})$, where the inverse is generalized. A test of the pricing errors inclusive of the contest can be derived from the joint distribution of $\sqrt{T}(\hat{\theta} - \theta)$ and $\sqrt{T} g_T(\hat{\theta})$. Since $S = \sum_{j=-\infty}^{\infty} E(u_t u'_{t-j})$, I use a variant of a VARHAC estimator for S. Due to limited sample size I only allow lags of an error to enter into the VAR equation for that error.

7.2 Direct GMM Estimation of the Model

7.2.1 Model without a Constant

Asympotic Theory Let

$$u_{1t}(b, \mu) = R_t^e[1 - (f_t - \mu)'b] \tag{26}$$

$$g_{1T}(b, \mu) = T^{-1} \sum_{t=1}^{T} u_{1t}(b, \mu) = \bar{R}^e(1 + \mu'b) - D_T b. \tag{27}$$

where $D_T = T^{-1} \sum_{t=1}^{T} R_t^e f_t'$ and $\bar{R}^e = T^{-1} \sum_{t=1}^{T} R_t^e$. Also define

$$u_{2t}(\mu) = f_t - \mu \tag{28}$$

$$g_{2T}(\mu) = T^{-1} \sum_{t=1}^{T} u_{2t}(\mu) = \bar{f} - \mu. \tag{29}$$

Finally, define the stacked vectors

$$u_t(b, \mu) = \begin{pmatrix} u_{1t}(b, \mu) \\ u_{2t}(\mu) \end{pmatrix} \qquad g_T(b, \mu) = \begin{pmatrix} g_{1T}(b, \mu) \\ g_{2T}(\mu) \end{pmatrix}$$

and the matrix

$$S = E\left[\sum_{j=-\infty}^{\infty} u_t(b_0, \mu_0) u_{t-j}(b_0, \mu_0)'\right].$$

The parameters b and μ are estimated by setting $a_T g_T = 0$, where

$$a_T = \begin{pmatrix} (D_T - \bar{R}^e \mu')' W_T & 0 \\ 0 & I_k \end{pmatrix},$$

and W_T is some weighting matrix. Given the definition of g_T this means the GMM estimator is the solution to

$$\begin{pmatrix} (D_T - \bar{R}^e \mu')' W_T & 0 \\ 0 & I_k \end{pmatrix} \begin{pmatrix} \bar{R}^e(1 + \mu'b) - D_T b \\ \bar{f} - \mu \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{30}$$

implying that

$$\hat{\mu} = \bar{f} \tag{31}$$

$$\hat{b} = (d_T' W_T d_T)^{-1} d_T' W_T \bar{R}^e, \tag{32}$$

where $d_T = D_T - \bar{R}^e \bar{f}'$.

In the first stage the weighting matrix is $W_T = I_n$. In the second stage, Lustig and Verdelhan follow Cochrane (2005) and set

$$W_T = \left[T^{-1} \sum_{t=1}^{T} u_{1t}(\hat{b}, \hat{\mu}) u_{1t}(\hat{b}, \hat{\mu})' \right]^{-1} \tag{33}$$

where $\hat{b} = (d_T' d_T)^{-1} d_T' \bar{R}^e$ and $\hat{\mu} = \bar{f}$ are the first stage estimates of the parameters. In this case

$$\text{plim } W_T = W = S_{11}^{-1} \text{ where } S_{11} = E[u_{1t}(b_0, \mu_0) u_{1t}(b_0, \mu_0)'].$$

Given (32), $\text{plim } \hat{b} = b_0$. This follows from the fact that $\text{plim } d_T = d \equiv E[R_t^e(f_t - \mu)']$ and that $\text{plim } \bar{R}^e = E(R^e)$. We then get $\text{plim } \hat{b} = (d'Wd)^{-1} d'W E(R^e)$. The model implies that $E(R^e) = db_0$. Hence $\text{plim } \hat{b} = b_0$. So the first and second stage estimates of b are obviously consistent.

The derivation of the asymptotic distribution of $(\hat{b}, \hat{\mu})$ relies on deriving the distance between $g_T(\hat{b}, \hat{\mu})$ and $g_T(b_0, \mu_0)$. Using (27) and the consistency of $\hat{b}$ and $\hat{\mu}$ we can argue that there is a pair $(\bar{b}, \bar{\mu})$ between (b_0, μ_0) and $(\hat{b}, \hat{\mu})$ such that

$$g_{1T}(\hat{b}, \hat{\mu}) = g_{1T}(b_0, \mu_0) + \left(\bar{R}^e \bar{\mu}' - D_T \right)(\hat{b} - b_0) + \bar{R}^e \bar{b}'(\hat{\mu} - \mu_0). \tag{34}$$

From (29) we also have

$$g_{2T}(\hat{\mu}) = g_{2T}(\mu_0) - (\hat{\mu} - \mu_0). \tag{35}$$

Premultiplying (34) by $d_T' W_T$ one obtains

$$0 = d_T' W_T g_{1T}(\hat{b}, \hat{\mu}) = d_T' W_T[g_{1T}(b_0, \mu_0) + \left(\bar{R}^e \bar{\mu}' - D_T \right)(\hat{b} - b_0) + \bar{R}^e \bar{b}'(\hat{\mu} - \mu_0)] \tag{36}$$

We can rewrite (35) and (36) together as

$$0 = \hat{a}_T \left[g_T(b_0, \mu_0) - \Delta_T \begin{pmatrix} \hat{b} - b_0 \\ \hat{\mu} - \mu_0 \end{pmatrix} \right]. \tag{37}$$

33

where

$$\hat{a}_T = \begin{pmatrix} d_T' W_T & 0 \\ 0 & I_k \end{pmatrix} \quad \Delta_T = \begin{pmatrix} (D_T - \bar{R}^e \bar{\mu}') & -\bar{R}^e \bar{b}' \\ 0 & I_k \end{pmatrix}.$$

We have plim $\hat{a}_T = a$ and plim $\Delta_T = \Delta$ where

$$a = \begin{pmatrix} d'W & 0 \\ 0 & I_k \end{pmatrix} \quad \Delta = \begin{pmatrix} d & -db_0 b_0' \\ 0 & I_k \end{pmatrix},$$

and I have used the fact that plim $\bar{R}^e = E(R^e) = db_0$. Hence

$$\sqrt{T} \begin{pmatrix} \hat{b} - b_0 \\ \hat{\mu} - \mu_0 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} (d'Wd)^{-1} & b_0 b_0' \\ 0 & I_k \end{pmatrix} \begin{pmatrix} d'W & 0 \\ 0 & I_k \end{pmatrix} \sqrt{T} g_T(b_0, \mu_0).$$

Thus we have

$$\sqrt{T}(\hat{b} - b_0) \quad \xrightarrow{d} \left((d'Wd)^{-1} d'W \quad b_0 b_0' \right) \sqrt{T} g_T(b_0, \mu_0) = B \sqrt{T} g_T(b_0, \mu_0)$$

$$\sqrt{T}(\hat{\mu} - \mu_0) \quad \xrightarrow{d} \left(0 \quad I_k \right) \sqrt{T} g_T(b_0, \mu_0) = \sqrt{T} g_{2T}(b_0, \mu_0)$$

and the asymptotic covariance matrix of $\sqrt{T}(\hat{b} - b_0)$ is

$$V_b = BSB'. \tag{38}$$

The fact that μ is estimated affects V_b. If μ was known the covariance matrix would reduce to $(d'Wd)^{-1} d'W S_{11} Wd (d'Wd)^{-1}$.

To get a test of the pricing errors, Cochrane (2005) follows Hansen (1982) in showing that the asymptotic distribution of $\sqrt{T} g_T(\hat{b}, \hat{\mu})$ is normal with covariance matrix

$$[I - \Delta(a\Delta)^{-1} a] S [I - a'(a\Delta)^{-1} \Delta'].$$

Some algebra shows that this implies that $\sqrt{T} g_{1T}(\hat{b}, \hat{\mu})$ is normal with covariance matrix

$$V_0 = [I - d (d'Wd)^{-1} d'W] S_{11} [I - Wd (d'Wd)^{-1} d'].$$

This is the same expression as one obtains when μ is known. Since the test of the pricing errors is obtained as $T g_{1T}(\hat{b}, \hat{\mu})' V_T^{-1} g_{1T}(\hat{b}, \hat{\mu})$, where the inverse is generalized, and V_T is a consistent estimate of V_0, the fact that μ is estimated has no effect on the statistic as compared to the case where μ is treated as known.

Factor Risk Premia The GMM estimator produces estimates of b and μ. To obtain an estimate of λ we can use the expression $\lambda = \Sigma_f b$. This requires estimation of Σ_f. This can be done by adding moment conditions that identify the unique elements of Σ_f:

$$E[(f_{it} - \mu_i) (f_{jt} - \mu_j) - \Sigma_{f,ij}], \quad i = 1, \ldots, k, \quad j = i, \ldots, k. \tag{39}$$

The estimate $\hat{\Sigma}_f$ then corresponds to the sample covariance matrix of f_t. Of course, standard errors for $\hat{\lambda}$ should take into account estimation of Σ_f.

Equivalence Between the First Stage of GMM and the Two-Pass Procedure The
first stage estimate of b based on $W = I_n$ is $\hat{b} = (d'_T d_T)^{-1} d'_T \bar{R}^e$. The matrix d_T is the sample
covariance between R_t and f_t. Hence $\hat{\beta} = d_T \hat{\Sigma}_f^{-1}$, where $\hat{\Sigma}_f$ is the sample covariance of the
risk factors. Therefore $d_T = \hat{\beta}\hat{\Sigma}_f$ and $\hat{b} = (\hat{\Sigma}_f \hat{\beta}' \hat{\beta} \hat{\Sigma}_f)^{-1} \hat{\Sigma}_f \hat{\beta}' \bar{R}^e = \hat{\Sigma}_f^{-1} \hat{\lambda}$. Since $\hat{\lambda}_{GMM} \equiv \hat{\Sigma}_f \hat{b}$,
$\hat{\lambda}_{GMM} = \hat{\lambda}$ from the two-pass procedure.

VARHAC Spectral Density Matrix Since $S = \sum_{j=-\infty}^{\infty} E(u_t u'_{t-j})$, I estimate it as
follows. Define u_{1t} and u_{2t} as in (26) and (28). I use a VARHAC estimator for S, imposing
the restriction that $E u_{1t} u'_{t-j} = 0$ for $j \geq 1$. This means that the VARHAC estimator for
S_{11}, the sub-block of S equal to $\sum_{j=-\infty}^{\infty} E(u_{1t} u'_{1t-j})$, is the same as the HAC estimator for
S_{11}. But this is not true for the S_{12}, S_{21} and S_{22} sub-blocks. In practice, the VARHAC
procedure typically finds persistence in some elements of u_{2t} because these are the GMM
errors corresponding to $f_t - \hat{\mu}$. Since some of the risk factors are persistent it is important
to allow for this possibility, which is not ruled out by theory.

Equivalence of the Pricing Error Test at the First and Second Stages of GMM
At the first stage of GMM we have

$$\hat{b}_1 = (d'_T d_T)^{-1} d'_T \bar{R}^e$$

so the pricing errors are $\hat{\alpha}_1 = \bar{R}^e - d_T \hat{b}_1 = M_d \bar{R}^e$ where $M_d = I - d_T (d'_T d_T)^{-1} d'_T$. The
estimated covariance matrix of $\hat{\alpha}_1$ is $V_T = M_d \hat{S}_{11} M_d$, so the test statistic is

$$T \left(\bar{R}^e \right)' M_d (M_d \hat{S}_{11} M_d)^{-1} M_d \bar{R}^e$$

where the inverse is generalized.

At the second stage of GMM we have

$$\hat{b}_2 = (d'_T W_T d_T)^{-1} d'_T W_T \bar{R}^e$$

so the pricing errors are $\hat{\alpha}_2 = \bar{R}^e - d_T \hat{b}_2 = M_W \bar{R}^e$ where $M_W = [I - d_T (d'_T W_T d_T)^{-1} d'_T W_T] \bar{R}^e$.
The estimated covariance matrix of $\hat{\alpha}_2$ is $V_T = M_W \hat{S}_{11} M'_W$ so the test statistic is

$$T \left(\bar{R}^e \right)' M'_W (M_W \hat{S}_{11} M'_W)^{-1} M_W \bar{R}^e.$$

Because $M_d (M_d \hat{S}_{11} M_d)^{-1} M_d = M'_W (M_W \hat{S}_{11} M'_W)^{-1} M_W$, when $W = \hat{S}_{11}^{-1}$, the two statistics
are the same.

7.2.2 Model with a Constant

Asympotic Theory Let

$$
\begin{align}
u_{1t}(b, \mu, \gamma) &= \bar{R}_t^e[1 - (f_t - \mu)'b] - \gamma \tag{40} \\
g_{1T}(b, \mu, \gamma) &= T^{-1} \sum_{t=1}^{T} u_{1t}(b, \mu, \gamma) = \bar{R}^e(1 + \bar{\mu}'\bar{b}) - \bar{D}_T \bar{b}. \tag{41}
\end{align}
$$

where $\bar{b} = (\ \gamma\quad b'\ )'$, $\bar{\mu} = (\ 0\quad \mu'\ )'$ and $\bar{D}_T = (\ \iota_{n\times 1}\quad D_T\ )$.

Define u_{2t} and g_{2T} as in (28) and (29). Define the stacked vectors

$$u_t(b, \mu, \gamma) = \begin{pmatrix} u_{1t}(b, \mu, \gamma) \\ u_{2t}(\mu) \end{pmatrix} \qquad g_T(b, \mu, \gamma) = \begin{pmatrix} g_{1T}(b, \mu, \gamma) \\ g_{2T}(\mu) \end{pmatrix}$$

and the matrix

$$S = E\Big[\sum_{j=-\infty}^{\infty} u_t(b_0, \mu_0, \gamma_0) u_{t-j}(b_0, \mu_0, \gamma_0)'\Big].$$

The parameters $\bar{b}$ and μ are estimated by setting $a_T g_T = 0$, where

$$a_T = \begin{pmatrix} (\bar{D}_T - \bar{R}^e\bar{\mu}')'W_T & 0 \\ 0 & I_k \end{pmatrix},$$

and W_T is some weighting matrix. Given the definition of g_T this means the GMM estimator is the solution to

$$\begin{pmatrix} (\bar{D}_T - \bar{R}^e\bar{\mu}')'W_T & 0 \\ 0 & I_k \end{pmatrix}\begin{pmatrix} \bar{R}^e(1 + \bar{\mu}'\bar{b}) - \bar{D}_T\bar{b} \\ \bar{f} - \mu \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{42}$$

implying that

$$\hat{\mu} = \bar{f} \tag{43}$$
$$\widehat{\bar{b}} = (\bar{d}_T'W_T\bar{d}_T)^{-1}\bar{d}_T'W_T\bar{R}^e, \tag{44}$$

where $\bar{d}_T = \bar{D}_T - \bar{R}^e\widehat{\bar{\mu}}' = (\ \iota_{n\times 1}\quad D_T\ ) - \bar{R}(\ 0\quad \bar{f}'\ ) = (\ \iota_{n\times 1}\quad d_T\ )$.

The first and second stage estimates are calculated as in the case with the constant. In the first stage $W_T = I_n$. In the second stage, W_T is the inverse of a consistent estimator for $S_{11} = E[u_{1t}(b_0, \mu_0, \gamma_0)u_{1t}(b_0, \mu_0, \gamma_0)']$.

Equivalence Between the First Stage of GMM and the Two-Pass Procedure
The first stage estimate of $\bar{b}$ based on $W = I_n$ is $\widehat{\bar{b}} = (\bar{d}_T'\bar{d}_T)^{-1}\bar{d}_T'\bar{R}^e$. The matrix $\bar{d}_T = (\ \iota_{n\times 1}\quad d_T\ )$, which can be rewritten as $\bar{d}_T = (\ \iota_{n\times 1}\quad \hat{\beta}\hat{\Sigma}_f\ )$. Hence

$$\begin{aligned}
\widehat{\bar{b}} &= \begin{pmatrix} \iota'\iota & \iota'\hat{\beta}\hat{\Sigma}_f \\ \hat{\Sigma}_f\hat{\beta}'\iota & \hat{\Sigma}_f\hat{\beta}'\hat{\beta}\hat{\Sigma}_f \end{pmatrix}^{=1}\begin{pmatrix} \iota'\bar{R}^e \\ \hat{\Sigma}_f\hat{\beta}'\bar{R}^e \end{pmatrix} \\
&= \Big[\begin{pmatrix} 1 & 0 \\ 0 & \hat{\Sigma}_f \end{pmatrix}\begin{pmatrix} \iota'\iota & \iota'\hat{\beta} \\ \hat{\beta}'\iota & \hat{\beta}'\hat{\beta} \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & \hat{\Sigma}_f \end{pmatrix}\Big]^{-1}\begin{pmatrix} \iota'\bar{R}^e \\ \hat{\Sigma}_f\hat{\beta}'\bar{R}^e \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & \hat{\Sigma}_f^{-1} \end{pmatrix}\begin{pmatrix} \iota'\iota & \iota'\hat{\beta} \\ \hat{\beta}'\iota & \hat{\beta}'\hat{\beta} \end{pmatrix}^{-1}\begin{pmatrix} \iota'\bar{R}^e \\ \hat{\beta}'\bar{R}^e \end{pmatrix}.
\end{aligned}$$

The two-step estimator of γ and λ is

$$\begin{pmatrix} \hat{\gamma} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} \iota'\iota & \iota'\hat{\beta} \\ \hat{\beta}'\iota & \hat{\beta}'\hat{\beta} \end{pmatrix}^{-1}\begin{pmatrix} \iota'\bar{R}^e \\ \hat{\beta}'\bar{R}^e \end{pmatrix}$$

Hence

$$\widehat{b} = \begin{pmatrix} 1 & 0 \\ 0 & \hat{\Sigma}_f^{-1} \end{pmatrix} \begin{pmatrix} \hat{\gamma} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} \hat{\gamma} \\ \hat{\Sigma}_f^{-1}\hat{\lambda} \end{pmatrix}.$$

So the GMM estimator of γ is identical to the two-step estimator of γ. Also $\hat{\lambda}_{GMM} \equiv \hat{\Sigma}_f \hat{b}$, $\hat{\lambda}_{GMM} = \hat{\lambda}$ from the two-pass procedure.

7.3 Accounting for Differences in Results

Some of my results differ from Lustig and Verdelhan's. I discuss each of these differences in turn and in the order they appear in the tables.

7.3.1 Table 3(a), OLS Case

There is only one difference here. The p-value on the test statistic for the pricing errors when the constant is included in the model's predicted mean returns. In their Table 14, Lustig and Verdelhan report a p-value of 0.628, while I report a p-value of 0.483. The test statistic is 3.4666. When a constant is included in the model, the covariance matrix of the error vector has rank $n - k - 1 = 4$. The p-value for a statistic of 3.4666, with 4 degrees of freedom is 0.483. If one incorrectly uses the $n - k = 5$ as the degrees of freedom for the test, one obtains Lustig and Verdelhan's p-value, 0.628.

7.3.2 Table 3(a), Shanken Case

The numbers here that directly correspond to numbers given in Lustig and Verdelhan's paper are the Shanken standard errors for the model with a constant, which they report in their Table 14. For consumption growth, durables growth and the market return, I report standard errors of 2.11, 2.42 and 18.8. They report slightly larger standard errors: 2.15, 2.52 and 19.8. I believe that this may be due to them using the formula $[1+(\lambda'\Sigma_f^{-1}\lambda)](A\Sigma A'+\tilde{\Sigma}_f)$ in computing the standard errors instead of $[1 + (\lambda'\Sigma_f^{-1}\lambda)]A\Sigma A' + \tilde{\Sigma}_f$ (the meaning of these expressions is explained in a previous section of the appendix). When I use the incorrect formula, I reproduce their standard errors to within one decimal place.

7.3.3 Mapping from b to the Structural Parameters

As reported in the main text the structural parameters in the model map to the bs according to

$$b_1 = \kappa(1/\sigma + \alpha(1/\rho - 1/\sigma)) \tag{45}$$
$$b_2 = \kappa\alpha(1/\sigma - 1/\rho) \tag{46}$$
$$b_3 = 1 - \kappa \tag{47}$$

where $\kappa = (1 - \gamma)/(1 - 1/\sigma)$. This corresponds to equation (19) is Yogo (2006).

Given estimates of the bs, Yogo's approach is to set a value for ρ and then solve the three equations above for α, γ and σ. The solutions Yogo (2006) states in his paper near the bottom of page 557 are

$$\sigma = \frac{1 - b_3}{b_1 + b_3} \tag{48}$$

$$\gamma = b_1 + b_2 + b_3 \tag{49}$$

$$\alpha = \frac{b_2}{b_1 + b_2 + (b_3 - 1)/\rho}. \tag{50}$$

The expression for σ (48) is wrong and should be:

$$\sigma = \frac{1 - b_3}{b_1 + b_2} \tag{51}$$

With $b_c = -21.0$, $b_d = 130$ and $b_r = 4.46$ I obtain $\sigma = -0.032$ using (51). Using the incorrect formula in (48) gives $\sigma = 0.21$, as in Lustig and Verdelhan's paper. This error does not affect values of structural parameters given in Yogo (2006), as the error appears to only be in the text, not in calculations.

7.3.4   Table 4(b)

Table 4(b) presents estimates obtained using GMM. Lustig and Verdelhan's second stage results (which are presented in their Table 14). The point estimates in my table are the same as theirs, but the HAC standard errors are slightly different than theirs. I believe that this difference may be due to them using an incorrect expression for the standard errors which ignores the sampling uncertainty due to μ being estimated.

When μ_0 is known, the expression in (38) is simpler, and reduces to

$$V_k = (d'Wd)^{-1} d'W S_{11} W d (d'Wd)^{-1}. \tag{52}$$

In the second stage of GMM $W_0 = S_{11}^{-1}$ so (52) reduces to

$$V_k = \left(d' S_{11}^{-1} d\right)^{-1}. \tag{53}$$

I believe that Lustig and Verdelhan base their GMM standard errors on (53). However, this is inappropriate when μ_0 must be estimated. This is because V_b, given in (38), does not reduce to V_k unless $b_0 = 0$ or μ_0 is known. This problem does not bias the standard errors sharply in a consistent direction, and the differences it induces are small.