The Forward Premium is Still a Puzzle
Appendix

Craig Burnside*

June 2011

*Duke University and NBER.

1 Results Alluded to in the Main Text

1.1 First-Pass Estimates of Betas

Table 1 provides first-pass estimates of the betas for the base case discussed in Section 2 of the main text. The first pass is a time series regression of each portfolio's excess return on the vector of risk factors:

$$R_{it}^e = a_i + \mathbf{f}_t'\boldsymbol{\beta}_i + \epsilon_{it}, \quad t = 1, \ldots, T, \text{ for each } i = 1, \ldots, n. \tag{1}$$

Here $\boldsymbol{\beta}_i'$ represents the ith row in $\boldsymbol{\beta}$. The system of equations represented by (1) is estimated using equation-by-equation OLS.

In Section 2 I also mention that tests of the SDF covariances lead to the same conclusions as tests of the SDF betas. The p-values associated with the the null hypothesis that $\text{cov}(R_{it}^e, m_t) = c$ for all i are 0.94, 0.91, and 0.89 for SDF models (i), (ii) and (iii), respectively. The p-values associated with the the null hypothesis that $\text{cov}(R_{it}^e, m_t) = 0$ for all i are 0.81, 0.70, and 0.70 for SDF models (i), (ii) and (iii), respectively.

Table 2 provides first-pass estimates of the betas for the case discussed in Section 4.3 of the main text.

1.2 Weak Identification

In this section I argue that tests of the pricing errors fail to reject LV's model at low levels of significance due to an identification problem. In the second-pass regression with the constant, the parameters γ and $\boldsymbol{\lambda}$ are identified under the assumption that $\boldsymbol{\beta}^+ = \begin{pmatrix} \boldsymbol{\iota} & \boldsymbol{\beta} \end{pmatrix}$ has full column rank, where $\boldsymbol{\iota}$ is an $n \times 1$ vector of ones. In the second-pass regression with no constant, $\boldsymbol{\lambda}$ is identified if $\boldsymbol{\beta}$ has full column rank. The same conditions must hold for the GMM procedure used to estimate $\mathbf{b}$. When the rank conditions fail, conventional inference drawn from second pass regressions and GMM is unreliable because standard asymptotic theory does not apply. As Burnside (2010) discusses, t-statistics for $\hat{\boldsymbol{\lambda}}$ and $\hat{\mathbf{b}}$ have non-standard distributions, and, most importantly, pricing-error tests cannot reliably detect model misspecification.

It is straightforward to test whether LV's model is identified using a rank test from Jonathan H. Wright (2003). Table 3 presents tests of the null hypothesis that $\boldsymbol{\beta}$ has reduced rank. Since there are three risk factors in the model, $\boldsymbol{\beta}$ has reduced rank if $\text{rank}(\boldsymbol{\beta}) < 3$. As Table 3 indicates, it is not possible to reject that $\text{rank}(\boldsymbol{\beta}) = 1$ at conventional significance levels. In fact, it is only when VARHAC standard errors are used that we can reject the null hypothesis that rank($\boldsymbol{\beta}$) = 0, which is equivalent to the null that *every* element of $\boldsymbol{\beta}$ is zero. Similar results are obtained when I test the rank of $\boldsymbol{\beta}^+$.

A standard tool for conducting inference under weak identification is to construct confidence sets for weakly identified parameters using the objective function corresponding to the continuously updated (CU) GMM estimator.[1] To put this method into practice, I construct an m_t series using the following values of $\mathbf{b}$: (i) the vector corresponding to LV's two pass estimates of $\boldsymbol{\lambda}$: $b_c = -21$, $b_d = 130$ and $b_r = 4.5$, (ii) the vector corresponding to LV's GMM estimates of $\mathbf{b}$: $b_c = 37$, $b_d = 75$ and $b_r = 4.7$, (iii) the vector corresponding to the calibrated model discussed in section I.E of LV's paper: $b_c = 6.7$, $b_d = 23$ and $b_r = 0.31$. I then treat $-m_t$ as a risk factor, and construct a robust confidence set for λ_m, the price of SDF risk, using the CU-GMM objective function corresponding to the two-pass regression method (with no constant included in the regression). The reason for treating $-m_t$ as a risk factor is that this reduces the dimensionality of the confidence set to one parameter. The objective function is evaluated at each possible value of λ_m, and the corresponding p-value is calculated. Those values of λ_m for which the p-value exceeds 0.05 lie in the 95 percent confidence set for λ_m. The difference between this approach and standard GMM is that the weighting matrix is recalculated at each λ_m, and no degrees of freedom are lost because λ_m is not estimated. The resulting confidence sets are the entire real line in cases (i) and (iii). In case (ii) the confidence set is $(-\infty, 0.75) \cup (0.98, +\infty)$.[2] While the rank tests are the definitive indication that the data are uninformative about $\boldsymbol{\lambda}$, the fact that the confidence sets are so vast helps to illustrate the problem.

When confidence sets for parameters are constructed using the CU-GMM objective function, lack of information about the parameters goes hand-in-hand with inability to reject the model. Given the confidence sets I have constructed, of course, we cannot formally reject the model. It is easy to find parameter values for which the test of the over-identifying restrictions fails to reject. But this hardly seems like a signal of the model's success.

Table 4 presents results of rank tests for the cases where the six Fama-French equity portfolios and the five Fama bond portfolios are added to the set of test assets. It also presents results of rank tests for the cases where LV's differenced currency portfolios are used

[1] See James H. Stock and Wright (2000), Stock, Wright and Yogo (2002) and Yogo (2004) for details.

[2] The test for $\lambda_m = 0$ is equivalent to a test that the vector $E(R^e)$ is equal to zero, if the GMM errors are demeaned before the long-run covariance matrix is computed. I do not demean the GMM errors, but doing so does not change the fact that the confidence region for λ_m encompasses most of the real line.

as the test assets. The results of the tests suggest that the model remains underidentified in these cases.

1.3 Estimates of the Models with Additional Test Assets

Table 5 provides GMM estimates of the model with no constant for the case where the currency portfolios and six Fama-French equity portfolios are used as test assets. Table 6 provides GMM estimates of the model with no constant for the case where the currency portfolios, six Fama-French equity portfolios, and five Fama bond portfolios are used as test assets.

2 Calculation of Standard Errors

2.1 Standard Errors for the Two-Pass Procedure

This section discusses the computation of standard errors for the two-pass regression 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).

2.1.1 Betas are Known

Equation (1) can be rewritten as

$$\mathbf{R}_t^e = \mathbf{a} + \boldsymbol{\beta}\mathbf{f}_t + \boldsymbol{\epsilon}_t$$

where $\mathbf{a}$ is an $n \times 1$ vector formed from the individual a_i, and $\boldsymbol{\epsilon}_t$ 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 $\mathrm{var}(\mathbf{f}_t) = \boldsymbol{\Sigma}_\mathbf{f}$ and $\mathrm{var}(\boldsymbol{\epsilon}_t) = \boldsymbol{\Sigma}$, but these assumptions can be generalized. Taking averages over time:

$$\bar{\mathbf{R}}^e = \mathbf{a} + \boldsymbol{\beta}\bar{\mathbf{f}} + \bar{\boldsymbol{\epsilon}}, \tag{2}$$

where $\bar{\mathbf{R}}^e$, $\bar{\mathbf{f}}$ and $\bar{\boldsymbol{\epsilon}}$ are the sample means of $\mathbf{R}_t^e$, $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t$.

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

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

Given that the beta representation implies that $E(\bar{\mathbf{R}}^e) = \boldsymbol{\beta}\boldsymbol{\lambda}$, it follows that

$$\text{plim}\,\hat{\boldsymbol{\alpha}} = \mathbf{M}_\beta E(\bar{\mathbf{R}}^e) = \mathbf{M}_\beta \boldsymbol{\beta}\boldsymbol{\lambda} = \mathbf{0}.$$

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

$$\boldsymbol{\Omega}_{\hat{\boldsymbol{\alpha}}} = \mathbf{M}_\beta \boldsymbol{\Omega}_{\bar{\mathbf{R}}} \mathbf{M}_\beta$$

where $\boldsymbol{\Omega}_{\bar{\mathbf{R}}}$ is the asymptotic covariance matrix of $\sqrt{T}(\bar{\mathbf{R}}^e - E\mathbf{R}^e)$. Given (2) and the assumptions made above:

$$\boldsymbol{\Omega}_{\bar{\mathbf{R}}} = \boldsymbol{\beta}\boldsymbol{\Sigma}_{\mathbf{f}}\boldsymbol{\beta}' + \boldsymbol{\Sigma}$$

hence

$$\boldsymbol{\Omega}_{\hat{\boldsymbol{\alpha}}} = \mathbf{M}_\beta \left(\boldsymbol{\beta}\boldsymbol{\Sigma}_{\mathbf{f}}\boldsymbol{\beta}' + \boldsymbol{\Sigma}\right) \mathbf{M}_\beta = \mathbf{M}_\beta \boldsymbol{\Sigma} \mathbf{M}_\beta.$$

Since $\boldsymbol{\Omega}_{\hat{\boldsymbol{\alpha}}}$ has rank $n-k$, $C = T\hat{\boldsymbol{\alpha}}'\boldsymbol{\Omega}_{\hat{\boldsymbol{\alpha}}}^{-1}\hat{\boldsymbol{\alpha}}$ 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 $\sqrt{T}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})$ is

$$\begin{aligned}\boldsymbol{\Omega}_{\hat{\boldsymbol{\lambda}}} &= (\boldsymbol{\beta}'\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{\Omega}_{\bar{\mathbf{R}}}\boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{\beta})^{-1} \\ &= \boldsymbol{\Sigma}_{\mathbf{f}} + (\boldsymbol{\beta}'\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{\Sigma}\boldsymbol{\beta}(\boldsymbol{\beta}'\boldsymbol{\beta})^{-1}.\end{aligned}$$

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

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

where $\boldsymbol{\theta} = \begin{pmatrix} \gamma & \boldsymbol{\lambda}' \end{pmatrix}'$, $\mathbf{X} = \begin{pmatrix} \iota & \boldsymbol{\beta} \end{pmatrix}$ and ι is an $n \times 1$ vector of ones. Therefore,

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

The beta representation states that $E(\mathbf{R}^e) = \gamma + \boldsymbol{\beta}\boldsymbol{\lambda} = \mathbf{X}\boldsymbol{\theta}$ where $\gamma = 0$. So

$$E(\hat{\mathbf{u}}) = \mathbf{M}_{\mathbf{X}} E(\bar{\mathbf{R}}^e) = \mathbf{M}_{\mathbf{X}}\mathbf{X}'\boldsymbol{\theta} = \mathbf{0}.$$

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

$$\boldsymbol{\Omega}_{\hat{\mathbf{u}}} = \mathbf{M}_{\mathbf{X}}\boldsymbol{\Omega}_{\bar{\mathbf{R}}}\mathbf{M}_{\mathbf{X}}.$$

The term $\boldsymbol{\beta}'\boldsymbol{\Sigma}_{\mathbf{f}}\boldsymbol{\beta}$ can be written as $\mathbf{X}\tilde{\boldsymbol{\Sigma}}_{\mathbf{f}}\mathbf{X}'$ with

$$\tilde{\boldsymbol{\Sigma}}_{\mathbf{f}} = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{f}} \end{pmatrix}.$$

Therefore we can rewrite $\mathbf{\Omega}_{\bar{\mathbf{R}}}$ as $\mathbf{X}\tilde{\mathbf{\Sigma}}_{\mathbf{f}}\mathbf{X}' + \mathbf{\Sigma}$ so that:

$$\mathbf{\Omega}_{\hat{\mathbf{u}}} = \mathbf{M}_{\mathbf{X}}(\mathbf{X}\tilde{\mathbf{\Sigma}}_{\mathbf{f}}\mathbf{X}' + \mathbf{\Sigma})\mathbf{M}_{\mathbf{X}} = \mathbf{M}_{\mathbf{X}}\mathbf{\Sigma}\mathbf{M}_{\mathbf{X}}.$$

Since $\mathbf{\Omega}_{\hat{\mathbf{u}}}$ has rank $n-k-1$, $C = T\hat{\mathbf{u}}'\mathbf{\Omega}_{\hat{\mathbf{u}}}^{-1}\hat{\mathbf{u}}$ 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 $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is

$$\begin{aligned}
\mathbf{\Omega}_{\hat{\boldsymbol{\theta}}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}_{\bar{\mathbf{R}}}\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\tilde{\mathbf{\Sigma}}_{\mathbf{f}}\mathbf{X}' + \mathbf{\Sigma})\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1} \\
&= \tilde{\mathbf{\Sigma}}_{\mathbf{f}} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}\mathbf{X}'(\mathbf{X}'\mathbf{X})^{-1}.
\end{aligned}$$

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 $\hat{\boldsymbol{\theta}}$. Alternatively one might also consider a reformulated χ^2 test based on

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

Letting

$$\mathbf{P} = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix}$$

we have

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

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

$$\mathbf{\Omega}_{\hat{\boldsymbol{\alpha}}} = \mathbf{H}(\mathbf{X}\tilde{\mathbf{\Sigma}}_{\mathbf{f}}\mathbf{X}' + \mathbf{\Sigma})\mathbf{H}' = \mathbf{H}\mathbf{\Sigma}\mathbf{H}'.$$

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

2.1.2 Shanken Corrections (Betas are Estimated)

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

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

where $\mathbf{R}_i^e$ is a $T \times 1$ vector with elements R_{it}^e and $\hat{\mathbf{f}}$ is a $T \times k$ matrix with rows equal to $(\mathbf{f}_t - \bar{\mathbf{f}})'$. Given the model, $\mathbf{R}_i^e = \mathbf{a}_i + \mathbf{f}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$, where $\mathbf{f}$ is an $T \times k$ matrix with rows equal to $\mathbf{f}_t'$ and $\boldsymbol{\epsilon}_i$ is a $T \times 1$ vector with elements ϵ_{it}. Hence

$$\begin{aligned}\hat{\boldsymbol{\beta}}_i &= (\hat{\mathbf{f}}'\hat{\mathbf{f}})^{-1}\hat{\mathbf{f}}'(\mathbf{a}_i + \mathbf{f}\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i) \\ &= \boldsymbol{\beta}_i + (\hat{\mathbf{f}}'\hat{\mathbf{f}})^{-1}\hat{\mathbf{f}}'\boldsymbol{\epsilon}_i.\end{aligned}$$

Assuming that $\mathbf{f}_t$ and $\boldsymbol{\epsilon}_t$ are independent, the asymptotic covariance between $\sqrt{T}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)$ and $\sqrt{T}(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j)$ is given by $\sigma_{ij}\boldsymbol{\Sigma}_{\mathbf{f}}^{-1}$ where σ_{ij} is the covariance between ϵ_{it} and ϵ_{jt}. If $\hat{\boldsymbol{\beta}}$ is rearranged into a $nk \times 1$ *stacked vector*,

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

the asymptotic covariance matrix of $\sqrt{T}(\hat{\boldsymbol{\beta}}_v - \boldsymbol{\beta}_v)$ is $\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}_{\mathbf{f}}^{-1}$.

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

$$\bar{\boldsymbol{\lambda}} = \boldsymbol{\lambda} + \bar{\mathbf{f}} - \boldsymbol{\mu}. \tag{3}$$

The model implies that $E\mathbf{R}^e = \mathbf{a} + \boldsymbol{\beta}\boldsymbol{\mu} = \boldsymbol{\beta}\boldsymbol{\lambda}$. Hence we can write $\mathbf{a} = \boldsymbol{\beta}(\boldsymbol{\lambda} - \boldsymbol{\mu})$. Substituting this result into (2) we get

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

Using (3) we have

$$\begin{aligned}\bar{\mathbf{R}}^e &= \boldsymbol{\beta}\boldsymbol{\lambda} + \bar{\boldsymbol{\epsilon}} \\ &= \hat{\boldsymbol{\beta}}\bar{\boldsymbol{\lambda}} + \bar{\boldsymbol{\epsilon}} - (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\bar{\boldsymbol{\lambda}}. \end{aligned} \tag{4}$$

Premultiplying (4) by $\hat{\mathbf{A}}$ we get

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

so that

$$\hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}} = \hat{\mathbf{A}}[\bar{\boldsymbol{\epsilon}} - (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\bar{\boldsymbol{\lambda}}]. \tag{5}$$

Now

$$\begin{aligned}\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda} &= (\hat{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}) + (\bar{\boldsymbol{\lambda}} - \boldsymbol{\lambda}) \\ &= \hat{\mathbf{A}}[\bar{\boldsymbol{\epsilon}} - (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\bar{\boldsymbol{\lambda}}] + (\bar{\mathbf{f}} - \boldsymbol{\lambda}).\end{aligned}$$

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

$$\boldsymbol{\Omega}_{\hat{\boldsymbol{\lambda}}} = \mathbf{A}[\boldsymbol{\Sigma} + (\mathbf{I}_n \otimes \boldsymbol{\lambda}') \left(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}_{\mathbf{f}}^{-1}\right) (\mathbf{I}_n \otimes \boldsymbol{\lambda})]\mathbf{A}' + \boldsymbol{\Sigma}_{\mathbf{f}}.$$

Using the rules for Kronecker products this reduces to

$$\boldsymbol{\Omega}_{\hat{\boldsymbol{\lambda}}} = (1 + \boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\mathbf{f}}^{-1}\boldsymbol{\lambda})\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' + \boldsymbol{\Sigma}_{\mathbf{f}}.$$

The pricing errors are

$$\begin{aligned}
\hat{\boldsymbol{\alpha}} &= \bar{\mathbf{R}}^e - \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\lambda}} = \left[\mathbf{I} - \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}})^{-1}\hat{\boldsymbol{\beta}}'\right]\bar{\mathbf{R}}^e = \mathbf{M}_{\hat{\beta}}\bar{\mathbf{R}}^e \\
&= \mathbf{M}_{\hat{\beta}}\left[\hat{\boldsymbol{\beta}}\bar{\boldsymbol{\lambda}} + \bar{\boldsymbol{\epsilon}} - (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\bar{\boldsymbol{\lambda}}\right] \\
&= \mathbf{M}_{\hat{\beta}}\left[\bar{\boldsymbol{\epsilon}} - (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\bar{\boldsymbol{\lambda}}\right]
\end{aligned}$$

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

$$\boldsymbol{\Omega}_{\hat{\boldsymbol{\alpha}}} = (1 + \boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\mathbf{f}}^{-1}\boldsymbol{\lambda})\mathbf{M}_{\beta}\boldsymbol{\Sigma}\mathbf{M}_{\beta}.$$

Since $\boldsymbol{\Omega}_{\hat{\boldsymbol{\alpha}}}$ has rank $n-k$, $C = T\hat{\boldsymbol{\alpha}}'\boldsymbol{\Omega}_{\hat{\boldsymbol{\alpha}}}^{-1}\hat{\boldsymbol{\alpha}}$ 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

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

where $\boldsymbol{\theta} = (\begin{array}{cc}\gamma & \boldsymbol{\lambda}'\end{array})'$, $\hat{\mathbf{X}} = (\begin{array}{cc}\boldsymbol{\iota}_n & \hat{\boldsymbol{\beta}}'\end{array})$ and $\boldsymbol{\iota}_n$ is an $n \times 1$ vector of ones. If $\hat{\mathbf{X}}$ is rearranged into a $n(k+1) \times 1$ *stacked vector*,

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

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

$$\boldsymbol{\Xi} = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{f}}^{-1} \end{pmatrix}.$$

We have $\hat{\boldsymbol{\theta}} = \hat{\mathbf{A}}\bar{\mathbf{R}}^e$, where $\hat{\mathbf{A}} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'$. The model implies that $E\mathbf{R}^e = \mathbf{a} + \boldsymbol{\beta\mu} = \gamma + \boldsymbol{\beta\lambda} = \boldsymbol{\beta\lambda}$ (since, under the null, $\gamma = 0$). Hence we can write $\mathbf{a} = \boldsymbol{\beta}(\boldsymbol{\lambda} - \boldsymbol{\mu})$. Substituting this result into (2) we get

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

Defining $\bar{\boldsymbol{\theta}} \equiv \begin{pmatrix} 0 & \bar{\boldsymbol{\lambda}}' \end{pmatrix}'$ we can then write this as

$$\begin{aligned} \bar{\mathbf{R}}^e &= \mathbf{X}\bar{\boldsymbol{\theta}} + \bar{\boldsymbol{\epsilon}} \\ &= \hat{\mathbf{X}}\bar{\boldsymbol{\theta}} + \bar{\boldsymbol{\epsilon}} - (\hat{\mathbf{X}} - \mathbf{X})\bar{\boldsymbol{\theta}}. \end{aligned} \tag{6}$$

Premultiplying (6) by $\hat{\mathbf{A}}$ we get

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

so that

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

Now

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

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

$$\boldsymbol{\Omega}_{\hat{\boldsymbol{\theta}}} = \mathbf{A}[\boldsymbol{\Sigma} + (\mathbf{I}_n \otimes \boldsymbol{\theta}')(\boldsymbol{\Sigma} \otimes \boldsymbol{\Xi})(\mathbf{I}_n \otimes \boldsymbol{\theta})]\mathbf{A}' + \tilde{\boldsymbol{\Sigma}}_{\mathbf{f}}.$$

Using the rules for Kronecker products this reduces to

$$\boldsymbol{\Omega}_{\hat{\boldsymbol{\theta}}} = (1 + \boldsymbol{\theta}'\boldsymbol{\Xi}\boldsymbol{\theta})\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' + \tilde{\boldsymbol{\Sigma}}_{\mathbf{f}},$$

but because of the form of $\boldsymbol{\Xi}$ it can also be written as

$$\boldsymbol{\Omega}_{\hat{\boldsymbol{\theta}}} = (1 + \boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\mathbf{f}}^{-1}\boldsymbol{\lambda})\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}' + \tilde{\boldsymbol{\Sigma}}_{\mathbf{f}}.$$

The pricing errors are

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

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

$$\boldsymbol{\Omega}_{\hat{\mathbf{u}}} = (1 + \boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\mathbf{f}}^{-1}\boldsymbol{\lambda})\mathbf{M}_{\mathbf{X}}\boldsymbol{\Sigma}\mathbf{M}_{\mathbf{X}}.$$

Since $\boldsymbol{\Omega}_{\hat{\mathbf{u}}}$ has rank $n-k-1$, $C = T\hat{\mathbf{u}}'\boldsymbol{\Omega}_{\hat{\mathbf{u}}}^{-1}\hat{\mathbf{u}}$ 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 $\hat{\boldsymbol{\theta}}$. Alternatively one might also consider a reformulated χ^2 test based on

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

Letting

$$\mathbf{P} = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix}$$

we have

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

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

$$\boldsymbol{\Omega}_{\hat{\boldsymbol{\alpha}}} = (1 + \boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\mathbf{f}}^{-1}\boldsymbol{\lambda})\mathbf{H}\boldsymbol{\Sigma}\mathbf{H}'.$$

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

2.1.3 GMM Standard Errors (Betas are Estimated)

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

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

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

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

the $n(k+2) \times 1$ vector $\mathbf{g}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{u}_t(\boldsymbol{\theta})$, and the $[n(k+1)+k] \times [n(k+2) \times 1]$ matrix

$$\mathbf{a}_T = \begin{pmatrix} \mathbf{I}_{n(k+1)} & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{\beta}}'_{OLS} \end{pmatrix}.$$

The GMM estimator that sets $\mathbf{a}_T\mathbf{g}_T = \mathbf{0}$ reproduces the two-pass estimates of $\mathbf{a}$, $\boldsymbol{\beta}$, and $\boldsymbol{\lambda}$. Define

$$\mathbf{d}_T = \frac{\partial \mathbf{g}_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \begin{pmatrix} -\mathbf{I}_n \otimes \mathbf{M}_{\tilde{\mathbf{f}}} & \mathbf{0}_{n(k+1)\times k} \\ -\mathbf{I}_n \otimes \begin{pmatrix} 0 & \hat{\boldsymbol{\lambda}}' \end{pmatrix} & -\hat{\boldsymbol{\beta}}_{OLS} \end{pmatrix}$$

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

Let $\mathbf{a} = \text{plim}\, \mathbf{a}_T$ and $\mathbf{d} = \text{plim}\, \mathbf{d}_T$. The covariance matrix of $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is

$$\mathbf{V}_{\boldsymbol{\theta}} = (\mathbf{ad})^{-1}\mathbf{aSa}'\left[(\mathbf{ad})^{-1}\right]'$$

and the covariance matrix of $\sqrt{T}\mathbf{g}_T(\hat{\boldsymbol{\theta}})$ is

$$\mathbf{V}_{\mathbf{g}} = [\mathbf{I} - \mathbf{d}(\mathbf{ad})^{-1}\mathbf{a}]\mathbf{S}[\mathbf{I} - \mathbf{d}(\mathbf{ad})^{-1}\mathbf{a}]'$$

where $\mathbf{S}$ is the asymptotic covariance matrix of $\sqrt{T}\mathbf{g}_T(\boldsymbol{\theta})$. These results follow from the facts that $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} (\mathbf{ad})^{-1}\mathbf{a}\sqrt{T}\mathbf{g}_T(\boldsymbol{\theta})$ and $\sqrt{T}\mathbf{g}_T(\hat{\boldsymbol{\theta}}) \xrightarrow{d} [\mathbf{I} - \mathbf{d}(\mathbf{ad})^{-1}\mathbf{a}]\sqrt{T}\mathbf{g}_T(\boldsymbol{\theta})$. The test statistic for the pricing errors is just $T\mathbf{g}_T(\hat{\boldsymbol{\theta}})'\mathbf{V}_{\mathbf{g}}^{-1}\mathbf{g}_T(\hat{\boldsymbol{\theta}})$, where the inverse is generalized. Since $\mathbf{S} = \sum_{j=-\infty}^{\infty} E(\mathbf{u}_t\mathbf{u}_{t-j}')$, I use a variant of a VARHAC estimator for $\mathbf{S}$: 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 $E(R^e_{it} - a_i - \boldsymbol{\beta}_i'\mathbf{f}_t) = 0$, $E[(R^e_{it} - a_i - \boldsymbol{\beta}_i'\mathbf{f}_t)\mathbf{f}_t'] = \mathbf{0}$, and $E(R^e_{it} - \gamma - \boldsymbol{\beta}_i'\boldsymbol{\lambda}) = 0$, $i = 1, \ldots, n$. Now let

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

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

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

the $n(k+2)\times 1$ vector $\mathbf{g}_T(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_t(\boldsymbol{\theta})$,and the $(n+1)(k+1)\times[n(k+2)\times 1]$ matrix

$$\mathbf{a}_T = \begin{pmatrix} \mathbf{I}_{n(k+1)} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{X}}' \end{pmatrix}.$$

where $\hat{\mathbf{X}} = \begin{pmatrix} \boldsymbol{\iota}_{n\times 1} & \hat{\boldsymbol{\beta}}_{OLS} \end{pmatrix}$. The GMM estimator that sets $\mathbf{a}_T\mathbf{g}_T = \mathbf{0}$ reproduces the two-pass estimates of $\mathbf{a}$, $\boldsymbol{\beta}$, γ and $\boldsymbol{\lambda}$. Define

$$\mathbf{d}_T = \frac{\partial \mathbf{g}_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \begin{pmatrix} -\mathbf{I}_n \otimes \mathbf{M}_{\tilde{\mathbf{f}}} & \mathbf{0}_{n(k+1)\times(k+1)} \\ -\mathbf{I}_n \otimes \begin{pmatrix} 0 & \hat{\boldsymbol{\lambda}}' \end{pmatrix} & -\hat{\mathbf{X}} \end{pmatrix}.$$

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

$$\mathbf{V}_\theta = (\mathbf{ad})^{-1}\mathbf{aSa}'\left[(\mathbf{ad})^{-1}\right]'$$

and the covariance matrix of $\sqrt{T}\mathbf{g}_T(\hat{\boldsymbol{\theta}})$ is

$$\mathbf{V}_{\mathbf{g}} = [\mathbf{I} - \mathbf{d}(\mathbf{ad})^{-1}\mathbf{a}]\mathbf{S}[\mathbf{I} - \mathbf{d}(\mathbf{ad})^{-1}\mathbf{a}]'$$

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

2.2 GMM Estimation of the Model

2.2.1 Model without a Constant

Asympotic Theory Let

$$\mathbf{u}_{1t}(\mathbf{b}, \boldsymbol{\mu}) = \mathbf{R}^e_t[1 - (\mathbf{f}_t - \boldsymbol{\mu})'\mathbf{b}] \tag{7}$$

$$\mathbf{g}_{1T}(\mathbf{b}, \boldsymbol{\mu}) = T^{-1}\sum_{t=1}^{T}\mathbf{u}_{1t}(\mathbf{b}, \boldsymbol{\mu}) = \bar{\mathbf{R}}^e(1 + \boldsymbol{\mu}'\mathbf{b}) - \mathbf{D}_T\mathbf{b}. \tag{8}$$

where $\mathbf{D}_T = T^{-1}\sum_{t=1}^T \mathbf{R}_t^e \mathbf{f}_t'$ and $\bar{\mathbf{R}}^e = T^{-1}\sum_{t=1}^T \mathbf{R}_t^e$. Also define

$$\mathbf{u}_{2t}(\boldsymbol{\mu}) = \mathbf{f}_t - \boldsymbol{\mu} \tag{9}$$

$$\mathbf{g}_{2T}(\boldsymbol{\mu}) = T^{-1}\sum_{t=1}^T \mathbf{u}_{2t}(\boldsymbol{\mu}) = \bar{\mathbf{f}} - \boldsymbol{\mu}. \tag{10}$$

Finally, define the stacked vectors

$$\mathbf{u}_t(\mathbf{b},\boldsymbol{\mu}) = \begin{pmatrix} \mathbf{u}_{1t}(\mathbf{b},\boldsymbol{\mu}) \\ \mathbf{u}_{2t}(\boldsymbol{\mu}) \end{pmatrix} \qquad \mathbf{g}_T(\mathbf{b},\boldsymbol{\mu}) = \begin{pmatrix} \mathbf{g}_{1T}(\mathbf{b},\boldsymbol{\mu}) \\ \mathbf{g}_{2T}(\boldsymbol{\mu}) \end{pmatrix}$$

and the matrix

$$\mathbf{S} = E[\sum_{j=-\infty}^{\infty} \mathbf{u}_t(\mathbf{b}_0,\boldsymbol{\mu}_0)\mathbf{u}_{t-j}(\mathbf{b}_0,\boldsymbol{\mu}_0)'].$$

The parameters $\mathbf{b}$ and $\boldsymbol{\mu}$ are estimated by setting $\mathbf{a}_T\mathbf{g}_T = \mathbf{0}$, where

$$\mathbf{a}_T = \begin{pmatrix} (\mathbf{D}_T - \bar{\mathbf{R}}^e\boldsymbol{\mu}')'\mathbf{W}_T & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix},$$

and $\mathbf{W}_T$ is some weighting matrix. Given the definition of $\mathbf{g}_T$ this means the GMM estimator is the solution to

$$\begin{pmatrix} (\mathbf{D}_T - \bar{\mathbf{R}}^e\boldsymbol{\mu}')'\mathbf{W}_T & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix} \begin{pmatrix} \bar{\mathbf{R}}^e(1+\boldsymbol{\mu}'\mathbf{b}) - \mathbf{D}_T\mathbf{b} \\ \bar{\mathbf{f}} - \boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \tag{11}$$

implying that

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{f}} \tag{12}$$

$$\hat{\mathbf{b}} = \left(\mathbf{d}_T'\mathbf{W}_T\mathbf{d}_T\right)^{-1}\mathbf{d}_T'\mathbf{W}_T\bar{\mathbf{R}}^e, \tag{13}$$

where $\mathbf{d}_T = \mathbf{D}_T - \bar{\mathbf{R}}^e\bar{\mathbf{f}}'$.

In the first stage the weighting matrix is $\mathbf{W}_T = \mathbf{I}_n$. In the second stage, Lustig and Verdelhan follow Cochrane (2005) and set

$$\mathbf{W}_T = \left[T^{-1}\sum_{t=1}^T \mathbf{u}_{1t}(\hat{\mathbf{b}},\hat{\boldsymbol{\mu}})\mathbf{u}_{1t}(\hat{\mathbf{b}},\hat{\boldsymbol{\mu}})'\right]^{-1} \tag{14}$$

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

$$\operatorname{plim}\mathbf{W}_T = \mathbf{W} = \mathbf{S}_{11}^{-1} \text{ where } \mathbf{S}_{11} = E[\mathbf{u}_{1t}(\mathbf{b}_0,\boldsymbol{\mu}_0)\mathbf{u}_{1t}(\mathbf{b}_0,\boldsymbol{\mu}_0)'].$$

Given (13), $\operatorname{plim}\hat{\mathbf{b}} = \mathbf{b}_0$. This follows from the fact that $\operatorname{plim}\mathbf{d}_T = \mathbf{d} \equiv E[\mathbf{R}_t^e(\mathbf{f}_t - \boldsymbol{\mu})']$ and that $\operatorname{plim}\bar{\mathbf{R}}^e = E(\mathbf{R}^e)$. We then get $\operatorname{plim}\hat{\mathbf{b}} = \left(\mathbf{d}'\mathbf{W}\mathbf{d}\right)^{-1}\mathbf{d}'\mathbf{W}E(\mathbf{R}^e)$. The model implies that $E(\mathbf{R}^e) = \mathbf{d}\mathbf{b}_0$. Hence $\operatorname{plim}\hat{\mathbf{b}} = \mathbf{b}_0$. So the first and second stage estimates of $\mathbf{b}$ are obviously consistent.

The derivation of the asymptotic distribution of $(\hat{\mathbf{b}}, \hat{\boldsymbol{\mu}})$ relies on deriving the distance between $\mathbf{g}_T(\hat{\mathbf{b}}, \hat{\boldsymbol{\mu}})$ and $\mathbf{g}_T(\mathbf{b}_0, \boldsymbol{\mu}_0)$. Using (8) and the consistency of $\hat{\mathbf{b}}$ and $\hat{\boldsymbol{\mu}}$ we can argue that there is a pair $(\bar{\mathbf{b}},\bar{\mu})$ between $(\mathbf{b}_0, \boldsymbol{\mu}_0)$ and $(\hat{\mathbf{b}}, \hat{\boldsymbol{\mu}})$ such that

$$\mathbf{g}_{1T}(\hat{\mathbf{b}}, \hat{\boldsymbol{\mu}}) = \mathbf{g}_{1T}(\mathbf{b}_0, \boldsymbol{\mu}_0) + \left(\bar{\mathbf{R}}^e \bar{\boldsymbol{\mu}}' - \mathbf{D}_T\right)(\hat{\mathbf{b}} - \mathbf{b}_0) + \bar{\mathbf{R}}^e \bar{\mathbf{b}}'(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0). \tag{15}$$

From (10) we also have

$$\mathbf{g}_{2T}(\hat{\boldsymbol{\mu}}) = \mathbf{g}_{2T}(\boldsymbol{\mu}_0) - (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0). \tag{16}$$

Premultiplying (15) by $\mathbf{d}_T'\mathbf{W}_T$ one obtains

$$\mathbf{0} = \mathbf{d}_T'\mathbf{W}_T\mathbf{g}_{1T}(\hat{\mathbf{b}}, \hat{\boldsymbol{\mu}}) = \mathbf{d}_T'\mathbf{W}_T[\mathbf{g}_{1T}(\mathbf{b}_0, \boldsymbol{\mu}_0) + \left(\bar{\mathbf{R}}^e \bar{\boldsymbol{\mu}}' - \mathbf{D}_T\right)(\hat{\mathbf{b}} - \mathbf{b}_0) + \bar{\mathbf{R}}^e \bar{\mathbf{b}}'(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)] \tag{17}$$

We can rewrite (16) and (17) together as

$$\mathbf{0} = \hat{\mathbf{a}}_T \left[\mathbf{g}_T(\mathbf{b}_0, \boldsymbol{\mu}_0) - \boldsymbol{\Delta}_T \begin{pmatrix} \hat{\mathbf{b}} - \mathbf{b}_0 \\ \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0 \end{pmatrix}\right]. \tag{18}$$

where

$$\hat{\mathbf{a}}_T = \begin{pmatrix} \mathbf{d}_T'\mathbf{W}_T & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix} \quad \boldsymbol{\Delta}_T = \begin{pmatrix} \left(\mathbf{D}_T - \bar{\mathbf{R}}^e \bar{\boldsymbol{\mu}}'\right) & -\bar{\mathbf{R}}^e \bar{\mathbf{b}}' \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix}.$$

We have $\operatorname{plim}\hat{\mathbf{a}}_T = \mathbf{a}$ and $\operatorname{plim}\boldsymbol{\Delta}_T = \boldsymbol{\Delta}$ where

$$\mathbf{a} = \begin{pmatrix} \mathbf{d}'\mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix} \qquad \boldsymbol{\Delta} = \begin{pmatrix} \mathbf{d} & -\mathbf{d}\mathbf{b}_0\mathbf{b}_0' \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix},$$

and I have used the fact that $\operatorname{plim}\bar{\mathbf{R}}^e = E(\mathbf{R}^e) = \mathbf{d}\mathbf{b}_0$. Hence

$$\sqrt{T}\begin{pmatrix} \hat{\mathbf{b}} - \mathbf{b}_0 \\ \hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} (\mathbf{d}'\mathbf{W}\mathbf{d})^{-1} & \mathbf{b}_0\mathbf{b}_0' \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix} \begin{pmatrix} \mathbf{d}'\mathbf{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix} \sqrt{T}\mathbf{g}_T(\mathbf{b}_0, \boldsymbol{\mu}_0).$$

Thus we have

$$\begin{aligned} \sqrt{T}(\hat{\mathbf{b}} - \mathbf{b}_0) &\xrightarrow{d} \begin{pmatrix} (\mathbf{d}'\mathbf{W}\mathbf{d})^{-1}\mathbf{d}'\mathbf{W} & \mathbf{b}_0\mathbf{b}_0' \end{pmatrix} \sqrt{T}\mathbf{g}_T(\mathbf{b}_0, \boldsymbol{\mu}_0) = \mathbf{B}\sqrt{T}\mathbf{g}_T(\mathbf{b}_0, \boldsymbol{\mu}_0) \\ \sqrt{T}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0) &\xrightarrow{d} \begin{pmatrix} \mathbf{0} & \mathbf{I}_k \end{pmatrix} \sqrt{T}\mathbf{g}_T(\mathbf{b}_0, \boldsymbol{\mu}_0) = \sqrt{T}\mathbf{g}_{2T}(\mathbf{b}_0, \boldsymbol{\mu}_0) \end{aligned}$$

and the asymptotic covariance matrix of $\sqrt{T}(\hat{\mathbf{b}} - \mathbf{b}_0)$ is

$$\mathbf{V}_{\mathbf{b}} = \mathbf{B}\mathbf{S}\mathbf{B}'. \tag{19}$$

The fact that $\boldsymbol{\mu}$ is estimated affects $\mathbf{V_b}$. If $\boldsymbol{\mu}$ was known the covariance matrix would reduce to $(\mathbf{d'Wd})^{-1}\,\mathbf{d'WS}_{11}\mathbf{Wd}\,(\mathbf{d'Wd})^{-1}$.

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

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

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

$$\mathbf{V}_0=[\mathbf{I}-\mathbf{d}\,(\mathbf{d'Wd})^{-1}\,\mathbf{d'W}]\mathbf{S}_{11}[\mathbf{I}-\mathbf{Wd}\,(\mathbf{d'Wd})^{-1}\,\mathbf{d}'].$$

This is the same expression as one obtains when $\boldsymbol{\mu}$ is known. Since the test of the pricing errors is obtained as $T\mathbf{g}_{1T}(\hat{\mathbf{b}},\hat{\boldsymbol{\mu}})'\mathbf{V}_T^{-1}\mathbf{g}_{1T}(\hat{\mathbf{b}},\hat{\boldsymbol{\mu}})$, where the inverse is generalized, and $\mathbf{V}_T$ is a consistent estimate of $\mathbf{V}_0$, the fact that $\boldsymbol{\mu}$ is estimated has no effect on the statistic as compared to the case where $\boldsymbol{\mu}$ is treated as known.

Factor Risk Premia The GMM estimator produces estimates of $\mathbf{b}$ and $\boldsymbol{\mu}$. To obtain an estimate of $\boldsymbol{\lambda}$ we can use the expression $\boldsymbol{\lambda}=\boldsymbol{\Sigma}_\mathbf{f}\mathbf{b}$. This requires estimation of $\boldsymbol{\Sigma}_\mathbf{f}$. This can be done by adding moment conditions that identify the unique elements of $\boldsymbol{\Sigma}_\mathbf{f}$:

$$E[(f_{it}-\mu_i)\left(f_{jt}-\mu_j\right)-\boldsymbol{\Sigma}_{\mathbf{f},ij}],\quad i=1,\ldots,k,\quad j=i,\ldots,k. \tag{20}$$

The estimate $\hat{\boldsymbol{\Sigma}}_\mathbf{f}$ then corresponds to the sample covariance matrix of $\mathbf{f}_t$. Of course, standard errors for $\hat{\boldsymbol{\lambda}}$ should take into account estimation of $\boldsymbol{\Sigma}_\mathbf{f}$.

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

VARHAC Spectral Density Matrix Since $\mathbf{S}=\sum_{j=-\infty}^{\infty}E(\mathbf{u}_t\mathbf{u}_{t-j}')$, I estimate it as follows. Define $\mathbf{u}_{1t}$ and $\mathbf{u}_{2t}$ as in (7) and (9). I use a VARHAC estimator for $\mathbf{S}$, imposing the restriction that $E\mathbf{u}_{1t}\mathbf{u}_{t-j}'=\mathbf{0}$ for $j\geq 1$. This means that the VARHAC estimator for $\mathbf{S}_{11}$, the sub-block of $\mathbf{S}$ equal to $\sum_{j=-\infty}^{\infty}E(\mathbf{u}_{1t}\mathbf{u}_{1t-j}')$, is the same as the HAC estimator for $\mathbf{S}_{11}$. But this is not true for the $\mathbf{S}_{12}$, $\mathbf{S}_{21}$ and $\mathbf{S}_{22}$ sub-blocks. In practice, the VARHAC procedure typically finds persistence in some elements of $\mathbf{u}_{2t}$ because these are the GMM errors corresponding to $\mathbf{f}_t - \hat{\boldsymbol{\mu}}$. 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

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

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

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

where the inverse is generalized.

At the second stage of GMM we have

$$\hat{\mathbf{b}}_2 = \left(\mathbf{d}_T'\mathbf{W}_T\mathbf{d}_T\right)^{-1}\mathbf{d}_T'\mathbf{W}_T\bar{\mathbf{R}}^e$$

so the pricing errors are $\hat{\boldsymbol{\alpha}}_2 = \bar{\mathbf{R}}^e - \mathbf{d}_T\hat{\mathbf{b}}_2 = \mathbf{M}_\mathbf{W}\bar{\mathbf{R}}^e$ where $\mathbf{M}_\mathbf{W} = [\mathbf{I} - \mathbf{d}_T\left(\mathbf{d}_T'\mathbf{W}_T\mathbf{d}_T\right)^{-1}\mathbf{d}_T'\mathbf{W}_T]\bar{\mathbf{R}}^e$. The estimated covariance matrix of $\hat{\boldsymbol{\alpha}}_2$ is $\mathbf{V}_T = \mathbf{M}_\mathbf{W}\hat{\mathbf{S}}_{11}\mathbf{M}_\mathbf{W}'$ so the test statistic is

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

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

2.2.2 Model with a Constant

Asymptotic Theory Let

$$\mathbf{u}_{1t}(\mathbf{b},\boldsymbol{\mu},\gamma) = \bar{\mathbf{R}}^e[1 - (\mathbf{f}_t - \boldsymbol{\mu})'\mathbf{b}] - \gamma \tag{21}$$

$$\mathbf{g}_{1T}(\mathbf{b},\boldsymbol{\mu},\gamma) = T^{-1}\sum_{t=1}^{T}\mathbf{u}_{1t}(\mathbf{b},\boldsymbol{\mu},\gamma) = \bar{\mathbf{R}}^e(1 + \bar{\boldsymbol{\mu}}'\bar{\mathbf{b}}) - \bar{\mathbf{D}}_T\bar{\mathbf{b}}. \tag{22}$$

where $\bar{\mathbf{b}} = \begin{pmatrix}\gamma & \mathbf{b}'\end{pmatrix}'$, $\bar{\boldsymbol{\mu}} = \begin{pmatrix}0 & \boldsymbol{\mu}'\end{pmatrix}'$ and $\bar{\mathbf{D}}_T = \begin{pmatrix}\boldsymbol{\iota}_{n\times 1} & \mathbf{D}_T\end{pmatrix}$.

Define $\mathbf{u}_{2t}$ and $\mathbf{g}_{2T}$ as in (9) and (10). Define the stacked vectors

$$\mathbf{u}_t(\mathbf{b},\boldsymbol{\mu},\gamma) = \begin{pmatrix}\mathbf{u}_{1t}(\mathbf{b},\boldsymbol{\mu},\gamma)\\ \mathbf{u}_{2t}(\boldsymbol{\mu})\end{pmatrix} \qquad \mathbf{g}_T(\mathbf{b},\boldsymbol{\mu},\gamma) = \begin{pmatrix}\mathbf{g}_{1T}(\mathbf{b},\boldsymbol{\mu},\gamma)\\ \mathbf{g}_{2T}(\boldsymbol{\mu})\end{pmatrix}$$

and the matrix

$$\mathbf{S} = E[\sum_{j=-\infty}^{\infty} \mathbf{u}_t(\mathbf{b}_0, \boldsymbol{\mu}_0, \gamma_0)\mathbf{u}_{t-j}(\mathbf{b}_0, \boldsymbol{\mu}_0, \gamma_0)'].$$

The parameters $\bar{\mathbf{b}}$ and $\boldsymbol{\mu}$ are estimated by setting $\mathbf{a}_T\mathbf{g}_T = \mathbf{0}$, where

$$\mathbf{a}_T = \begin{pmatrix} (\bar{\mathbf{D}}_T - \bar{\mathbf{R}}^e\bar{\boldsymbol{\mu}}')'\mathbf{W}_T & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix},$$

and $\mathbf{W}_T$ is some weighting matrix. Given the definition of $\mathbf{g}_T$ this means the GMM estimator is the solution to

$$\begin{pmatrix} (\bar{\mathbf{D}}_T - \bar{\mathbf{R}}^e\bar{\boldsymbol{\mu}}')'\mathbf{W}_T & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k \end{pmatrix} \begin{pmatrix} \bar{\mathbf{R}}^e(1+\bar{\boldsymbol{\mu}}'\bar{\mathbf{b}}) - \bar{\mathbf{D}}_T\bar{\mathbf{b}} \\ \bar{\mathbf{f}} - \boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \tag{23}$$

implying that

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{f}} \tag{24}$$

$$\widehat{\bar{\mathbf{b}}} = \left(\bar{\mathbf{d}}_T'\mathbf{W}_T\bar{\mathbf{d}}_T\right)^{-1}\bar{\mathbf{d}}_T'\mathbf{W}_T\bar{\mathbf{R}}^e, \tag{25}$$

where $\bar{\mathbf{d}}_T = \bar{\mathbf{D}}_T - \bar{\mathbf{R}}^e\widehat{\bar{\boldsymbol{\mu}}}' = \begin{pmatrix} \boldsymbol{\iota}_{n\times1} & \mathbf{D}_T \end{pmatrix} - \bar{\mathbf{R}}^e\begin{pmatrix} \mathbf{0} & \bar{\mathbf{f}}' \end{pmatrix} = \begin{pmatrix} \boldsymbol{\iota}_{n\times1} & d_T \end{pmatrix}$.

The first and second stage estimates are calculated as in the case with the constant. In the first stage $\mathbf{W}_T = \mathbf{I}_n$. In the second stage, $\mathbf{W}_T$ is the inverse of a consistent estimator for $\mathbf{S}_{11} = E[\mathbf{u}_{1t}(\mathbf{b}_0, \boldsymbol{\mu}_0, \gamma_0)\mathbf{u}_{1t}(\mathbf{b}_0, \boldsymbol{\mu}_0, \gamma_0)']$.

Equivalence Between the First Stage of GMM and the Two-Pass Procedure The first stage estimate of $\bar{\mathbf{b}}$ based on $\mathbf{W} = \mathbf{I}_n$ is $\widehat{\bar{\mathbf{b}}} = \left(\bar{\mathbf{d}}_T'\bar{\mathbf{d}}_T\right)^{-1}\bar{\mathbf{d}}_T'\bar{\mathbf{R}}^e$. The matrix $\bar{\mathbf{d}}_T = \begin{pmatrix} \boldsymbol{\iota}_{n\times1} & \mathbf{d}_T \end{pmatrix}$, which can be rewritten as $\bar{\mathbf{d}}_T = \begin{pmatrix} \boldsymbol{\iota}_{n\times1} & \hat{\boldsymbol{\beta}}\hat{\boldsymbol{\Sigma}}_{\mathbf{f}} \end{pmatrix}$. Hence

$$\begin{aligned}
\widehat{\bar{\mathbf{b}}} &= \begin{pmatrix} \boldsymbol{\iota}'\boldsymbol{\iota} & \boldsymbol{\iota}'\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\Sigma}}_{\mathbf{f}} \\ \hat{\boldsymbol{\Sigma}}_{\mathbf{f}}\hat{\boldsymbol{\beta}}'\boldsymbol{\iota} & \hat{\boldsymbol{\Sigma}}_{\mathbf{f}}\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\Sigma}}_{\mathbf{f}} \end{pmatrix}^{=1} \begin{pmatrix} \boldsymbol{\iota}'\bar{\mathbf{R}}^e \\ \hat{\boldsymbol{\Sigma}}_{\mathbf{f}}\hat{\boldsymbol{\beta}}'\bar{\mathbf{R}}^e \end{pmatrix} \\
&= \left[\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{\Sigma}}_{\mathbf{f}} \end{pmatrix}\begin{pmatrix} \boldsymbol{\iota}'\boldsymbol{\iota} & \boldsymbol{\iota}'\hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\beta}}'\boldsymbol{\iota} & \hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}} \end{pmatrix}\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{\Sigma}}_{\mathbf{f}} \end{pmatrix}\right]^{-1} \begin{pmatrix} \boldsymbol{\iota}'\bar{\mathbf{R}}^e \\ \hat{\boldsymbol{\Sigma}}_{\mathbf{f}}\hat{\boldsymbol{\beta}}'\bar{\mathbf{R}}^e \end{pmatrix} \\
&= \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{\Sigma}}_{\mathbf{f}}^{-1} \end{pmatrix}\begin{pmatrix} \boldsymbol{\iota}'\boldsymbol{\iota} & \boldsymbol{\iota}'\hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\beta}}'\boldsymbol{\iota} & \hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{\iota}'\bar{\mathbf{R}}^e \\ \hat{\boldsymbol{\beta}}'\bar{\mathbf{R}}^e \end{pmatrix}.
\end{aligned}$$

The two-step estimator of γ and $\boldsymbol{\lambda}$ is

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

Hence

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

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

3 Replication of Lustig and Verdelhan's Results

Some results in my paper are directly comparable to results presented in LV's original article. In some cases differences arise, and I try to explain these differences in this section.

Their Table 5 (column EZ-DCAPM) and Table 14 (panel B, column C) are directly comparable to my Table 3. The point estimates of $\boldsymbol{\lambda}$, OLS standard errors, and R^2 are identical. The p-values for the pricing error test are different. They report a p-value of 0.628, while I report a p-value of 0.483. The value of my 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 LV's p-value, 0.628. The Shanken standard errors are different. 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 $[1 + (\boldsymbol{\lambda}'\mathbf{\Sigma}_{\mathbf{f}}^{-1}\boldsymbol{\lambda})](\mathbf{A\Sigma A}' + \tilde{\mathbf{\Sigma}}_{\mathbf{f}})$ in computing the standard errors instead of $[1 + (\boldsymbol{\lambda}'\mathbf{\Sigma}_{\mathbf{f}}^{-1}\boldsymbol{\lambda})]\mathbf{A\Sigma A}' + \tilde{\mathbf{\Sigma}}_{\mathbf{f}}$ (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. We report different types of GMM standard errors (they use HAC, while I used VARHAC), so they are not directly comparable. The reported mean absolute pricing errors also differ. I cannot account for this difference.

LV's Table 6 reports individual factor betas, not partial factor betas obtained in the first pass regressions, so it is not comparable to Table 1 in this appendix.

Table 8 of LV's paper appears to repeat a typo contained in Yogo (2006). The structural parameters of the Yogo model map to the bs according to

$$b_1 = \kappa(1/\sigma + \alpha(1/\rho - 1/\sigma)) \quad (26)$$

$$b_2 = \kappa\alpha(1/\sigma - 1/\rho) \quad (27)$$

$$b_3 = 1 - \kappa \quad (28)$$

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

$$\sigma = \frac{1-b_3}{b_1+b_3} \tag{29}$$
$$\gamma = b_1+b_2+b_3 \tag{30}$$
$$\alpha = \frac{b_2}{b_1+b_2+(b_3-1)/\rho}. \tag{31}$$

The expression for σ (29) is wrong and should be:

$$\sigma = \frac{1-b_3}{b_1+b_2} \tag{32}$$

With $b_c = -21.0$, $b_d = 130$ and $b_r = 4.46$ I obtain $\sigma = -0.032$ using (32). Using the incorrect formula in (29) 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.

LV's Table 14 (panel A, column C) is directly comparable to my Tables 7 and 8. Our reported point estimates are identical. They appear to report the MAE for the first stage of GMM rather than the second stage. We report different types of GMM standard errors, so they are not directly comparable. However, I believe their reported HAC standard errors to be incorrect. My code can be used for the HAC case, and I do not replicate their results. I believe their code ignores the sampling uncertainty induced by $\boldsymbol{\mu}$ being estimated.

When $\boldsymbol{\mu}_0$ is known, the expression in (19) is simpler, and reduces to

$$\mathbf{V}_k = (\mathbf{d}'\mathbf{W}\mathbf{d})^{-1}\,\mathbf{d}'\mathbf{W}\mathbf{S}_{11}\mathbf{W}\mathbf{d}\,(\mathbf{d}'\mathbf{W}\mathbf{d})^{-1}. \tag{33}$$

In the second stage of GMM $\mathbf{W}_0 = \mathbf{S}_{11}^{-1}$ so (33) reduces to

$$\mathbf{V}_k = \left(\mathbf{d}'\mathbf{S}_{11}^{-1}\mathbf{d}\right)^{-1}. \tag{34}$$

I believe that LV base their GMM standard errors on (34). However, this is inappropriate when $\boldsymbol{\mu}_0$ must be estimated. This is because $\mathbf{V}_{\mathbf{b}}$, given in (19), does not reduce to $\mathbf{V}_k$ unless $\mathbf{b}_0 = 0$ or $\boldsymbol{\mu}_0$ is known. This problem does not bias the standard errors sharply in a consistent direction, and the differences it induces are small.

Finally, my Tables 5(b) and 6(b) in this appendix are not directly comparable to LV's Table 14 (panel A, columns E/C and E/B/C). The reason is that LV do not provide the equity and bond portfolio data in their archive. Consequently, differences between our data series for these test assets probably account for any differencs in results.

Appendix References not Included in the Main Article

Stock, James H. and Jonathan H. Wright (2000) "GMM with Weak Identification," *Econometrica* 68, 1055–96.

Stock, James H., Jonathan H. Wright and Motohiro Yogo (2002) "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments" Journal of Business and Economic Statistics 20: 518–29.

Yogo, Motohiro (2004) "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak," *Review of Economics and Statistics* 86: 797–810.

TABLE 1 (Part 1): First-Pass Estimates of the Betas, Portfolios P1–P4

Portfolio	Factor (f_j)			Test of
(Pi)	Δc	Δd	r_W	$\beta_{ij} = 0\ \forall j$
P1	0.201	0.028	−0.068	
	(0.852)	(0.612)	(0.055)	(0.600)
	[0.643]	[0.563]	[0.049]	[0.492]
	$\langle -1.4, 2.0 \rangle$	$\langle -1.4, 1.0 \rangle$	$\langle -0.19, 0.03 \rangle$	
P2	0.740	0.091	−0.034	
	(0.889)	(0.638)	(0.058)	(0.579)
	[0.620]	[0.470]	[0.062]	[0.164]
	$\langle -1.0, 2.1 \rangle$	$\langle -0.8, 1.6 \rangle$	$\langle -0.16, 0.13 \rangle$	
P3	−0.639	0.962	0.019	
	(0.882)	(0.633)	(0.057)	(0.464)
	[1.026]	[0.814]	[0.058]	[0.645]
	$\langle -2.6, 1.5 \rangle$	$\langle -0.8, 1.6 \rangle$	$\langle -0.13, 0.11 \rangle$	
P4	−0.546	0.982	−0.089	
	(1.095)	(0.786)	(0.071)	(0.156)
	[1.075]	[0.749]	[0.069]	[0.065]
	$\langle -2.6, 2.3 \rangle$	$\langle -1.3, 1.7 \rangle$	$\langle -0.25, 0.04 \rangle$	

Notes: Annual data, 1953–2002. The regression equation is $R^e_{it} = a_i + \mathbf{f}_t'\boldsymbol{\beta}_i + \epsilon_{it}$, where R^e_{it} is the excess return of portfolio i at time t, $\mathbf{f}_t = (\ \Delta c_t \quad \Delta d_t \quad r_{Wt}\ )'$, 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 United States, where P1 and P8 are the portfolios with, respectively, the smallest and largest interest differentials. The table reports $\hat{\beta}_{ij}$ and p-values for tests of the hypotheses that $\beta_{ij} = 0\ \forall j$ for each portfolio i. For estimates of betas, OLS standard errors are in parentheses, and GMM-VARHAC standard errors are in square brackets. Bootstrapped 95 percent confidence regions appear in angled brackets. For test statistics, corresponding p-values are presented.

TABLE 1 (Part 2): First-Pass Estimates of the Betas, Portfolios P5–P8

Portfolio (Pi)	Factor (f_j) Δc	Δd	r_W	Test of $\beta_{ij}=0\ \forall j$
P5	0.180	0.485	0.009	
	(1.006)	(0.722)	(0.065)	(0.740)
	[0.754]	[0.714]	[0.065]	[0.746]
	$\langle -1.5, 2.4\rangle$	$\langle -1.4, 1.5\rangle$	$\langle -0.15, 0.14\rangle$	
P6	−0.755	1.079	0.023	
	(1.089)	(0.781)	(0.071)	(0.556)
	[0.958]	[0.833]	[0.068]	[0.595]
	$\langle -2.2, 2.5\rangle$	$\langle -1.6, 1.6\rangle$	$\langle -0.14, 0.15\rangle$	
P7	0.036	1.234	−0.027	
	(1.044)	(0.749)	(0.068)	(0.101)
	[0.797]	[0.730]	[0.063]	[0.126]
	$\langle -1.2, 2.8\rangle$	$\langle -1.1, 1.7\rangle$	$\langle -0.21, 0.09\rangle$	
P8	−1.342	1.426	0.079	
	(1.674)	(1.201)	(0.108)	(0.684)
	[1.646]	[1.225]	[0.114]	[0.700]
	$\langle -6.0, 1.9\rangle$	$\langle -1.2, 2.8\rangle$	$\langle -0.18, 0.34\rangle$	

Notes: Annual data, 1953–2002. The regression equation is $R^e_{it} = a_i + \mathbf{f}_t'\boldsymbol{\beta}_i + \epsilon_{it}$, where R^e_{it} is the excess return of portfolio i at time t, $\mathbf{f}_t = (\ \Delta c_t \quad \Delta d_t \quad r_{Wt}\ )'$, 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 United States, where P1 and P8 are the portfolios with, respectively, the smallest and largest interest differentials. The table reports $\hat{\beta}_{ij}$ and p-values for tests of the hypotheses that $\beta_{ij} = 0\ \forall j$ for each portfolio i. For estimates of betas, OLS standard errors are in parentheses, and GMM-VARHAC standard errors are in square brackets. Bootstrapped 95 percent confidence regions appear in angled brackets. For test statistics, corresponding p-values are presented.

TABLE 2: First-Pass Estimates of the Betas, Portfolios DQ2–DQ6

Portfolio	Factor (j)			Test of
(i)	Δc	Δd	r_W	$\beta_{ij} = 0\ \forall j$
DQ2	0.072	−0.522	0.081	
	(0.722)	(0.609)	(0.034)	(0.073)
	[0.750]	[0.579]	[0.060]	[0.165]
DQ3	0.264	0.147	0.082	
	(0.770)	(0.650)	(0.037)	(0.132)
	[0.939]	[0.695]	[0.054]	[0.500]
DQ4	−0.745	0.060	0.091	
	(0.978)	(0.825)	(0.047)	(0.249)
	[1.098]	[0.833]	[0.055]	[0.298]
DQ5	0.043	−0.333	0.156	
	(0.948)	(0.799)	(0.045)	(0.005)
	[1.054]	[0.760]	[0.062]	[0.030]
DQ6	−0.399	0.162	0.161	
	(1.087)	(0.917)	(0.052)	(0.021)
	[1.186]	[0.862]	[0.072]	[0.163]

Notes: Quarterly data, 1976Q2–2010Q1. The regression equation is $R^e_{it} = a_i + \mathbf{f}_t'\boldsymbol{\beta}_i + \epsilon_{it}$, where R^e_{it} is the excess return of portfolio DQi at time t, $\mathbf{f}_t = (\ \Delta c_t \quad \Delta d_t \quad r_{Wt}\ )'$, 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 excess return to portfolio DQi is the difference between the quarterly excess returns of portfolios Qi and Q1. At the monthly frequency the excess returns of the Q-portfolios are the payoffs to holding long equally-weighted one-month forward positions in foreign currencies sorted by forward discount versus the US dollar. Q1 and Q6 are the portfolios with, respectively, the smallest and largest forward discounts. The table reports $\hat{\beta}_{ij}$ and p-values for tests of the hypotheses that $\beta_{ij} = 0\ \forall j$ for each portfolio i. For estimates of betas, OLS standard errors are in parentheses, and GMM-VARHAC standard errors are in square brackets. For test statistics, corresponding p-values are presented.

TABLE 3: Tests of the Rank of the Factor Beta Matrix

Test of H_0: rank($\boldsymbol{\beta}$) $= r$		Test of H_0: rank($\boldsymbol{\beta}^+$) $= r+1$	
r	p-value	$r+1$	p-value
2	(0.615) [0.657]	3	(0.529) [0.699]
1	(0.643) [0.444]	2	(0.618) [0.326]
0	(0.639) [0.000]	1	(0.562) [0.003]

Notes: Annual data, 1953–2002. The matrix $\boldsymbol{\beta}$ is obtained by running the regressions described in Table 1. The matrix $\boldsymbol{\beta}$ must have full column rank (3) for $\boldsymbol{\lambda}$ to be identified in the two-pass procedure without the constant. The matrix $\boldsymbol{\beta}^+ = (\ \boldsymbol{\iota} \quad \boldsymbol{\beta}\ )$ must have full column rank (4) for γ and $\boldsymbol{\lambda}$ to be identified in the two-pass procedure with the constant. The table presents tests of the null hypothesis that these rank conditions fail. The p-values for the tests are presented in parentheses (OLS standard errors) and square brackets (GMM-VARHAC standard errors).

TABLE 4: Tests of the Rank of the Factor Beta Matrix with Different Test Assets

Tests of H_0: rank($\boldsymbol{\beta}$) = r			
r	Currencies & Equities	Currencies, Equities & Bonds	Differenced Currencies 1953-2002
2	(0.422) –	(0.732) –	(0.529) [0.706]
1	(0.415) –	(0.016) –	(0.618) [0.134]
0	(0.000) –	(0.000) –	(0.562) [0.001]

Tests of H_0: rank($\boldsymbol{\beta}^+$) = $r+1$			
$r+1$	Currencies & Equities	Currencies, Equities & Bonds	Differenced Currencies 1953-2002
3	(0.449) –	(0.728) –	(0.406) [0.626]
2	(0.497) –	(0.182) –	(0.612) [0.339]
1	(0.000) –	(0.000) –	(0.603) [0.005]

Notes: See the note to Table 3. P-values for rank tests are presented in parentheses (OLS standard errors) and square brackets (GMM-VARHAC standard errors). Dashes indicate cases where the GMM-VARHAC standard errors cannot be calculated because the number of parameters in the $\boldsymbol{\beta}$ matrix plus the number of equations (to account for the constants in the first pass regressions) exceeds the sample size.

TABLE 5: GMM Estimates of the Model with no Constant
Currency and Equity Portfolios as Test Assets

	(a) 1st Stage		(b) 2nd Stage		(c) Iterated GMM	
Factor	$\hat{\mathbf{b}}$	$\hat{\boldsymbol{\lambda}}$	$\hat{\mathbf{b}}$	$\hat{\boldsymbol{\lambda}}$	$\hat{\mathbf{b}}$	$\hat{\boldsymbol{\lambda}}$
Δc	113.9	2.38	84.6	2.57	60.7	2.53
	(82.1)	(1.23)	(55.4)	(0.62)	(51.9)	(0.98)
Δd	−4.89	1.80	33.2	2.68	60.7	3.50
	(65.8)	(2.03)	(27.9)	(0.88)	(45.6)	(1.59)
r_W	2.11	11.2	4.19	13.7	2.60	5.80
	(3.37)	(3.87)	(1.71)	(6.2)	(2.14)	(5.06)
R^2 for currencies	0.03		−0.06		−0.76	
MAE for currencies	1.44		1.37		1.88	

Notes: Annual data, 1953–2002. The table reports GMM estimates of $\mathbf{b}$ and $\boldsymbol{\lambda}$ obtained by exploiting the moment restrictions $E\{\mathbf{R}_t^e[1-(\mathbf{f}_t-\boldsymbol{\mu})'\,\mathbf{b}]\}=0$, $E(\mathbf{f}_t-\boldsymbol{\mu})=\mathbf{0}$ and $E[(\mathbf{f}_t-\boldsymbol{\mu})(\mathbf{f}_t-\boldsymbol{\mu})'-\boldsymbol{\Sigma}_f]=0$, where $\mathbf{R}_t^e$ is a vector of excess returns that includes the currency portfolios described in Table 1, as well as Fama and French's (1993) six equity portfolios created by sorting stocks on the basis of size and value, $\mathbf{f}_t=(\ \Delta c_t\quad \Delta d_t\quad r_{Wt}\ )'$, 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. GMM-VARHAC standard errors are in parentheses. The R^2 statistic and mean absolute pricing error (MAE) are presented for currency portfolios only, and are comparable to the statistics in Tables 7 and 8 in the main text.

TABLE 6: GMM ESTIMATES OF THE MODEL WITH NO CONSTANT
CURRENCY, EQUITY AND BOND PORTFOLIOS AS TEST ASSETS

	(a) 1st Stage		(b) 2nd Stage		(c) Iterated GMM	
Factor	$\hat{\mathbf{b}}$	$\hat{\boldsymbol{\lambda}}$	$\hat{\mathbf{b}}$	$\hat{\boldsymbol{\lambda}}$	$\hat{\mathbf{b}}$	$\hat{\boldsymbol{\lambda}}$
Δc	89.3	1.67	91.1	1.80	17.1	0.04
	(61.1)	(1.07)	(33.5)	(0.46)	(23.2)	(0.80)
Δd	−14.7	0.94	−12.2	1.00	−28.8	−1.33
	(52.5)	(1.73)	(20.7)	(0.67)	(22.8)	(0.74)
r_W	1.92	10.4	3.04	13.7	6.09	21.6
	(2.64)	(3.24)	(1.24)	(5.5)	(1.55)	(10.8)
R^2 for currencies	0.03		0.05		−1.35	
MAE for currencies	1.40		1.42		1.64	

Notes: Annual data, 1953–2002. The table reports GMM estimates of $\mathbf{b}$ and $\boldsymbol{\lambda}$ obtained by exploiting the moment restrictions $E\{\mathbf{R}_t^e[1-(\mathbf{f}_t-\boldsymbol{\mu})'\,\mathbf{b}]\}=0$, $E(\mathbf{f}_t-\boldsymbol{\mu})=\mathbf{0}$ and $E[(\mathbf{f}_t-\boldsymbol{\mu})(\mathbf{f}_t-\boldsymbol{\mu})'-\boldsymbol{\Sigma}_f]=0$, where $\mathbf{R}_t^e$ is a vector of excess returns that includes the currency portfolios described in Table 1, the equity portfolios described in Table 5, and five Fama bonds portfolios sorted by maturity (from the Center for Research in Securities Prices, 2007), $\mathbf{f}_t=(\ \Delta c_t \quad \Delta d_t \quad r_{Wt}\ )'$, 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. GMM-VARHAC standard errors are in parentheses. The R^2 statistic and mean absolute pricing error (MAE) are presented for currency portfolios only, and are comparable to the statistics in Tables 7 and 8 in the main text.