# Asset Pricing with Consumption and Robust Inference

Frank Kleibergen* Zhaoguo Zhan†

January 9, 2018

**Abstract**

Alongside the National Income and Product Accounts (NIPA), we consider four consumption measures recently proposed for asset pricing in the *Journal of Finance* and the *Journal of Political Economy*: Parker and Julliard (2005), Jagannathan and Wang (2007), Savov (2011) and Kroencke (2017). We show that their correlations with asset returns are not sufficiently large to precisely identify the risk premium and relative risk aversion. This invalidates traditional two-pass and GMM inference methods so the resulting findings cannot be trusted. Inference methods that remain valid irrespective of the magnitude of the correlations result in unbounded confidence sets for the risk premium and relative risk aversion.

JEL Classification: G12

Keywords: asset pricing; consumption; relative risk aversion; risk premium; robust inference

*Amsterdam School of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. Email: f.r.kleibergen@uva.nl.

†Department of Economics, Finance and Quantitative Analysis, Coles College of Business, Kennesaw State University, GA 30144, USA. Email: zzhan@kennesaw.edu.

# 1 Introduction

The classical consumption-based asset pricing theory relates asset returns to consumption risk. Yet a worrisome phenomenon is that there exist various consumption measures in the literature leading to incongruous empirical findings. It has been well documented that the canonical consumption measure from the National Income and Product Accounts (NIPA) induces small correlations between consumption growth and asset returns, which could be improved by adopting alternative consumption measures. See, e.g., the three-year consumption measure in Parker and Julliard (2005), the fourth-quarter to fourth-quarter consumption measure in Jagannathan and Wang (2007), the garbage measure in Savov (2011) and the unfiltered NIPA consumption measure in Kroencke (2017).

On the one hand, as consumption measures vary, we expect the resulting correlation between consumption growth and asset returns to exhibit subtle differences. The magnitude of this correlation further affects standard asset pricing tests in, e.g., the Fama-MacBeth (1973) (FM) two-pass procedure and the Generalized Method of Moments (GMM) framework of Hansen (1982). Outcomes of these tests therefore appear to favor some consumption measures for asset pricing over others. See, e.g., Kroencke (2017).

On the other hand, it is well-known that the reliability of standard asset pricing tests depends on the statistical quality of consumption measures, or more generally, risk factors. For instance, Kan and Zhang (1999) and Kleibergen (2009) warn that the $t$-test in the FM two-pass procedure can spuriously favor risk factors that are independent of or weakly correlated with asset returns, respectively. Similar spurious outcomes also exist in the GMM setting, see, e.g., Stock and Wright (2000), Kleibergen (2005), Gospodinov *et al*. (2017). It is therefore important not to misinterpret such spurious outcomes as evidence in support of factor pricing.

In this paper, we show that the correlation of the aforementioned five consumption measures with asset returns is not sufficiently large to rule out spurious outcomes of standard asset pricing tests. Put differently, consumption growth resulting from these measures is statistically similar to, albeit different from, the useless factor analyzed in Kan and Zhang (1999). It leads to seemingly promising empirical findings in both the FM two-pass procedure and GMM. These findings, however, are more likely caused by statistical failure of standard asset pricing tests, rather than the superior quality of consumption measures.

Instead of using standard asset pricing tests, we re-examine all the five consumption measures by means of identification robust econometric tests. Unlike standard tests, these robust tests remain valid regardless of the magnitude of the correlation between the consumption measure and asset returns. When these correlations are sizeable, the robust tests are asymptotically equivalent to the conventional $t$-test. When the correlations are small, this is, however, not so. The robust tests are then still reliable while the $t$-test is not. The robust tests are well-established in the econometrics literature but, to the best of our knowledge, their usage for comparing consumption measures for asset pricing is new.

In particular, we employ the robust Factor-Anderson-Rubin (FAR) test from Kleibergen (2009) for the risk premium on consumption growth in the FM two-pass procedure, jointly with the robust GMM-AR test from Stock and Wright (2000) for the relative risk aversion in the Euler equation. The non-robust counterparts of the FAR and GMM-AR tests are the conventional $t$-tests in the FM two-pass procedure and GMM, respectively. We focus on the risk premium and relative risk aversion, since both of them are essential to establish credibility of existing consumption measures for asset pricing. See, e.g., Savov (2011), Kroencke (2017).

Interestingly, for all consumption measures examined in this paper, we find that the 95% confidence sets from the FAR and GMM-AR tests for the risk premium and relative risk aversion are unbounded. These unbounded confidence sets obviously contain the corresponding estimates documented in the existing literature, but they also show that the correlations between the consumption measures and asset returns are not sufficiently large to precisely identify the risk premium and the relative risk aversion. Our findings thus cast doubt on the claimed ability of these consumption measures to explain the equity premium puzzle, the risk-free rate puzzle, etc. On the bright side, however, we find empirical evidence that the garbage measure in Savov (2011) and the unfiltered NIPA consumption in Kroencke (2017) minorly help to address the risk-free rate puzzle, while the other three measures do not.

The rest of the paper proceeds as follows. The consumption growth data used in our empirical analysis and our findings on the risk premium from the robust FAR test are presented in Section 2. Section 2 also addresses misspecification of the factor pricing moment equation which can as well be evaluated using the FAR test. Section 3 contains our results for the relative risk aversion from the robust GMM-AR test and the implications for the risk-free rate puzzle. Section 4 concludes. The necessary technical details are in the Appendix. Throughout the paper, we use $C_t$ for consumption at time $t$, $\triangle c_t = \ln(\frac{C_{t+1}}{C_t})$ for consumption growth and $R_t$ for asset returns.

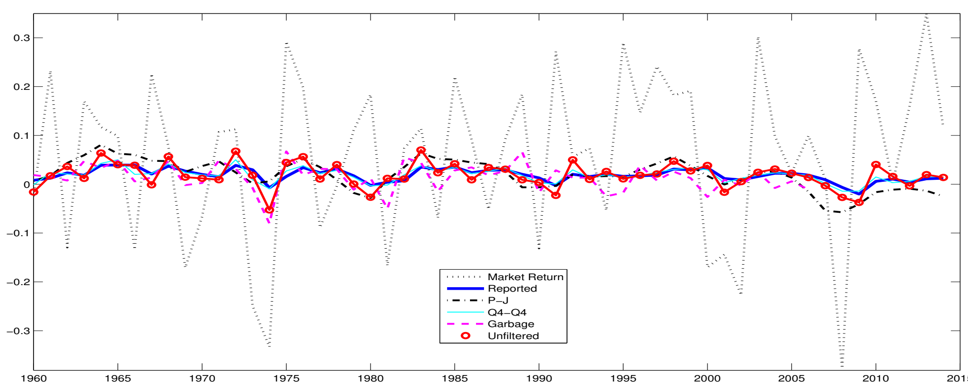
# 2 Risk premium on consumption growth

## 2.1 Data description

Various measures of consumption exist in the asset pricing literature. This paper considers five of them: (i) the consumption expenditure on nondurable goods and services reported by NIPA ("**Reported**"); (ii) the three-year consumption measure from Parker and Julliard (2005) ("**P-J**"); (iii) the fourth-quarter to fourth-quarter consumption measure from Jagannathan and Wang (2007) ("**Q4-Q4**"); (iv) the garbage measure in Savov (2011) ("**Garbage**"); (v) the unfiltered NIPA consumption measure in Kroencke (2017) ("**Unfiltered**").

To facilitate comparison, we use the same data as in Savov (2011) and Kroencke (2017).[1] The annual garbage series is from 1960-2007 as in Savov (2011), while we use the other four consumption measures in the postwar sample, 1960-2014, as in Kroencke (2017). Figure 1 presents the resulting five consumption growth rates, jointly with the excess market return for comparison.

Figure 1: Consumption growth and market return




As shown in Figure 1, all consumption growth rates are much less volatile than the market return, although some growth rates appear slightly more volatile than others. This is also reflected by the large standard error 16.89 of the market return, compared to the much smaller ones of consumption growth in the 1.30 ("**Reported**") $\sim$ 3.06 ("**P-J**") range.[2] Overall, Figure 1 indicates that the comovement between consumption growth and the market return is not large.

[1]We thank Alexi Savov and Tim A. Kroencke for sharing their data.

[2]These standard errors are scaled by 100, as reported in Kroencke (2017).

## 2.2 Risk premium: $\lambda_c$

The risk premium on consumption growth is commonly tested in the following beta representation implied by the consumption-based asset pricing model (see, e.g., Breeden *et al*. 1989):

$$E(R^e_{i,t+1}) = \lambda_0 + \beta_i \lambda_c, \; i = 1, ..., N, \tag{1}$$

where $\lambda_0$ is the zero-beta return, $\lambda_c$ is the risk premium, $\beta_i$ results from the covariance of consumption growth and the $i$-th asset return divided by the variance of consumption growth, and $R^e_{i,t+1}$ is the excess return on the $i$-th asset.

The FM two-pass procedure estimates $\lambda_c$ by regressing the average returns on the $N$-dimensional vector of ones, $\iota_N$, and the $N$-dimensional vector of $\beta$'s, $\beta = [\beta_1, ...\beta_N]'$, estimated in a first step time series regression of the portfolio returns on consumption growth. The regression estimating $\lambda_c$ is not allowed to be subject to multicollinearity which would result when $\iota_N$ and $\beta$ are proportional to each other, or put differently when the $N \times 2$ matrix $[\iota_N, \beta]$ is not of full rank.

It is widely acknowledged that consumption is poorly measured. The quality of the consumption data further jeopardizes the rank condition of the FM two-pass procedure that the $N \times 2$ matrix $[\iota_N, \beta]$ is of full rank. In the extreme case that the empirically used consumption growth measure is just uncorrelated noise, $\beta$ reduces to zero, so the rank condition fails. Similarly, when $\beta$ is non-zero but small, the FM two-pass procedure also breaks down, see Kleibergen (2009). As suggested in Kleibergen and Paap (2006), Burnside (2016), etc., a rank test on $[\iota_N, \beta]$ can be employed to diagnose if $[\iota_N, \beta]$ has a reduced rank value.

Following Kroencke (2017), we use $N = 31$ portfolios sorted by size, value, and investment, plus the equity premium as test assets. The estimated $\beta$'s from the first pass time series regression are presented in Table 1, using all five consumption measures, respectively. It is obvious that the reported values for $\beta$ in Table 1 exhibit substantial differences as consumption measures vary. In particular, the estimated $\beta$ from the reported NIPA consumption overall appears much smaller than its counterparts from alternative consumption measures such as the garbage measure in Savov (2011) and the unfiltered NIPA consumption in Kroencke (2017). This is also reflected by the pseudo-$R^2$ reported in Table 1, which measures the percentage of the total variation of asset returns that is explained by consumption growth.

Table 1: $\beta$ with 31 portfolio returns

| | **Reported** | | **P-J** | | **Q4-Q4** | | **Garbage** | | **Unfiltered** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\hat{\beta}$ | $t$-stat | $\hat{\beta}$ | $t$-stat | $\hat{\beta}$ | $t$-stat | $\hat{\beta}$ | $t$-stat | $\hat{\beta}$ | $t$-stat |
| (1) | -1.30 | -0.41 | 1.19 | 0.93 | 4.18 | 1.51 | 4.54 | 3.33 | 5.40 | 3.81 |
| (2) | -0.23 | -0.09 | 1.14 | 1.06 | 4.42 | 1.93 | 4.06 | 3.52 | 4.70 | 3.98 |
| (3) | 0.14 | 0.06 | 0.99 | 1.06 | 4.15 | 2.09 | 3.74 | 3.86 | 3.80 | 3.64 |
| (4) | -0.52 | -0.25 | 0.73 | 0.86 | 3.46 | 1.92 | 3.66 | 4.30 | 3.33 | 3.50 |
| (5) | -0.23 | -0.11 | 0.73 | 0.85 | 3.74 | 2.06 | 3.62 | 4.30 | 3.33 | 3.44 |
| (6) | -0.84 | -0.43 | 0.58 | 0.72 | 3.01 | 1.76 | 3.42 | 4.38 | 3.10 | 3.43 |
| (7) | -0.85 | -0.44 | 0.49 | 0.63 | 2.73 | 1.63 | 3.51 | 4.90 | 2.95 | 3.32 |
| (8) | -0.59 | -0.33 | 0.39 | 0.53 | 2.69 | 1.69 | 3.06 | 4.55 | 2.85 | 3.41 |
| (9) | -0.12 | -0.07 | 0.30 | 0.43 | 2.75 | 1.87 | 2.61 | 3.98 | 2.48 | 3.16 |
| (10) | 0.34 | 0.20 | 0.64 | 0.92 | 2.97 | 2.01 | 3.14 | 4.53 | 2.60 | 3.29 |
| (11) | -0.95 | -0.46 | 0.21 | 0.25 | 2.25 | 1.22 | 3.58 | 4.29 | 2.85 | 2.91 |
| (12) | -0.44 | -0.25 | 0.60 | 0.81 | 2.77 | 1.77 | 3.15 | 4.22 | 2.59 | 3.09 |
| (13) | -0.63 | -0.38 | 0.52 | 0.78 | 2.49 | 1.73 | 2.95 | 4.38 | 2.22 | 2.84 |
| (14) | -0.56 | -0.31 | 0.07 | 0.10 | 2.26 | 1.39 | 2.51 | 3.10 | 2.04 | 2.28 |
| (15) | 0.88 | 0.53 | 0.46 | 0.67 | 3.14 | 2.19 | 2.75 | 3.89 | 2.25 | 2.85 |
| (16) | 0.73 | 0.45 | 0.58 | 0.88 | 3.04 | 2.19 | 2.86 | 4.49 | 2.62 | 3.56 |
| (17) | 0.96 | 0.56 | 0.72 | 1.03 | 3.50 | 2.40 | 2.93 | 4.53 | 2.81 | 3.61 |
| (18) | 1.21 | 0.65 | 1.05 | 1.40 | 4.27 | 2.72 | 3.17 | 4.35 | 3.33 | 4.02 |
| (19) | 1.23 | 0.67 | 1.09 | 1.47 | 3.97 | 2.56 | 2.86 | 3.65 | 2.96 | 3.53 |
| (20) | 1.77 | 0.84 | 1.32 | 1.56 | 5.22 | 2.98 | 2.93 | 3.26 | 4.21 | 4.67 |
| (21) | -0.50 | -0.22 | 0.58 | 0.61 | 3.57 | 1.78 | 3.85 | 4.30 | 3.88 | 3.73 |
| (22) | -0.09 | -0.05 | 0.66 | 0.88 | 3.14 | 1.98 | 3.00 | 4.10 | 3.15 | 3.82 |
| (23) | 0.37 | 0.23 | 0.54 | 0.81 | 2.84 | 2.02 | 2.94 | 4.49 | 2.27 | 2.96 |
| (24) | 0.44 | 0.31 | 0.54 | 0.92 | 2.96 | 2.40 | 2.84 | 5.30 | 2.25 | 3.38 |
| (25) | -0.50 | -0.32 | 0.31 | 0.49 | 2.11 | 1.57 | 2.69 | 5.05 | 2.22 | 3.10 |
| (26) | 1.04 | 0.62 | 0.79 | 1.16 | 3.70 | 2.60 | 3.09 | 4.44 | 2.54 | 3.25 |
| (27) | 1.10 | 0.66 | 0.98 | 1.46 | 3.71 | 2.63 | 2.90 | 4.13 | 2.74 | 3.60 |
| (28) | 0.03 | 0.01 | 0.71 | 0.98 | 3.18 | 2.07 | 3.15 | 4.32 | 3.07 | 3.83 |
| (29) | -0.89 | -0.41 | 0.17 | 0.19 | 2.44 | 1.26 | 3.54 | 3.82 | 2.70 | 2.59 |
| (30) | -1.40 | -0.58 | 0.03 | 0.03 | 2.10 | 0.97 | 3.92 | 3.95 | 2.92 | 2.52 |
| (31) | 0.23 | 0.13 | 0.72 | 1.00 | 3.32 | 2.21 | 3.21 | 4.74 | 2.95 | 3.72 |
| Pseudo-$R^2$ | 0.003 | | 0.015 | | 0.067 | | 0.270 | | 0.182 | |
| $p$-value of testing $rank([\iota_N, \beta]) = 1$ | 0.37 | | 0.57 | | 0.50 | | 0.84 | | 0.53 | |

Note: The test assets are as in Kroencke (2017), i.e., the 31 portfolios sorted by size, value, and investment, plus the equity premium in 1960-2014. The reported pseudo-$R^2$ is a goodness of fit measure that reflects the percentage of the variation of asset returns that is explained by consumption growth, see Kleibergen and Zhan (2015). More details about the rank test can be found in Kleibergen and Paap (2006).

Table 1 also presents the $p$-value of the Kleibergen and Paap (2006) rank test for testing $H_0 : [\iota_N, \beta]$ has reduced rank one. The large $p$-values, however, imply that the rank condition of the FM two-pass procedure is under doubt for all five consumption measures. Note that if $\beta$ is close to zero, the rank condition is jeopardized. Similarly, if $\beta$ is sizeable and significant, but its elements are alike so $\beta$ is close to be spanned by $\iota_N$, as is the case for the "**Garbage**" and "**Unfiltered**" consumption measures, the rank condition is also at risk.

Table 2: Risk premium $\lambda_c$ with 31 portfolio returns

| | **Reported** | **P-J** | **Q4-Q4** | **Garbage** | **Unfiltered** |
|---|---|---|---|---|---|
| Estimate of $\lambda_c$ | 0.37 | 4.42 | 2.00 | 1.71 | 2.04 |
| FM $t$ | 0.73 | 3.04 | 2.98 | 1.21 | 2.18 |
| Shanken $t$ | 0.70 | 1.83 | 1.81 | 1.05 | 1.75 |
| $R^2$ | 0.03 | 0.59 | 0.64 | 0.15 | 0.67 |
| FACCHECK | 93% | 93% | 93% | 90% | 92% |
| 95% C.I. of $\lambda_c$, FAR | $(-\infty, \infty)$ | $(-\infty, \infty)$ | $(-\infty, \infty)$ | $(-\infty, \infty)$ | $(-\infty, \infty)$ |

Note: The test assets are as in Kroencke (2017), i.e., the 31 portfolios sorted by size, value, and investment, plus the equity premium in 1960-2014. The estimate of $\lambda_c$, $t$-statistics and the cross-sectional $R^2$ result from the Fama-MacBeth (1973) two-pass procedure. FACCHECK equals the percentage of the variation of residuals explained by the three largest principal components. The FAR test is from Kleibergen (2009).

The second pass cross-sectional estimation results for the risk premium $\lambda_c$ are presented in Table 2, where our estimates of $\lambda_c$, as well as the Shanken (1992) $t$-statistics and the cross-sectional $R^2$'s, are identical to those in Kroencke (2017). Table 2 shows that all estimates of $\lambda_c$ are positive under the five consumption measures, although their significance differs. However, our findings on $\beta$ in Table 1 put the estimates and $t$-statistics of $\lambda_c$ in Table 2 under doubt, since the sizeable weak factor literature has demonstrated that both the estimates and $t$-statistics can be spurious, if the rank condition is jeopardized. See Kan and Zhang (1999), Kleibergen (2009). Similarly, Kleibergen and Zhan (2015) show that the cross-sectional $R^2$ can also be spuriously large under weak or useless factors when a strong factor structure is present in the residuals of the first pass time series regression of the FM two-pass procedure. They propose a measure/check for the unexplained factor structure (with three factors) left in the residuals of the time series regression of the FM two-pass procedure. This measure equals the fraction of the total variation that is explained by the three largest principal components:

$$FACCHECK = \frac{v_1 + v_2 + v_3}{v_1 + ... + v_N} \tag{2}$$

with $v_1 > v_2 > ... > v_N$ the characteristic roots of the covariance matrix in descending order, see Kleibergen and Zhan (2015).[3]

The factor structure check for the $N = 31$ test assets equals 93%, so 93% of the variation in the test assets is explained by the three largest principal components. This implies that there is a strong factor structure. Table 2, however, shows that the factor structure checks for the residuals are also around 90% $\sim$ 93%, which implies that the five consumption growth rates do not explain any of the main factors of the asset returns. This shows that the large $R^2$'s in Table 2 do not have to indicate strong factor pricing.

### 2.3 Factor Anderson-Rubin (FAR) test

In Kleibergen (2009) the FAR statistic is proposed to conduct tests on the risk premia. Its expression and limiting distribution are stated in Appendix A. They show that the FAR test is easy to use in practice. Unlike the $t$-test, the limiting distribution of the FAR test does not depend on the rank of $[\iota_N, \beta]$. It therefore remains trustworthy when the quality of the consumption measures is unsatisfactory so the rank condition is jeopardized. When the full rank condition of $[\iota_N, \beta]$ is satisfied, i.e., consumption growth is a strong factor, the FAR test is similar to the $t$-test. On the other hand, when consumption growth is a weak factor that jeopardizes the rank condition, the FAR test remains, unlike the $t$-test, reliable. In addition, the FAR test also has power against the model misspecification analyzed by Kan and Zhang (1999) and Gospodinov *et al.* (2017), which we further discuss in the next subsection. For ease of illustration, we conduct a simple simulation experiment.

For a data generating process (D.G.P.), we consider

$$R_t = \iota_N \lambda_0 + \beta(\triangle c_t + \lambda_c) + \epsilon_t, \; t = 1, 2, ..., T, \tag{3}$$

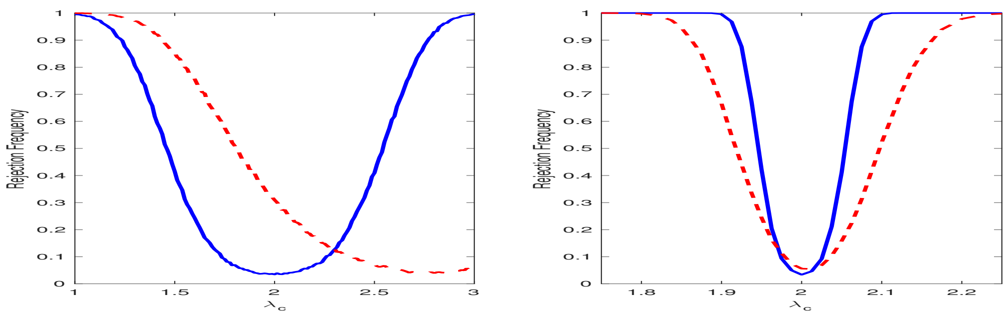
where $\triangle c_t \sim NID(0, V_c)$, $\epsilon_t \sim NID(0, \Omega_\epsilon)$, and $V_c$, $\Omega_\epsilon$, $\lambda_0$, $\beta$ are calibrated from consumption growth data in Parker and Julliard (2005) and the $N = 31$ test assets in Kroencke (2017). With the simulated data, we test $\lambda_c = 2$, which is close to the average of the five estimates reported in Table 2, by using the $t$-test with the Shanken (1992) correction and the FAR test.

Simulated power curves of $t$ and FAR tests are shown in Figure 2, with $T = 200$ and the number of Monte Carlo replications is 10000. For the left hand side panel of Figure 2,

[3] As explained in Kleibergen and Zhan (2015), this corresponds to using the trace norm of the covariance matrix as a measure of the total variation.

we calibrate $\beta$ to the estimate $\hat{\beta}$ that results from using consumption growth in Parker and Julliard (2005) and test assets in Kroencke (2017), while for the right hand side panel, we use $10 \times \hat{\beta}$ for the $\beta$ in the D.G.P.

Figure 2: Power curves of $t$ and FAR that test $\lambda_c = 2$ at the 5% level




Note: Power curves of FAR (solid) and $t$ (dash) with the Shanken (1992) correction. Left: $\beta = \hat{\beta}$; Right: $\beta = 10 \times \hat{\beta}$.

The left hand side panel shows that when $\beta$ is small, the $t$-test suffers from size distortion since the null is over rejected (i.e., its rejection frequency is 31% under the null $\lambda_c = 2$), while the FAR test remains size-correct (i.e., its rejection frequency is close to the nominal 5%). In contrast, when the magnitude of $\beta$ is amplified in the right hand side panel, both the $t$ and FAR tests are size-correct (i.e., under the null, their rejection frequencies are both close to the nominal 5%), and have comparable power.

Overall, Figure 2 indicates that the FAR test is more reliable than the $t$-test.[4] We refer to the malfunction of the $t$-test induced by the small magnitude of $\beta$ as a weak identification problem, which the FAR test is robust to.

## 2.4 Misspecification

Another related strand of literature concerns model misspecification, i.e., the moment condition in Equation (1) for the risk premium does not hold, alongside weak or no identification of the risk premia. See, e.g., Kan and Zhang (1999) and Gospodinov *et al*. (2017). To illustrate how the FAR test signals the presence or absence of misspecification in a weak

[4]More simulation results can be found in Kleibergen (2009).

identification setting, we consider an alternative D.G.P.:

$$\left[\begin{array}{c} \triangle c_t \\ R_t \end{array}\right] \sim NID\left(\left[\begin{array}{c} \mu_c \\ \mu_R \end{array}\right], \left[\begin{array}{cc} V_c & \theta \cdot V_{cR} \\ V_{Rc} \cdot \theta & V_R \end{array}\right]\right) \tag{4}$$

where the mean $\mu_c$, $\mu_R$, variance $V_c$, $V_R$ and covariance $V_{cR}$, $V_{Rc}$ are calibrated from consumption growth in Parker and Julliard (2005) and the $N = 31$ test assets in Kroencke (2017), $\theta$ is a scalar so the resulting $\beta$ equals $V_{Rc}V_c^{-1} \cdot \theta$.

The pseudo risk premia can be written as

$$\left[\begin{array}{c} \lambda_0 \\ \lambda_c \end{array}\right] = (X'X)^{-1}X'\mu_R, \ \textit{where } X = [\iota_N, \beta]. \tag{5}$$

Thus the moment condition $\mu_R = \iota_N \lambda_0 + \beta\lambda_c$ corresponds to $\mu_R = X(X'X)^{-1}X'\mu_R$. When this equation fails, there is misspecification.

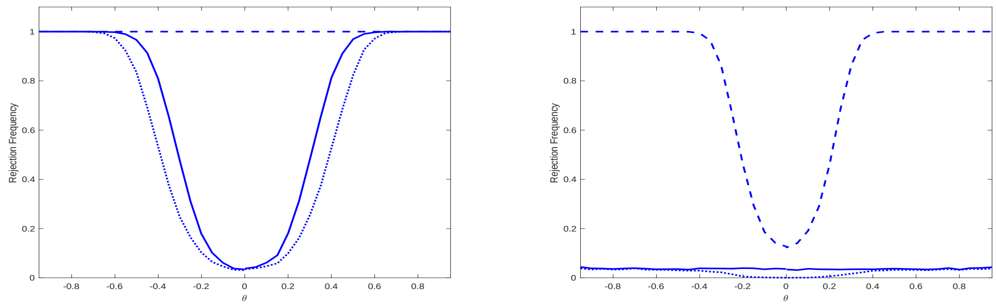
In the D.G.P. described by (4), misspecification exists, since $\mu_R$, $\beta$, and thus also $X$, are all calibrated to their estimates $\hat{\mu}_R$, $\hat{\beta}$, $\hat{X} = [\iota_N, \hat{\beta}]$ from the FM two-pass procedure so $\hat{\mu}_R \neq \hat{X}(\hat{X}'\hat{X})^{-1}\hat{X}'\hat{\mu}_R$. Scaling $\hat{\beta}$ by $\theta$ does not alter the inequality. Consequently, $\mu_R \neq X(X'X)^{-1}X'\mu_R$ in the current D.G.P. and there is misspecification.

To facilitate comparison, we also consider a correctly specified D.G.P. by explicitly imposing $\mu_R = \iota_N \lambda_0 + \beta\lambda_c$, or equivalently, $\mu_R = X(X'X)^{-1}X'\mu_R$. This is achieved by setting $\mu_R$ equal to $\hat{X}(\hat{X}'\hat{X})^{-1}\hat{X}'\hat{\mu}_R$.

We illustrate the power of the FAR test when the magnitude of $\beta$ varies using values of $\theta$ from $-1$ to 1, jointly with misspecification. With the simulated data, we test the hypothesis that $\lambda_c$ is equal to its pseudo true value denoted by $\lambda_{c,0}$, i.e., the second element in $(X'X)^{-1}X'\mu_R$. The resulting rejection frequencies of the FAR test are presented in Figure 3. Specifically, we use three different implementations of the FAR statistic: the $FAR(\lambda_{c,0})$ statistic (solid) as stated in (12) in Appendix A, $min_{\lambda_c}FAR(\lambda_c)$ (dotted) and $max_{\lambda_c}FAR(\lambda_c)$ (dash). We separately discuss each one of them:

$FAR(\lambda_{c,0})$ **(solid):** Panel (a) of Figure 3 shows that the FAR test primarily rejects the null hypothesis for values of $\theta$ away from zero and rejects 5% when $\theta$ is equal to zero. When $\theta$ equals zero, the risk premium is not identified so the tested value of $\lambda_c$ is infinite, since $\beta$ equals zero. In Appendix B, it is shown that the FAR statistic then converges to a $\chi^2_{N-1}$ distributed random variable when $T$ goes to infinity despite that there is misspecification.

Figure 3: Rejection frequencies of the FAR test against misspecification at the 5% level




(a) Misspecification (b) Correct Specification

Note: The FAR test is used to test $\lambda_c$ equal to its pseudo value $\lambda_{c,0}$ at the 5% level. $\theta$ is the scaling factor of $\beta$. Rejection frequencies are plotted for three statistics: $FAR(\lambda_{c,0})$ (solid), $min_{\lambda_c} FAR(\lambda_c)$ (dotted), $max_{\lambda_c} FAR(\lambda_c)$ (dash). Panel (a): the model is misspecified; Panel (b): the model is correctly specified. $T = 200$ and the number of Monte Carlo replications is 10000. Critical values result from $\chi^2_{N-1}$ for $FAR(\lambda_{c,0})$ and $max_{\lambda_c} FAR(\lambda_c)$ and from $\chi^2_{N-2}$ for $min_{\lambda_c} FAR(\lambda_c)$.

This explains the 5% rejection frequency when $\theta$ is equal to zero since we use a 5% significance level. When $\theta$ differs from zero, the FAR test rejects and appropriately signals that the model is misspecified.

There is no misspecification in the D.G.P. used for Panel (b) so the rejection frequency of the FAR test is close to the nominal 5% for every value of $\theta$. This is in line with Figure 2, which shows that the FAR test remains size-correct when the magnitude of $\beta$ varies.

$min_{\lambda_c} FAR(\lambda_c)$ **(dotted):** The minimal value of the FAR statistic over $\lambda_c$, $min_{\lambda_c} FAR(\lambda_c)$, equals the $J$-statistic used in GMM to test for misspecification, see, e.g., Hansen (1982). Theorem 2 of Gospodinov *et al.* (2017) then implies that $min_{\lambda_c} FAR(\lambda_c)$ is bounded from above by a $\chi^2_{N-2}$ distributed random variable when $\theta$ equals zero.[5] The $J$-statistic is therefore not able to detect misspecification when $\theta$ equals zero. Panel (a) of Figure 3 shows that for other values of $\theta$, the $J$-statistic appropriately indicates that there is misspecification.

The D.G.P. used for Panel (b) in Figure 3 has no misspecification so the $J$-statistic, or $min_{\lambda_c} FAR(\lambda_c)$, rejects at most 5% for all values of $\theta$.

$max_{\lambda_c} FAR(\lambda_c)$ **(dash):** The 95% confidence set for the risk premium that results from the FAR test contains all values of the risk premium for which the FAR test does not reject

[5]The $J$-statistic for testing misspecification in the linear factor model equals the smallest root of a characteristic polynomial. In Guggenberger *et al.* (2012), the limiting distribution of the $J$-statistic for the linear instrumental variables regression model, which equals the smallest root of an identical characteristic polynomial, is shown to be bounded in an equivalent manner.

at the 5% level. This confidence set can be unbounded when there is no misspecification and the risk premia are weakly identified. The maximal value of the FAR statistic over all values of $\lambda_c$, $max_{\lambda_c}FAR(\lambda_c)$, then shows if any values of the risk premium are rejected so if the confidence set of the risk premium is unbounded.

Panel (a) of Figure 3 shows that the rejection frequency of $max_{\lambda_c}FAR(\lambda_c)$ equals one for all values of $\theta$. Hence, the confidence set of the risk premium that results from the FAR statistic excludes specific values of the risk premium for all values of $\theta$. This even holds true when $\theta$ equals zero where the FAR and $J$-tests only have 5% rejection frequencies. The pseudo true value of the risk premium tested using the FAR test is very large under small $\theta$'s, so the 95% confidence set is peculiar since it includes large values of the risk premium but excludes much smaller ones. This is indicative of the misspecification issues at hand.

The D.G.P. used for Panel (b) of Figure 3 has no misspecification. The rejection frequencies of $max_{\lambda_c}FAR(\lambda_c)$ now indicate that when $\theta$ equals zero, over 80% of the generated samples have no value of the risk premium rejected using the FAR test at the 5% level. Hence, for all these cases the 95% confidence sets for the risk premium that result from the FAR test equal the real line.

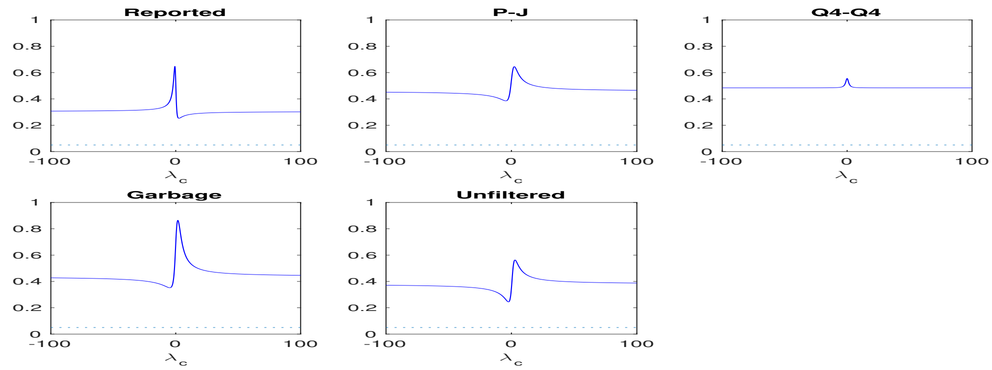
The above shows how the FAR test can be used to signal misspecification even when the risk premia are weakly identified. It also shows that a 95% confidence set of the risk premium from the FAR test which equals the real line does not occur in case of misspecification.

### 2.5 Analyzing consumption risk premia using the FAR test

Table 2 presents the 95% confidence sets for the risk premium that result from the FAR test, i.e., the tested values of $\lambda_{c,0}$ such that the FAR statistic does not exceed the upper 5% quantile of the $\chi^2_{N-1}$ distribution. These sets are found to be unbounded, hence little information about $\lambda_c$ is present in the consumption data. Their unboundedness also indicates that no misspecification is present. The robust FAR confidence sets result from the $p$-value plots of the FAR test for testing $\lambda_c$ equal to the value on the horizontal axis in Figure 4. A $p$-value larger than 0.05 implies that the hypothesized value of $\lambda_c$ on the horizontal axis is not rejected at the 5% significance level.

It is worth noting that the FAR 95% confidence sets contain the estimates and the conventional 95% confidence sets from the $t$-test for $\lambda_c$ in Table 2. Therefore, the findings from previous studies are not rejected by the FAR test. Instead, the unbounded sets from the FAR test indicate that the information on the risk premium under all five consumption

Figure 4: $p$-value of FAR with five consumption measures




Note: The test assets are as in Kroencke (2017), i.e., the 31 portfolios sorted by size, value, and investment, plus the equity premium in 1960-2014. The FAR test is from Kleibergen (2009). The null hypothesis is $\lambda_c$ equal to the value on the horizontal axis. The corresponding $p$-value of the test is plotted (solid), together with the 5% line (dotted).

measures is limited, so that none of them seems superior to the others.

It is also worth emphasizing that the (-$\infty$, $\infty$) confidence sets from the FAR test are not as surprising as they may appear. If the empirically used consumption growth is indeed just uncorrelated noise, $\beta$ is zero, so the risk premium $\lambda_c$ is unidentified in (1), which is reflected by its (-$\infty$, $\infty$) confidence set. Similarly, if $\beta$ is spanned by $\iota_N$, so the full rank condition of $[\iota_N, \beta]$ fails, $\lambda_c$ can not be identified either. Furthermore, since the FAR statistic remains insignificant as indicated by the (-$\infty$, $\infty$) confidence sets, we find little evidence that Equation (1) is misspecified.[6,7]

# 3 Relative risk aversion from GMM

## 3.1 Relative risk aversion: $\gamma$

Other than the cross-sectional tests in the last section, consumption measures are also commonly gauged in a GMM setting. Following the existing literature, we focus on the Euler

[6]As a robustness check, we also use the 25 Fama-French plus 10 industry portfolios in Savov (2011) as alternative test assets. The resulting FAR confidence sets for $\lambda_c$ remain unbounded.

[7]The limited sample size also helps explain the unbounded confidence sets. In particular, our simulation study for Figure 3, which is calibrated to "**P-J**" with $T = 200$, reports that the rejection frequency of $max_{\lambda_c} FAR(\lambda_c)$ (solid) is 100% under $\theta = 1$, which is no longer the case when we simulate with $T = 55$. In the empirical application, $T = 55$.

equation below as the moment condition for $\gamma$, the relative risk aversion:

$$E\left[\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} R^e_{m,t+1}\right] = 0 \tag{6}$$

where the discount factor $\delta$ is fixed at 0.95 for convenience, while $\frac{C_{t+1}}{C_t}$ is proxied by one plus consumption growth and $R^e_{m,t+1}$ is the excess market return. The value of $\gamma$ is of particular interest, since a high $\gamma$ implies implausibly large risk-free interest rates, see, e.g., Savov (2011).

Given the setup above, $\gamma$ is just-identified and its estimator can be derived by minimizing the following GMM objective function:

$$\min_{\gamma}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} R^e_{m,t+1}\right]\mathbb{V}(\gamma)^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} R^e_{m,t+1}\right] \tag{7}$$

where $\mathbb{V}(\gamma)$ is the variance of the GMM moment function.

The rank condition for GMM is that the derivative of the moment function has full rank, i.e., $E\left[\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} R^e_{m,t+1}\triangle c_t\right] \neq 0$. When $\delta$ is a fixed constant and $\frac{C_{t+1}}{C_t} \approx 1$, this rank condition is jeopardized due to the weak correlation between consumption growth and asset returns. Consequently, a similar weak identification problem for $\gamma$ exists, which is identical to the weak identification problem for the risk premium $\lambda_c$ discussed in the previous section.

Table 3: GMM estimates of relative risk aversion $\gamma$

| | **Reported** | **P-J** | **Q4-Q4** | **Garbage** | **Unfiltered** |
|---|---|---|---|---|---|
| Estimate of $\gamma$ | 136.15 | 42.35 | 64.05 | 15.63 | 22.53 |
| s.e. | 52.33 | 22.87 | 39.61 | 8.38 | 11.98 |
| 95% C.I. of $\gamma$, GMM-AR | $[37.51, \infty)$ | $[12.25, \infty)$ | $[16.74, \infty)$ | $[3.88, \infty)$ | $[4.86, \infty)$ |
| $r_f$ implied by lower bound | 63.95 | 66.71 | 33.68 | 10.01 | 13.34 |

Note: The excess market return used for the GMM estimation is in 1960-2014, as in Kroencke (2017). The GMM-AR test is from Stock and Wright (2000). $r_f$ is the risk-free rate.

With the data plotted in Figure 1, Table 3 presents the resulting GMM estimates of $\gamma$ and the associated standard errors, which are identical to those reported in Kroencke (2017).[8] Both Savov (2011) and Kroencke (2017) use the relatively small estimates of $\gamma$ to render credibility to their consumption measures ("**Garbage**" and "**Unfiltered**", respectively).

[8]The only exception is under "**Reported**" in Table 3, we report the GMM estimate, while Kroencke (2017) reports a value of $\gamma$ that minimizes the mean absolute error.

However, when $\gamma$ is weakly identified, the GMM estimates are not trustworthy. See Stock and Wright (2000), Kleibergen (2005).

### 3.2 GMM-AR test

Instead of using the GMM estimates, we test $H_0 : \gamma = \gamma_0$ using a test that does not depend on the GMM rank condition. Note that under the Euler equation and a central limit theorem, it follows, see Stock and Wright (2000),

$$
\begin{aligned}
& \mathit{GMM\text{-}AR}(\gamma_0) \\
= & \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0}R^e_{m,t+1}\right]\mathbb{V}(\gamma_0)^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_0}R^e_{m,t+1}\right] \\
\xrightarrow{d} & \ \chi^2_1.
\end{aligned} \tag{8}
$$

Therefore, a $100\times(1-\alpha)\%$ confidence set for $\gamma$ can be constructed in the following manner: If the test statistic $GMM\text{-}AR(\gamma_0)$ exceeds the upper $\alpha$ quantile of the $\chi^2_1$ distribution, reject $H_0 : \gamma = \gamma_0$ at the $100 \times \alpha\%$ level. The values of $\gamma_0$ that are not rejected constitute the $100 \times (1-\alpha)\%$ confidence set of $\gamma$. This corresponds to the so-called GMM-AR test in Stock and Wright (2000).

Table 3 shows that the 95% confidence sets for $\gamma$ from the GMM-AR test do not have an upper bound. Therefore, little information about $\gamma$ is available. These robust confidence sets result from the $p$-value plots for testing $\gamma$ equal to the value on the horizontal axis in Figure 5. A $p$-value larger than 0.05 implies that the null is not rejected at the 5% level. The largest $p$-value in Figure 5 corresponds to the GMM estimate of $\gamma$ reported in Table 3.

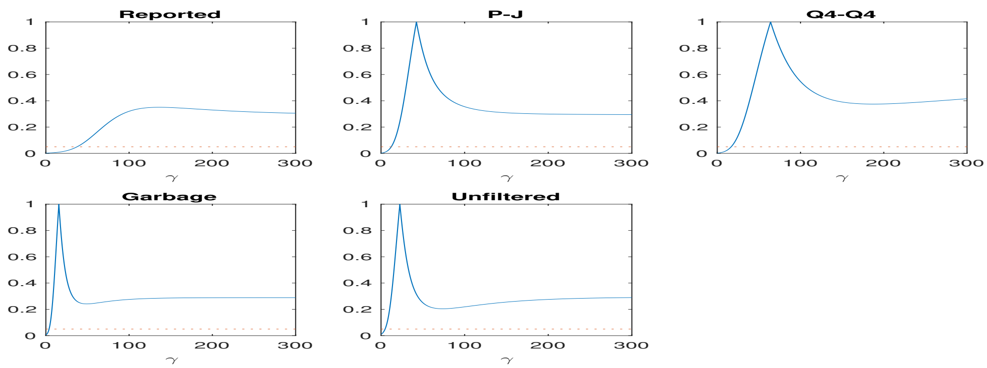
The risk-free rate $r_f$ is related to the relative risk aversion $\gamma$ by (see, e.g., Savov 2011):

$$r_f = -\log(\delta) + \gamma E[\log(C_{t+1}/C_t)] - \frac{\gamma^2}{2} Var[\log(C_{t+1}/C_t)]. \tag{9}$$

For a value of $\gamma$ within its robust confidence set constructed using the GMM-AR test, we can therefore compute the corresponding risk-free rate $r_f$. The resulting $r_f$ implied by the lower bound of $\gamma$ is reported using the GMM-AR confidence sets for $\gamma$ in Table 3.

Table 3 shows that the NIPA consumption measure ("**Reported**"), the three-year consumption measure in Parker and Julliard (2005) ("**P-J**") and the fourth-quarter to fourth-quarter consumption measure in Jagannathan and Wang (2007) ("**Q4-Q4**") imply risk-free

Figure 5: $p$-value of GMM-AR with five consumption measures




Note: The excess market return used for the GMM-AR test is in 1960-2014, as in Kroencke (2017). The GMM-AR test is from Stock and Wright (2000). The null hypothesis is $\gamma$ equal to the value on the horizontal axis. The corresponding $p$-value of the test is plotted (solid), together with the 5% line (dotted).

rates that are implausibly large, when the smallest possible $\gamma$ is adopted. In contrast, the garbage measure in Savov (2011) and the unfiltered NIPA consumption in Kroencke (2017) show support for risk-free rates that are substantially smaller. From this perspective, the consumption measures in Savov (2011) and Kroencke (2017) appear to outperform the other measures, in terms of helping to address the risk-free rate puzzle. Nevertheless, since all confidence sets for $\gamma$ from the GMM-AR test are unbounded, we find no strong evidence that one consumption measure is superior to the others.[9]

## 3.3 Joint confidence sets for $\gamma$ and $\delta$

Another commonly used moment restriction on the relative risk aversion is:

$$E\left[\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} R_{f,t+1} - 1\right] = 0 \tag{10}$$

where $R_{f,t+1}$ is the gross risk-free rate. See, e.g., Savov (2011), Kroencke (2017).

Combining Equation (6) and Equation (10) yields two restrictions on $\gamma$ and $\delta$. We therefore do not have to fix $\delta$ at 0.95. Instead, we consider $(\gamma, \delta)$ as a pair of parameters.

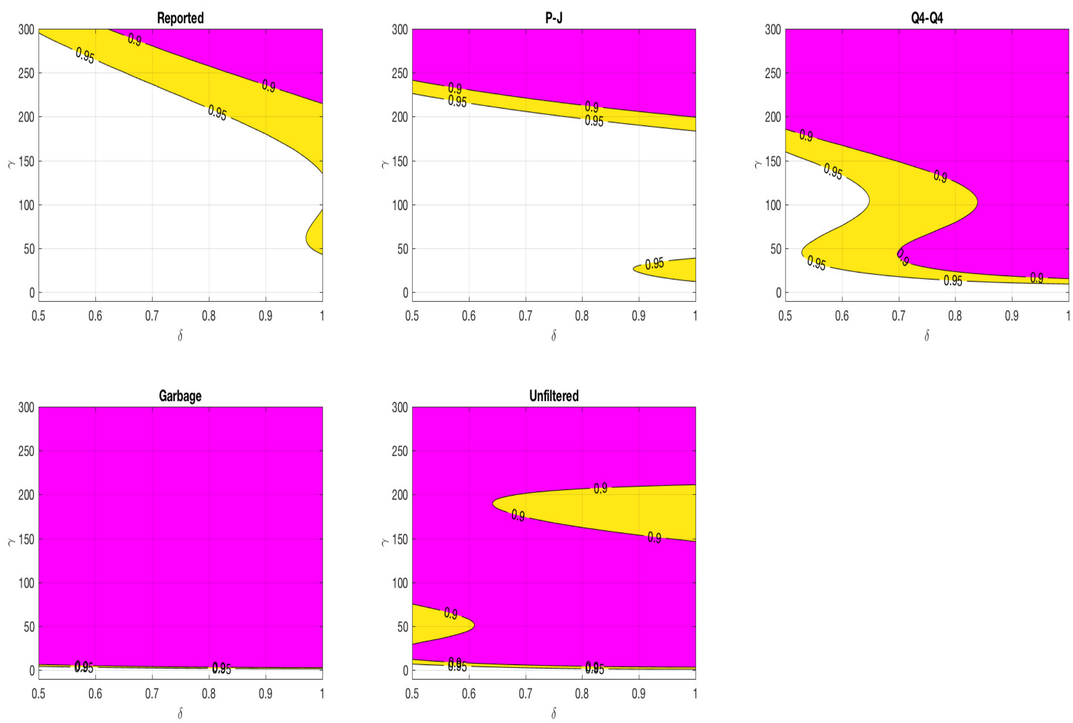
[9]As a robustness check, we also restrict the sample to 1960-2006, the sample period considered in Savov (2011). The resulting GMM-AR confidence sets for $\gamma$ are also unbounded.

The corresponding GMM-AR statistic for testing $H_0 : (\gamma, \delta) = (\gamma_0, \delta_0)$ reads:

$$
\begin{aligned}
& GMM\text{-}AR(\gamma_0, \delta_0) \\
= & \left\{ \begin{array}{c} \frac{1}{\sqrt{T}} \sum_t \delta_0 \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma_0} R^e_{m,t+1} \\ \frac{1}{\sqrt{T}} \sum_t \left[ \delta_0 \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma_0} R_{f,t+1} - 1 \right] \end{array} \right\}' \tilde{\mathbb{V}}(\gamma_0, \delta_0)^{-1} \left\{ \begin{array}{c} \frac{1}{\sqrt{T}} \sum_t \delta_0 \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma_0} R^e_{m,t+1} \\ \frac{1}{\sqrt{T}} \sum_t \left[ \delta_0 \left( \frac{C_{t+1}}{C_t} \right)^{-\gamma_0} R_{f,t+1} - 1 \right] \end{array} \right\} \\
\xrightarrow{d} & \ \chi^2_2
\end{aligned}
\tag{11}
$$

where $\tilde{\mathbb{V}}(\gamma_0, \delta_0)$ is the $2 \times 2$ covariance matrix of the two moment functions.

Figure 6: Joint confidence sets of $\gamma$ and $\delta$




Note: The joint confidence sets result from the GMM-AR test in Stock and Wright (2000). The excess market return and the gross risk-free rate used for the test are in 1960-2014.

Using this GMM-AR test with the five consumption measures, the resulting joint confidence sets for $\gamma$ and $\delta$ are presented in Figure 6. These confidence sets are made of the $(\gamma_0, \delta_0)$'s such that the $GMM\text{-}AR(\gamma_0, \delta_0)$ statistic does not exceed the critical values of the $\chi^2_2$ distribution. We focus on $\gamma$ in [0, 300] and $\delta$ in [0.5, 1] in Figure 6, since these ranges are mostly practically relevant. Figure 6 shows that all five consumption measures induce large $\gamma$'s in the shaded 90% and 95% confidence sets, so none of these measures is informative enough for the relative risk aversion.

Nevertheless, Figure 6 also illustrates that the garbage measure in Savov (2011) and the unfiltered NIPA consumption in Kroencke (2017) show support for small $\gamma$'s that are not favored by the consumption measures of NIPA, Parker and Julliard (2005), Jagannathan and Wang (2007). These findings are in line with the confidence sets from the GMM-AR test presented in Table 3, all of which imply that the consumption measures proposed by Savov (2011) and Kroencke (2017) minorly help to address the risk-free rate puzzle.

# 4 Conclusions

Our conclusions are two-fold. On the dark side, we find that the pricing information in the five different consumption measures appears to be similarly limited, when they are examined using the identification robust FAR and GMM-AR tests. On the bright side, however, we help bridge the gap of the seemingly conflicting empirical findings due to the adoption of different consumption measures.

The key insight is that the correlation between consumption growth and asset returns is not sufficiently strong under all our five consumption measures to meet the rank conditions in the FM two-pass procedure and the conventional GMM framework. Therefore, conventional estimates and $t$-statistics for the risk premium and relative risk aversion are not trustworthy and exhibit severe discrepancy from the confidence sets constructed by inverting robust tests.

# Appendix

## A. Factor Anderson-Rubin (FAR) statistic

Under the moment equation in (1) and $H_0 : \lambda_c = \lambda_{c,0}$, the FAR test statistic stated directly below in (12) converges to a $\chi^2_{N-1}$ distributed random variable, assuming homoscedasticity and other mild conditions that do not depend on the value of $\beta$:[10]

$$FAR(\lambda_{c,0}) = \frac{T}{1-\lambda_{c,0}^2/\hat{Q}(\lambda_{c,0})}(\bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{c,0})'\tilde{\Sigma}^{-1}(\bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{c,0}) \xrightarrow{d} \chi^2_{N-1} \tag{12}$$

where $T$ is the sample size, $\hat{Q}(\lambda_{c,0}) = \frac{1}{T}\sum_{t=1}^T(\overline{\triangle c}_t + \lambda_{c,0})^2$, $\overline{\triangle c}_t = \triangle c_t - \frac{1}{T}\sum_{t=1}^T \triangle c_t$, $\bar{\mathcal{R}} = \frac{1}{T}\sum_{t=1}^T \mathcal{R}_t$, $\mathcal{R}_t = R_{1t} - \iota_{N-1}R_{Nt}$, $R_t = (R'_{1t}, R'_{Nt})'$ is the $N \times 1$ vector of test assets with $R_{1t} : (N-1)\times 1$, $R_{Nt} : 1 \times 1$, $\tilde{\mathcal{B}} = \sum_{t=1}^T \mathcal{R}_t(\overline{\triangle c}_t + \lambda_{c,0})\left[\sum_{j=1}^T(\overline{\triangle c}_j + \lambda_{c,0})^2\right]^{-1}$, $\tilde{\Sigma} = \frac{1}{T-1}\sum_{t=1}^T(\mathcal{R}_t - \tilde{\mathcal{B}}(\overline{\triangle c}_t + \lambda_{c,0}))(\mathcal{R}_t - \tilde{\mathcal{B}}(\overline{\triangle c}_t + \lambda_{c,0}))'$. Since $\bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{c,0} = \bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_c + \tilde{\mathcal{B}}(\lambda_c - \lambda_{c,0})$, the FAR statistic tests both the correct specification of the moment equation in (1) and the hypothesized value $\lambda_{c,0}$, see Kleibergen (2009).

## B. Distribution of the FAR statistic under misspecification

Kleibergen (2009) establishes that the asymptotic distribution of the FAR statistic, $FAR(\lambda_{c,0})$, is $\chi^2_{N-1}$ under correct specification, regardless of $\beta$. In this appendix, we show that the asymptotic distribution of $FAR(\lambda_{c,0})$ for the D.G.P. used in our simulation experiment, which implies misspecification, is also $\chi^2_{N-1}$, when $\beta$ approaches zero. The tested value $\lambda_{c,0}$ reflects the pseudo risk premium which results from

$$\begin{bmatrix} \lambda_0 \\ \lambda_c \end{bmatrix} = (X'X)^{-1}X'\mu_R, \quad with\ X = [\iota_N, \beta],\ \beta = V_{Rc}V_c^{-1}\theta, \quad so \quad \begin{bmatrix} \lambda_0 \\ \lambda_c \end{bmatrix} =$$

$$\begin{pmatrix} N & \iota'_N V_{Rc}V_c^{-1}\theta \\ \iota'_N V_{Rc}V_c^{-1}\theta & V_c^{-2}V'_{Rc}V_{Rc}\theta^2 \end{pmatrix}^{-1} \begin{pmatrix} \iota'_N\mu_R \\ \theta V_c^{-1}V'_{Rc}\mu_R \end{pmatrix} = \begin{pmatrix} \frac{V_c^{-2}V'_{Rc}V_{Rc}\iota'_N\mu_R - \iota'_N V_{Rc}V_c^{-1}V_c^{-1}V'_{Rc}\mu_R}{NV_c^{-2}V'_{Rc}V_{Rc} - (\iota'_N V_{Rc}V_c^{-1})^2} \\ \theta^{-1}\frac{NV_c^{-1}V'_{Rc}\mu_R - \iota'_N V_{Rc}V_c^{-1}\iota'_N\mu_R}{NV_c^{-2}V'_{Rc}V_{Rc} - (\iota'_N V_{Rc}V_c^{-1})^2} \end{pmatrix}.$$

Hence, when $\theta$ goes to zero, $\beta$ goes to zero as well and the tested value $\lambda_{c,0}$ goes to infinity.

The moment condition for the risk premium can be re-written as:

$$\mu_R = \iota_N\lambda_0 + \beta\lambda_c + \mu_e$$

where $\mu_e$ is the specification error, which is non-zero iff misspecification exists. The accompanying linear factor model reads:

$$R_t = \iota_N\lambda_0 + \beta(\triangle c_t + \lambda_c) + \mu_e - \beta\mu_c + \epsilon_t$$

[10] The FAR statistic is identical to the Hotelling type statistic proposed in Beaulieu *et al*. (2013) which for the same null hypothesis reads: $H(\lambda_{c,0}) = T\left[\binom{1}{\lambda_{c,0}}'\left(\frac{1}{T}\sum_{t=1}^T\binom{1}{\Delta\bar{c}_t}\binom{1}{\Delta\bar{c}_t}'\right)^{-1}\binom{1}{\lambda_{c,0}}\right]^{-1}(\bar{\mathcal{R}} - \hat{\mathcal{B}}\lambda_{c,0})'\tilde{\Sigma}^{-1}(\bar{\mathcal{R}} - \hat{\mathcal{B}}\lambda_{c,0})$ with $\hat{\mathcal{B}} = \sum_{t=1}^T \mathcal{R}_t\overline{\triangle c}_t(\sum_{t=1}^T \overline{\triangle c}_t^2)^{-1}$.

where $\mu_c$ is the mean of $\triangle c_t$, $\epsilon_t$ is the mean zero error term. Define the $(N-1)\times N$ constant matrix $M = [I_{N-1} \vdots -\iota_{N-1}]$. Removing the $N$-th return in $R_t$ yields:

$$\mathcal{R}_t = \mathcal{B}(\triangle c_t + \lambda_c) + \tilde{\mu}_e - \mathcal{B}\mu_c + \tilde{\epsilon}_t$$

where $\mathcal{R}_t = MR_t$, $\mathcal{B} = M\beta$, $\tilde{\mu}_e = M\mu_e$, $\tilde{\epsilon}_t = M\epsilon_t$.

Recall the FAR statistic in (12). We rewrite $\hat{Q}(\lambda_{c,0}) = \frac{1}{T}\sum_{t=1}^{T}(\overline{\triangle c}_t + \lambda_{c,0})^2$ as $\hat{Q}(\lambda_{c,0}) = \lambda_{c,0}^2 + S_c^2$ with $S_c^2 = \frac{1}{T}\sum_{t=1}^{T}\overline{\triangle c}_t^2$. Therefore,

$$\hat{Q}(\lambda_{c,0}) - \lambda_{c,0}^2 = S_c^2 \xrightarrow{p} V_c, \ \textit{where}\ V_c = Var(\triangle c_t). \tag{13}$$

We specify $\hat{Q}(\lambda_{c,0})(\bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{c,0})$ as:

$$\begin{aligned}
\hat{Q}(\lambda_{c,0})(\bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{c,0}) &= \hat{Q}(\lambda_{c,0})\bar{\mathcal{R}} - \tfrac{1}{T}\textstyle\sum_{t=1}^{T}\mathcal{R}_t(\overline{\triangle c}_t + \lambda_{c,0})\lambda_{c,0} \\
&= S_c^2\bar{\mathcal{R}} - \tfrac{1}{T}\textstyle\sum_{t=1}^{T}\mathcal{R}_t\overline{\triangle c}_t\lambda_{c,0} \\
&= S_c^2(\bar{\mathcal{R}} - \hat{\mathcal{B}}\lambda_{c,0}),\ \textit{where}\ \hat{\mathcal{B}} = \textstyle\sum_{t=1}^{T}\mathcal{R}_t\overline{\triangle c}_t(\sum_{t=1}^{T}\overline{\triangle c}_t^2)^{-1}.
\end{aligned}$$

Note that $\bar{\mathcal{R}} \overset{a}{=} \mu_{\mathcal{R}} + T^{-1/2}\psi_{\mathcal{R}}$, where $\psi_R$ is normally distributed with mean 0 and variance $V_{\mathcal{R}}$. Similarly, note that $\hat{\mathcal{B}} \overset{a}{=} \mathcal{B} + T^{-1/2}\psi_{\mathcal{B}}$, where $\psi_{\mathcal{B}}$ is normally distributed with mean 0 and variance $V_c^{-1}V_{\tilde{\epsilon}}$, with $V_c$ the variance of $\triangle c_t$, $V_{\tilde{\epsilon}}$ is the variance of $\tilde{\epsilon}_t$. Therefore,

$$\begin{aligned}
\hat{Q}(\lambda_{c,0})(\bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{c,0}) &= S_c^2(\bar{\mathcal{R}} - \hat{\mathcal{B}}\lambda_{c,0}) \\
&\overset{a}{=} S_c^2(\mu_{\mathcal{R}} + T^{-1/2}\psi_{\mathcal{R}} - \mathcal{B}\lambda_{c,0} - T^{-1/2}\psi_{\mathcal{B}}\lambda_{c,0}) \\
&= S_c^2(\tilde{\mu}_e + T^{-1/2}\psi_{\mathcal{R}} - T^{-1/2}\psi_{\mathcal{B}}\lambda_{c,0})
\end{aligned}$$

When $\theta$ goes to zero, $\beta$ approaches zero as well and the hypothesized $\lambda_{c,0}$ goes to infinity, so the above three terms reduce to one:

$$\hat{Q}(\lambda_{c,0})(\bar{\mathcal{R}} - \tilde{\mathcal{B}}\lambda_{c,0}) \overset{a}{=} -V_cT^{-1/2}\psi_{\mathcal{B}}\lambda_{c,0}, \tag{14}$$

since $\lambda_{c,0}$ is of a much larger order of magnitude than the other components.

To complete the proof, we now only need to show

$$\lambda_{c,0}^{-2}\hat{Q}(\lambda_{c,0})\tilde{\Sigma} \xrightarrow{p} V_{\tilde{\epsilon}} \tag{15}$$

when $\lambda_{c,0}$ goes to infinity, which results from:

$$\begin{aligned}
\hat{Q}(\lambda_{c,0})\tilde{\Sigma} = &\ \tfrac{1}{T-1}\hat{Q}(\lambda_{c,0})\textstyle\sum_{t=1}^{T}\mathcal{R}_t\mathcal{R}_t' - \tfrac{1}{T-1}\sum_{t=1}^{T}\mathcal{R}_t(\overline{\triangle c}_t + \lambda_{c,0})\sum_{j=1}^{T}(\overline{\triangle c}_j + \lambda_{c,0})\mathcal{R}_j' \\
= &\ \lambda_{c,0}^2\left[\tfrac{1}{T-1}\textstyle\sum_{t=1}^{T}\left(\mathcal{R}_t - \bar{\mathcal{R}}\right)\left(\mathcal{R}_t - \bar{\mathcal{R}}\right)'\right] - \lambda_{c,0}\left[\bar{\mathcal{R}}\sum_{t=1}^{T}\overline{\triangle c}_t\mathcal{R}_t' + \left(\sum_{t=1}^{T}\mathcal{R}_t\overline{\triangle c}_t\right)\bar{\mathcal{R}}'\right] \\
&+ \tfrac{1}{T-1}S_c^2\textstyle\sum_{t=1}^{T}\mathcal{R}_t\mathcal{R}_t' - \tfrac{1}{T-1}\sum_{t=1}^{T}\mathcal{R}_t\overline{\triangle c}_t\sum_{j=1}^{T}\overline{\triangle c}_j\mathcal{R}_j'
\end{aligned}$$

since $\frac{1}{T-1}\sum_{t=1}^{T}\left(\mathcal{R}_t - \bar{\mathcal{R}}\right)\left(\mathcal{R}_t - \bar{\mathcal{R}}\right)' \xrightarrow{p} V_{\tilde{\epsilon}}$ when $\beta$ goes to zero.

Combining (13), (14), (15) implies $FAR(\lambda_{c,0}) \xrightarrow{d} \chi_{N-1}^2$, when $\theta$ goes to zero in the D.G.P. so $\beta$ also goes to zero and the hypothesized value $\lambda_{c,0}$ goes to infinity.

# References


[1] Marie-Claude Beaulieu, Jean-Marie Dufour, and Lynda Khalaf. Identification-robust estimation and testing of the zero-beta CAPM. *Review of Economic Studies*, 80(3):892–924, 2013.

[2] Douglas T Breeden, Michael R Gibbons, and Robert H Litzenberger. Empirical tests of the consumption-oriented CAPM. *Journal of Finance*, 44(2):231–262, 1989.

[3] Craig Burnside. Identification and inference in linear stochastic discount factor models with excess returns. *Journal of Financial Econometrics*, 14(2):295–330, 2016.

[4] E.F. Fama and J.D. MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, 81(3):607–636, 1973.

[5] Nikolay Gospodinov, Raymond Kan, and Cesare Robotti. Spurious inference in unidentified asset-pricing models. *Econometrica*, Forthcoming, 2017.

[6] Patrik Guggenberger, Frank Kleibergen, Sophocles Mavroeidis, and Linchun Chen. On the asymptotic sizes of subset anderson–rubin and lagrange multiplier tests in linear instrumental variables regression. *Econometrica*, 80(6):2649–2666, 2012.

[7] L.P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica*, pages 1029–1054, 1982.

[8] Ravi Jagannathan and Yong Wang. Lazy investors, discretionary consumption, and the cross-section of stock returns. *Journal of Finance*, 62(4):1623–1661, 2007.

[9] R. Kan and C. Zhang. Two-pass tests of asset pricing models with useless factors. *Journal of Finance*, pages 203–235, 1999.

[10] F. Kleibergen. Testing parameters in GMM without assuming that they are identified. *Econometrica*, pages 1103–1123, 2005.

[11] F. Kleibergen. Tests of risk premia in linear factor models. *Journal of Econometrics*, 149(2):149–173, 2009.

[12] F. Kleibergen and R. Paap. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133(1):97–126, 2006.

[13] F. Kleibergen and Z. Zhan. Unexplained factors and their effects on second pass R-squared's. *Journal of Econometrics*, 189:101–116, 2015.

[14] Tim A Kroencke. Asset pricing without garbage. *Journal of Finance*, 72(1):47–98, 2017.

[15] Jonathan A Parker and Christian Julliard. Consumption risk and the cross section of expected returns. *Journal of Political Economy*, 113(1):185–222, 2005.

[16] Alexi Savov. Asset pricing with garbage. *Journal of Finance*, 66(1):177–201, 2011.

[17] J. Shanken. On the estimation of beta-pricing models. *Review of Financial Studies*, pages 1–33, 1992.

[18] J.H. Stock and J.H. Wright. GMM with weak identification. *Econometrica*, pages 1055–1096, 2000.