Learning about Beta: Time-Varying Factor Loadings, Expected Returns, and the Conditional CAPM

Tobias Adrian and Francesco Franzoni *

February 3, 2009

ABSTRACT

We amend the conditional CAPM to allow for unobservable long-run changes in risk factor loadings. In this environment, investors rationally "learn" the long-run level of factor loadings from the observation of realized returns. As a consequence of this assumption, we model conditional betas using the Kalman filter. Because of its focus on low-frequency variation in betas, our approach circumvents recent criticisms of the conditional CAPM. When tested on portfolios sorted by size and book-to-market, our learning-augmented conditional CAPM passes the specification tests.

*Federal Reserve Bank of New York and Swiss Finance Institute at the University of Lugano, respectively.

Fama and French (1992) present compelling evidence that the unconditional captial asset pricing model (CAPM) does not account for returns on portfolios sorted by size and book-to-market (B/M). Since then, the literature on asset pricing has developed alternative theories, which depart from the original model along several dimensions. Promising avenues of research, which preserve the single-factor structure, have been conditional versions of the CAPM. The idea behind this approach is that, even though CAPM holds conditionally on time t information, it does not hold unconditionally. Accordingly, the poor empirical performance of CAPM might be due to its failure to account for time variation in conditional moments.

Among the many studies of the conditional CAPM, recent implementations are put forward by Jagannathan and Wang (1996), Ferson and Harvey (1999), and Lettau and Ludvigson (2001). Earlier papers include Ferson, Kandel, and Stambaugh (1987), Bollerslev, Engle, and Wooldridge (1988), and Harvey (1989). In all cases, the authors explicitly model the evolution of the conditional distribution of returns as a function of lagged state variables. They specify the covariance between the market return and portfolio returns as affine functions of these variables. This specification is estimated as a multifactor model, in which the additional factors are the interactions between the market return and the state variables.

A recent paper by Lewellen and Nagel (2006) casts doubts on the empirical success of this approach. While acknowledging that betas vary considerably over time, these authors suggest that the covariation between sample estimates of betas and the market risk premium is not large enough to justify the deviations from the unconditional CAPM observed for value and momentum portfolios (Fama and French (1993); Jegadeesh and Titman (1993)). They argue that the good empirical performance of previous conditional studies is due to their cross-sectional design—which ignores key theoretical restrictions on the estimated slope coefficients—and suggest time-series regressions instead.

In this paper, we complement the conditional CAPM literature by modeling a new type of time variation in conditional betas. There is substantial evidence that the risk of some asset classes has experienced long-run movements. For example, CAPM regressions on CRSP

data suggest that value stocks had higher betas than growth stocks between 1926 and 1963. The opposite is true in the later sample.[1] This low frequency evolution is different from the covariance of beta with the equity premium that is typically modeled in conditional CAPM applications.

Our goal is to explore the implications of long-run changes in factor loadings for the tests of conditional models. We expect this effect to be important for the assets that have displayed major swings in their betas, such as portfolios sorted by size and B/M.

We tackle this issue by assuming that betas change over time following a mean-reverting process. Also important, we postulate that investors are unaware of the long-run level of a given asset's risk. Hence, they need to infer both the level of beta and its long-run mean from the history of realized returns. Through a stylized model, in which the conditional CAPM holds, we make the point that the risk loading that determines expected returns is the expectation of beta that results from the learning process. We model this expectation via a Kalman filter, in which the factor loading is treated as a latent variable, and argue that betas should be estimated accordingly in empirical applications. Consistent with this theoretical argument, we use Kalman-filtered betas to explain the returns of the twenty-five portfolios sorted by size and B/M.

Our first set of results concerns the CAPM, augmented with learning, but without conditioning variables. When we use a mispricing measure borrowed from Campbell and Vuolteenaho (2004), the sole contribution of learning is a reduction of about 45% in mispricing relative to the unconditional CAPM. Low-frequency movements of beta play a crucial role in this result. Investors' inferences about the long-run level of beta can cause a significant difference between the ex-ante expected level of risk and ex-post estimates from typical OLS regressions. This mechanism is particularly relevant for portfolios of value and small stocks that have experienced considerable long-run variation in beta. We argue that the wedge between investors' ex-ante expectations of beta and ex-post OLS estimates can account for a large fraction of the unconditional alpha in standard OLS time-series regres-

[1] See, for example, Franzoni (2002 and 2008), Campbell and Vuolteenaho (2004), Fama and French (2006), and Ang and Chen (2007).

sions. In other words, the mismeasurement of expectations of beta and, hence, of equilibrium expected returns can be the source of the apparent mispricing.

Our second set of results concerns the conditional CAPM with different sets of scaling variables. We confirm Lewellen and Nagel's (2006) finding that conditioning variables alone do not improve the performance of CAPM much in time-series tests. However, once we introduce learning in conditional CAPMs, the model is no longer rejected, and the aggregate pricing error is reduced by 45% (it decreases by 65% when compared to the unconditional CAPM). In every specification, we find that the model without learning is rejected, whereas the model with learning is not. Pricing errors of our (one-factor) learning-augmented conditional CAPM are comparable in magnitude to the ones from the Fama-French three-factor model.

The intuition behind the empirical success of our learning-augmented CAPM in pricing B/M and size-sorted portfolios is as follows. Consider the value premium. The failure of CAPM to price value stocks in an OLS framework is due to value portfolios having high average returns but low estimated betas. In our approach based on the Kalman-filter, given the decrease in systematic risk of value stocks, the high level of the past factor loading affects today's estimates and makes them larger than OLS betas. A high estimate of beta is thus matched with high average returns, and the estimated alpha of value stocks is reduced. A similar intuition applies to small stocks, which have also experienced a decline in systematic risk. The situation is reversed in the case of large and growth stocks, whose levels of risk have increased over time.

Our results provide a support for the conditional CAPM that is not subject to Lewellen and Nagel's (2006) critique. The type of variability in beta that affects our estimates is low-frequency, idiosyncratic variation, experienced over a long time horizon. Lewellen and Nagel, instead, focus on the cyclical covariation between beta and market risk premium. Consistent with their prediction, we find that the standard conditioning variables do not lead to success of CAPM when the model is tested in the time series via OLS. However, the inclusion of these conditioning variables in our learning framework reestablishes the success

of the conditional CAPM in pricing size- and B/M-sorted portfolios.

Following much of the literature on asset pricing since the early 1970s, our test assets are portfolios of stocks. In the context of conditional asset pricing models, this procedure implicitly assumes that investors price stocks according to the risk of the asset class to which they belong. The asset class is defined by the relevant characteristic according to which portfolios are sorted. In our empirical application, we use the Kalman filter to estimate betas of size- and B/M-sorted portfolios rather than individual stocks. In this sense, we proceed as if investors estimated a unique beta for all the stocks in the same portfolio. While this assumption has the flavor of some "behavioral" specification, such as "style investing" (Barberis and Shleifer (2003)), it can also be motivated on rational grounds and anecdotal evidence. First, it makes sense to believe *a priori* that when the time series of a firm's returns is not long enough to allow a reliable estimation of its beta, the rational way to predict its risk is to compare the company to other firms with similar characteristics, for which a longer time series is available. Second, we believe the spirit of this assumption is in line with an important feature of the actual investment process. For example, Barra, a provider of beta estimates, reports a "fundamental measure" of a stock's beta. Its estimate is the weighted average of the betas of a set of characteristic-based portfolios to which a stock belongs.[2] Our empirical focus on the size and B/M characteristics is a clear simplification of the multidimensionality of the problem. Still, we feel that we replicate a salient feature of institutional investors' approach in computing betas.

The paper is organized as follows. Section 1 formalizes our theoretical argument through a stylized model, which we label Learning-CAPM. Section 2 implements the Kalman-filter methodology to estimating betas and discusses the assumptions behind our approach. Section 3 contains the asset pricing tests of different specifications of the Learning-CAPM and compares the performance of this model to other CAPM specifications. Section 4 contains a review of additional related literature. Section 5 draws the conclusions of this work.

[2]Barra says that this fundamental beta is superior to the historical beta in predicting future risk. See, for example, Barra's newsletter from September 1988: "What's New About Beta?"

1. Conditional CAPM with Learning

Since there is substantial evidence that the risk of some asset classes varies over time, investors may not know the precise riskiness of assets when they make their portfolio choices. In a world with uncertainty about relevant parameters, rational agents have to infer the factor loadings from the available information.

We denote excess returns for assets $i = 1...N$ by R^i_{t+1}. The previous discussion explains that we focus on assets that differ by some characteristic that captures an investment style. For example, the heterogeneity in these assets could originate from their size and B/M characteristics. While a clear simplification of the actual dimensionality of the problem, this interpretation is not distant from the top-down approach that characterizes many institutional investors.

In Appendix 1, we derive the following conditional CAPM from no-arbitrage and a simple factor structure for returns:

$$E_t\left[R^i_{t+1}\right] = \beta^{ie}_{t+1|t} E_t\left[R^M_{t+1}\right], \tag{1}$$

where $\beta^{ie}_{t+1|t} = E_t\left[\beta^i_{t+1}\right] = cov_t\left(R^i_{t+1}, R^M_{t+1}\right)/var_t\left(R^M_{t+1}\right)$ is both the time t expectation of the risk factor loading that applies to time $t+1$ returns and the conditional beta. Also, from Equation (A-4) in the appendix, β^i_{t+1} is loading on the market factor for time $t+1$ on asset i. We call this conditional CAPM the Learning-CAPM, as the expected return is proportional to the expected risk factor loading $\beta^{ie}_{t+1|t}$. The expected risk factor loading is a conditional expectation, and its evolution depends on the stochastic specification of unobserved beta's evolution. We assume that β^i_{t+1} evolves according to an autoregressive process conditional on a vector of stationary exogenous state variables y_t:

$$\beta^i_{t+1} = \left(1 - F^i\right) B^i + F^i \beta^i_t + \phi^{i\prime} y_t + u^i_{t+1}. \tag{2}$$

Without loss of generality, we assume that the average of the conditioning variables y_t is zero

over time, so that we can interpret B^i as the long-run mean of the factor loading β^i_{t+1}. We assume that this long-run mean is unobserved, so that investors have to form expectations about the current level of riskiness of asset i, β^i_{t+1}, as well as the long-run level of risk, B^i.[3] Finally, u^i_{t+1} is an independent normal shock. As a consequence, the innovations to the factor loadings are uncorrelated with expected risk premia. It is worth stressing that we are not assuming that the evolution of beta is independent of the equity premium, which would be counterfactual. In fact, the correlation between β^i_{t+1} and $E_t\left[R^M_{t+1}\right]$ is achieved through the presence of the conditioning variables y_t in Equation (2).

In Appendix 1, we also derive an expression for time $t+1$ excess returns:

$$R^i_{t+1} = \beta^i_{t+1} R^M_{t+1} + \eta_{t+1}, \tag{3}$$

where $\eta_{t+1} = \left(\beta^{ie}_{t+1|t} - \beta^i_{t+1}\right) E_t\left[R^M_{t+1}\right] + \varepsilon^i_{t+1}$. Also, ε^i_{t+1} is an idiosyncratic shock to R^i_{t+1}, the return on asset i, and is defined in Equation (A-2) of the appendix. Given the assumption of idiosyncratic innovations to β^i_{t+1}, one can show that η_{t+1} is orthogonal to R^M_{t+1} (the proof is in Appendix 3). This orthogonality is a necessary condition for applying the Kalman filter to estimate β^i_{t+1} in Equation (3).

Rationality implies that changes in the expectation of the factor loading are determined by Bayes' rule. We assume that the shocks u^i_{t+1} and η^i_{t+1} are conditionally normal. Hence, the conditional expectation of β^i_{t+1} evolves according to the Kalman filter. In Appendix 2, we show that the dynamics of investors' expectations follow:

$$\beta^{ie}_{t+1|t} = \left(1 - F^i\right) B^{ie}_{t-1} + F^i \beta^{ie}_{t|t-1} + \phi^{i\prime} y_t + k^i_t \left(R^i_t - E_{t-1}\left[R^i_t\right]\right), \tag{4}$$

where B^{ie}_{t-1} and $\beta^{ie}_{t|t-1}$ are time $t-1$ forecasts of B^i and β^i_t, respectively, and they are made using the Kalman filter. The optimal rule is to use the unexpected part of the current return realization to update the previous period's estimate of the factor loading k^i_t is the "gain"

[3] One can let the long run mean of the factor loading B vary over time. Our results still go through in this more general case. Hence, to keep the model parsimonious, we focus on the case with constant B.

and can be interpreted as a time-varying regression coefficient. Equation (4) states that the one-period-ahead forecast of the factor loading is a combination of the long-run behavior of β_t^i, as captured by the expectation of B^i, and the current estimate of the level of risk.

The updating equation for expectations about B^i is

$$B_t^{ie} = B_{t-1}^{ie} + K_t^i \left(R_t^i - E_{t-1} \left[R_t^i \right] \right). \tag{5}$$

Hence, although the long-run mean of beta is constant, its expectation changes over time.[4] In particular, Equation (5) implies that B_t^{ie} is a martingale under investors' information set. This feature introduces a slow-moving component in risk factor loadings and required risk premia. Ultimately, our results depend on the persistence of shocks to the expectation of beta's long-run mean. Because of this persistence, levels of risk from a distant past obtain a non-negligible weight in today's beliefs about beta. Hence, in our story, the high risk premia required from value and small stocks are related to their high perceived risk.[5]

The specification presented in Equation (2) nests popular conditional CAPM models as special cases. Harvey (1989), Schwert and Seguin (1990), Jagannathan and Wang (1996), Lettau and Ludvigson (2001), and Ferson and Harvey (1999) all assume that betas are deterministic, affine functions of state variables. This is the special case of Equation (2) when $u_{t+1}^i = 0$, $F^i = 0$, and B^i is observable. Our specification in Equation (2) also nests the models of time-varying betas proposed by Chan and Chen (1991), Jostova and Philipov (2005), and Ang and Chen (2007), all of whom assume autoregressive processes for β_{t+1}^i, but do not have conditioning variables ($\phi^{i\prime} = 0$), and—most importantly—assume that the long-run mean of beta (B^i) is an observable parameter.

One can see the impact of learning on observed mispricing by computing population

[4]The gain matrix for the filter of B^i, denoted K_t^i, can also be interpreted as a time-varying regression coefficient and its expression is given in Appendix 2.

[5]We did experiment with a specification, in which B^i is an observable parameter rather than a latent variable. In this case, B^i is estimated using maximum likelihood in the same way as the other parameters of the Kalman filter. This alternative specification, which is available upon request, produces substantially larger pricing errors than the ones we present in the paper. Hence, our results really hinge on the assumption that investors are unaware of an asset's underlying long-run level of risk.

counterpart to the OLS estimate of alpha, as in Lewellen and Nagel (2006). Denote the unconditional OLS beta by β^i_{OLS} and the unconditional OLS alpha estimated on the same sample by α^i_{OLS}. After simple steps that are described in Appendix 4, one obtains an expression for the population value of the OLS alpha:

$$\alpha^i_{OLS} = \underbrace{E\left[\beta^i_{t+1} - \beta^i_{OLS}\right] E\left[R^M_{t+1}\right]}_{\text{Learning component}} + \underbrace{Cov\left[\beta^i_{t+1}, R^M_{t+1}\right]}_{\text{Covariance component}}. \tag{6}$$

2. Estimation of the Learning-CAPM

The size- and B/M-related anomalies pointed out by Fama and French (1992 and 1993) represent the type of mispricing most likely related to learning, as the betas of these portfolios have displayed strong time variation (Franzoni, 2002 and 2009). This consideration motivates the adoption of size- and B/M-sorted portfolios in our empirical application. Our methodology is not subject to Lewellen, Nagel, and Shanken's (2007) critique of this choice of test assets because we avoid cross-sectional tests.

We use the twenty-five portfolios that result from double-sorting the stocks of NYSE, Amex, and Nasdaq along the size and B/M dimensions.[6] The portfolios, which are constructed at the end of each June, arise from the intersection of five portfolios formed on size (market equity) and five portfolios formed on the ratio of book equity to market equity. The size breakpoints for year t are the NYSE market equity quintiles at the end of June of t. The B/M for June of year t is the book equity for the last fiscal year-end in $t-1$ divided by market capitalization for December of $t-1$. The B/M breakpoints are NYSE quintiles as well. The portfolio returns are value-weighted averages of returns on the stocks in each group. More details on the portfolio formation procedure are provided in Davis, Fama, and French (2000). To save space, and given that these portfolios are widely used in the literature, we omit providing summary statistics for their returns.

We focus our empirical exercises on explaining average returns between the third quarter

[6]The portfolio returns are from Ken French's website.

of 1963 and the last quarter of 2004, because it is in this sample that CAPM experienced its strongest failures. However, we assume that investors start learning about the underlying factor loading as soon as portfolio data become available—that is in July 1926, which is the beginning of the monthly portfolio return series based on CRSP. The use of quarterly data is imposed by the fact that Lettau and Ludvigson's (2001) CAY, which is one of the conditioning variables in our analysis, is available only at quarterly frequency. It must be said, however, that the results that do not involve CAY (which we do not report to save space) are unaffected by the use of monthly frequency.

Our model suggests that variables that predict factor loadings also predict the equity premium.[7] We thus choose conditioning variables whose ability to predict beta or the equity premium has been demonstrated in previous studies. We follow Campbell and Vuolteenaho (2004) and use the return to the value-weighted market portfolio, the term spread, and the value spread as conditioning variables. We construct the term spread as the difference between the ten-year and the three-month constant-maturity Treasury yield (reported in the *Federal Reserve Bulletin*). For the value spread, we use the return to the HML factor from Fama and French (1993). Finally, we use Lettau and Ludvigson's (2001) CAY. This variable captures the innovations to the cointegrating relationship between consumption, the stock market, and labor income. Lettau and Ludvigson show that CAY predicts the equity premium and, when used as scaling variable, improves the pricing performance of the CAPM and the Consumption CAPM. Different subsets of these conditioning variables are included as lags in the updating Equation (4).

We also experimented with other conditioning variables such as the dividend yield, the price-earnings ratio, various Treasury yields, Moody's BAA-AAA credit spread, inflation, and the growth rate of industrial production (see Chen, Roll and Ross (1986) and Campbell (1996)). We found that these additional variables do not significantly improve the cross-sectional pricing performance of the learning model.

[7] The equity premium is pinned down by our model:
$E_t[R^M_{t+1}] = \bar{b}_t Var_t\left[R^M_{t+1}\right] R^f_t$.
Hence, the expected return on the market is proportional to $\bar{b}_t$, the cross-sectional average of the factor loadings, which in turn is a function of the conditioning variables.

To derive the filtered betas for the twenty-five portfolios, we apply the Kalman filtering procedure for a model with time-varying coefficients to the following state space:

$$R^i_{t+1} = \beta^i_{t+1} R^M_{t+1} + \eta^i_{t+1} \tag{7}$$

$$\beta^i_{t+1} = \left(1 - F^i\right) B^i + F^i \beta^i_t + \phi^{i\prime} y_t + u^i_{t+1} \tag{8}$$

$$B^i \quad : \quad \text{unobserved.} \tag{9}$$

The updating equations are provided in Appendix 2. In Section 1, we argue that η^i_{t+1} is orthogonal to R^M_{t+1}, which is a necessary for applying the Kalman filter. To improve efficiency, we impose the restriction from our model and estimate Equation (7) without a constant. Some readers were concerned that omitting the intercept would play in favor of the null (that is, the learning-CAPM), as the mean of the estimated betas would capture the component of average returns that is not explained by the market risk factor. In fact, our results are insensitive to the inclusion of an intercept in Equation (7).

The filter is estimated between 1926:Q3 and 2004:Q4. All conditioning variables except CAY are available in the whole sample. CAY's series extends between 1951:Q4 and 2003:Q2. When CAY is not available, the filter is estimated using the other data. We use diffuse priors as initial condition for the forecast error and a prior of 1 for both $b^{ie}_{t+1|t}$ and $B^{ie}_{t+1|t}$ for all portfolios. The Kalman filter is described in detail in Hamilton (1994), and we adopt the maximum likelihood estimation techniques from Koopman and Durbin (2001).

One identifying assumption of our model is that the innovations u^i_t are uncorrelated with the innovations η^i_t for each asset and that error terms are uncorrelated across assets. Correlation of betas across portfolios and correlations of betas with the expected market risk premium are thus fully captured by the set of state variables. The autoregressive parameter F^i, the standard deviations of error terms $\left(\sigma^i_\eta\right)^2$ and $\left(\sigma^i_u\right)^2$, and the loadings on the conditioning variables ϕ^i are estimated using maximum likelihood on the whole history of portfolio returns, the market return, and the state variables.

The parameter estimates of the system (7)-(9) are reported in Table 1. We estimate four

specifications of the learning model. In the first specification in Table 1, no conditioning variables are included. We then present estimates where CAY is the only conditioning variable: this specification is a learning augmented version of Lettau and Ludvigson (2001). The third model uses the conditioning variables from Campbell and Vuolteenaho (2004): the term spread, the market return, and the value spread (HML). In the last specification, we include all four conditioning variables in the estimation. In order to keep Table 1 readable, we report the parameter estimates for only four portfolios: small growth, small value, large growth, and large value.

Estimates of the idiosyncratic return variance $\left(\sigma_\eta^i\right)^2$ and the idiosyncratic beta variance $\left(\sigma_u^i\right)^2$ vary little across different specifications (reading Table 1 vertically), but do vary substantially across different portfolios (reading Table 1 horizontally). The idiosyncratic return variance $\left(\sigma_\eta^i\right)^2$ is highest for the small growth portfolio and lowest for the large growth portfolio. Small stocks and growth stocks have higher variance of idiosyncratic shocks to betas $\left(\sigma_u^i\right)^2$, except in the last specification. The coefficient of auto-regression F^i does vary substantially across different model specifications, suggesting that conditioning variables pick up some of the persistence in betas. Value stock betas are positively autocorrelated with estimates of F^i between 20% and 35% in the model without conditioning variables, whereas growth stock betas have an insignificantly negative autocorrelation around the long-run mean B^i. Deviations of value stocks' beta from the long-run mean B^i are thus more persistent. An interesting question for future research is what characteristic of value and growth firms' technology determines the different degree of persistence of their risk loadings.

For the four portfolios reported in Table 1, only the term spread is a conditioning variable that is consistently significant, whereas CAY, the market, and HML never appear significant. Each of the conditioning variables, however, is significant for at least one of the portfolios that we do not report here. Furthermore, we show in the next section that pricing errors are changing substantially with different conditioning variables, meaning that these instruments are relevant conditioning information. Betas of the large-growth portfolio depend negatively on the term premium, whereas the betas of the other portfolios depend positively on the term

premium. The term spread predicts recessions: in the postwar period, every time the term spread became negative, a recession followed (see Harvey (1989) and Estrella and Hardouvelis (1991)). This finding suggests that expected returns of large-growth stocks increase relative to other portfolios when the market expects a recession. The betas of growth stocks load negatively on CAY, whereas value stocks load positively on CAY. Lettau and Ludvigson (2001) argue that CAY is high during recessions and low in expansions, suggesting that the value premium is partially explained by the increase in betas during recessions. In our context, loadings of betas on HML and the market are small in magnitude and insignificant, which makes them hard to interpret.

Incidentally, the reader may wonder why we assume that agents learn about B^i and estimate it in a recursive way, while we obtain F^i from the whole sample of data. The estimation of F^i from the whole sample corresponds to the hypothesis that agents know F^i, which in turn is based on the implicit assumption that this parameter is stable over time. Consistent with this assumption, the econometrician needs to estimate F^i efficiently, that is, using the entire data set. In general, the Kalman filter methodology postulates that the parameters governing the evolution of the state variable are constant. Assuming that they are time-varying would just superimpose another layer of filtering. At some point, one has to assume that some structural parameters are constant. In the specific application of this paper, the assumption of constant F^i seems to be borne out by the data. In unreported analysis, we assume that beta follows an autoregressive process and estimate the model using five-year rolling windows of monthly data. The auto-regression parameter appears to be fairly stable over time. Hence, we feel reassured that our preferred specification captures the salient features of the data.

Let us now turn directly to the filtered series of the factor loadings. Figure 1 plots the one-step-ahead forecast level $b^{ie}_{t+1|t}$ and forecast long-run mean $B^{ie}_{t+1|t}$ series resulting from the Kalman filter for the four portfolios of Table 1. The graphs show a pattern that has been pointed out in Franzoni (2002 and 2009): starting from the late 1940s, there was a

downward trend in the betas of value and small stocks (see portfolios 1:1, 1:5, and 5:1). The flip side of the coin is that the beta of large-growth stocks displays an upward trend. However, relative to the rolling-window OLS estimates of Franzoni, our Kalman-filter betas vary more slowly. This fact is due to the Bayesian updating contained in the Kalman filter and to the assumption of mean reversion, both of which cause some positive weight to be attached to past levels of beta. The graphs also show how the forecast level of beta (solid line) is anchored to the forecast level of the long-run mean, which in turn evolves in a smooth fashion (dashed line).

Table 2 provides evidence on the relationship between OLS beta estimates and Kalman-filtered betas. Panel A contains the OLS beta estimates from time-series regressions in the sample between 1963:Q3 and 2004:Q4, while Panels B and C have summary statistics on the $\beta^{ie}_{t+1|t}$ series in the same period. The crucial point to notice is that, when there is evidence of a descending trend in beta, as for value and small portfolios, the two filtered beta series tend to lie above the OLS estimates. This occurs because the mean-reversion assumption embedded into the filter causes the current estimate of beta to be affected by the high levels of loading from the past. Consider, for example, portfolio 1:5 (small value): the OLS estimate of beta is 1.19, while the Kalman-filtered level of beta across models has a mean value always above 1.50. The opposite happens for large-growth portfolios, although to a smaller extent.

The wedge that learning introduces between Kalman-filtered betas and OLS estimates is at the core of our account of the mispricing. The fact that, for value portfolios, the Kalman filtered beta is higher than the OLS estimate partly explains the value premium in terms of expected return.

We need to point out that for the wedge between Kalman filtered betas and OLS estimates to exist, it is important that we let the Kalman filter start at the beginning of the longer sample. It is in the years up to the early 1960s that the betas of value and small stocks had very large realizations, which affected the forecast of the long-run mean thereafter. Accordingly, the assumption of mean reversion is also crucial for obtaining a wedge. We have estimated a separate model in which we allow for non-stationarity in beta. In this case,

the anchoring effect of the long-run mean disappears, and the filtered beta is very close to the betas that can be obtained from rolling-window OLS estimation.

The estimates of OLS and Kalman-filtered betas in Table 2 allow an approximate quantification of the learning component of the unconditional alpha, which is contained in equation (6). As an example, let us focus on small–value stocks (portfolio 1:5). In the 1963:Q3-2004:Q4 sample the OLS beta of this portfolio is 1.19 (see Panel A of Table 2). Instead, the average Kalman-filtered beta is between 1.51 and 1.55, depending on the specification. We can approximate the forecast of the long-run mean with the Kalman filtered beta, as the two are very close. Given an estimate from average returns of 1.56 for the quarterly equity premium in this sample, one can calibrate the learning component to be around

$$(1.52 - 1.19) \times 1.56 = 0.51,$$

which is about one-quarter of the quarterly OLS alpha for this portfolio. Table 3 contains the results for all twenty-five size- and B/M-sorted portfolios. Panel A has the unconditional OLS alphas, whereas Panel B reports the learning component using the Kalman-filtered betas from Panel E of Table 2. For portfolios that are less mispriced than portfolio 1:5, the fraction of alpha due to learning is even more important.

Next, we quantify the covariance component of the unconditional alpha in Equation (6). As in Lewellen and Nagel (2006), we estimate the unconditional covariance between the conditional beta and the market risk premium as the sample covariance between estimated betas and realized excess market returns. The implicit assumption is that the sampling error in betas is orthogonal to market returns. We use the Kalman filter estimates of beta that rely on the largest set of conditioning variables. The results are in Panel C of Table 3. Like Lewellen and Nagel, we find that the covariance component is not large enough to explain the unconditional alpha. This finding underscores the importance of learning to ultimately explain our results. Unlike these authors, however, we find that the covariance component has the correct sign to explain the unconditional alphas. That is, value-stock betas positively covary with the market risk premium, while the opposite holds for growth

stocks' betas. This evidence is consistent with Lettau and Ludvigson (2001). The difference with respect to Lewellen and Nagel may be due to the fact that Kalman-filtered estimates of conditional betas are possibly less noisy than these authors' high-frequency estimates.

To conclude, we point out that the learning and covariance components in Panels B and C of Table 3 are not expected to sum to the estimate of the unconditional alpha in Panel A. Indeed, equation (6) is derived using population moments under the null hypothesis that our version of the conditional CAPM holds. The sum of the learning and covariance components in Table 3 is certainly different from the estimated alphas because of sampling error. Hence, a thorough statistical test of the pricing ability of our learning-augmented version of the conditional CAPM is in order. We carry out this task in the next section.

3. Pricing Errors of the Learning-CAPM

In this section, we use the learning-augmented version of the conditional CAPM to explain average returns on the twenty-five size- and B/M-sorted portfolios.

The returns to be explained span the interval between the third quarter of 1963 and the last quarter of 2004. There are two reasons to give more attention to start the sample in 1963. First, the major failures of CAPM, the "value premium" and the "small firm effect," are detected in these data. As we wish to investigate the extent to which learning about betas can account for these anomalies, it is this sample that interests us. Second, the key element in our approach is the wedge between true riskiness, which is not observed, and expected riskiness, which determines the premium required by investors. This wedge positively depends on past changes in the underlying factor loadings. So the wedge is more likely to be a significant determinant of expected returns in the second sub-sample, as the more drastic changes in the beta of value and small stocks occurred in the initial forty years of data.

Lewellen and Nagel (2006) suggest that the good performance of conditional asset pricing models may result from the cross-sectional design adopted for the test of these models and

the failure to impose theoretical constraints in the estimation process. They argue that time-series tests are therefore more suitable to these models. Accordingly, we choose to assess the performance of the learning-augmented version of CAPM in a time-series framework.

The fact that betas are estimated with a Kalman filter raises two methodological issues, which are not present in standard time-series tests of CAPM. First, portfolio mispricing cannot be computed as the intercept in a time-series OLS regression. Hence, we proceed as follows. Each quarter we estimate a conditional alpha for portfolio i as:

$$\hat{\alpha}^i_{t+1} = R^i_{t+1} - \hat{\beta}^{i\ kf}_{t+1|t} R^M_{t+1}, \tag{10}$$

where $\hat{\beta}^{i\ kf}_{t+1|t}$ is the one-period-ahead forecast of beta made at time t, which results from the Kalman filter. Then, the final estimate of the pricing error for portfolio i, $\hat{\alpha}_i$, equals the time-series mean of $\hat{\alpha}^i_{t+1}$. Analogously to the Fama and MacBeth (1973) approach, the standard error for this estimate of mispricing is computed as the standard error of the mean. In Appendix 5, we show that the probability limit of $\hat{\alpha}_i$ is zero under the null hypothesis that the Learning-CAPM holds.

The second issue has to do with the tests for the joint significance of the pricing errors, such as the Gibbons, Ross, and Shanken (1989) finite sample test and the asymptotic chi-squared tests (see Cochrane (2001)), which cannot be applied to our estimates of pricing errors because the above mentioned tests are valid only for OLS estimates. Therefore, we provide bootstrapped p-values of two summary statistics of aggregate pricing error. The first summary measure is simply the square root of the mean-squared pricing errors (RMSE). While this statistic gives equal weight to each pricing error, it has the appealing feature of corresponding to the objective function that is minimized in a least-square problem, and it therefore has an intuitive interpretation. Following Campbell and Vuolteenaho (2004), our second statistic is the composite pricing error (CPE), which is defined as $\hat{\alpha}'\hat{\Omega}^{-1}\hat{\alpha}$, where $\hat{\alpha}$ is an N-dimensional vector of the portfolio pricing errors $\hat{\alpha}_i$, and $\hat{\Omega}$ is a diagonal matrix with estimated return variances on the main diagonal. This measure of aggregate pricing error gives less weight to the alphas of more volatile portfolios. The authors suggest that

this statistic is better behaved than the Hansen and Jagannathan (1997) distance measure, in which a freely estimated variance-covariance matrix of returns is replaced for $\hat{\Omega}$, because the large number of test assets magnifies the estimation error in the inverse matrix. The bootstrapping procedure of these statistics adjusts returns to be consistent with the pricing model before random samples are generated. Furthermore, the bootstrapped distribution is conditioned on the Kalman-filtered beta series. Accounting for estimation error in betas would increase the variance of the statistic. Given that our asset pricing tests fail to reject the null hypothesis of zero pricing errors, a more dispersed distribution would only confirm this conclusion. Finally, the choice of the identity matrix and the diagonal matrix $\hat{\Omega}$ as weighting matrices in the quadratic forms of the two summary statistics allows comparisons of the performance among different asset pricing models. This would not have been possible if we had used the variance-covariance of pricing errors, which is affected by the choice of the asset pricing model.

We consider four specifications of the learning-augmented version of CAPM. The first specification does not include any conditional variable in the information set for the estimation of beta besides the history of realized returns. The performance of this model is assessed against that of the corresponding CAPM formulation without learning—that is, the unconditional CAPM. The other three models specify three alternative sets of conditioning variables in the state equation that defines the evolution of beta (the y_t variables in Equation (4)). These models are contrasted to the corresponding conditional CAPM specification without learning on beta. The choice of state variables reflects the results in the literature on conditional models and is discussed in the previous section.

Although we consider learning about unobservable factor loadings as the appropriate complement to conditional models, we want to start our analysis from two models that do not include conditioning variables in order to assess learning's sole contribution in reducing the pricing errors. Table 4 compares the pricing errors generated by the unconditional CAPM (Panel A) with those from the Learning-CAPM with no conditioning variables (Panel B). In Panel A, the pricing errors are the intercept from time-series regressions, and their t-statistics

are computed accordingly. The computation of the pricing errors and t-statistics for the learning model in Panel B is the outcome of the procedure described above. At the bottom of the table, we provide the two summary measures of the aggregate mispricing (RMSE and CPE) along with their bootstrapped p-values.[8] Panel A of Table 4 confirms known results in asset pricing. Value stocks display positive and significant mispricing, which in the case of the small-value portfolio is about 8% annually. On the other hand, growth stocks have a discount in returns, which is marginally significant for the small-growth portfolio.

Moving from Panel A to Panel B shows that pricing errors tend to decrease. In particular, accounting for learning reduces the alpha of the small-value portfolio by about 25%, from 2.05% to 1.63% quarterly. The reduction in the mispricing of the other value portfolios is even larger, and the alpha of the large-value portfolio decreases to about zero. Also on the growth side of the spectrum, the alphas are generally reduced by the learning model, except for the small-growth portfolio. The lack of a decrease in the absolute mispricing of this group of stocks originates from the fact that while this portfolio has a discount in returns, its beta displays a decreasing trend (see Figure 1). Hence, there is no wedge between the OLS and the Kalman filter estimates of beta. The small-growth portfolio has proven to be problematic in the context of other studies based on these test assets (Fama and French (1993), and Campbell and Vuolteenaho (2004)), suggesting that the small size of these stocks, and probably their low liquidity, imposes a separate account of their returns.

The general reduction in pricing errors achieved by the learning model in Table 4 is confirmed by looking at the two aggregate measures. In moving from the unconditional CAPM to the Learning-CAPM, the RMSE decreases by about 25%, whereas the CPE drops by about 45%. Most importantly, the bootstrapped p-values for these two statistics indicate that, in the case of the learning model, they are only marginally significant.

The first two graphs in Figure 2 give a visual impression of the reduction in pricing errors achieved by the learning model relative to the unconditional CAPM. Moreover, the standard error bands provide direct evidence that the variance of the estimates does not increase from

[8] In the case of the OLS models, the bootstrapping procedure reestimates beta at each draw. The sampling error in the estimation of beta is therefore taken into account.

one model to the other, suggesting that the larger p-values for the summary measures of mispricing in Panel B are not the result of increased sampling error.

Among the conditional asset pricing models, the one by Lettau and Ludvigson (2001) has recently received much attention due to its success in substantially improving the pricing performance of the CAPM and the Consumption CAPM. As mentioned above, Lewellen and Nagel (2006) argue that the correct framework for testing this model is a time-series one. Accordingly, Panel A of Table 5 presents the time-series tests of a conditional asset pricing model, in which beta is assumed to be a linear function of Lettau and Ludvigson's CAY. This approach is equivalent to a time-series regression of portfolio returns on two factors: 1) the market and 2) the market scaled by lagged CAY. It is evident that, although the model decreases mispricing relative to the unconditional CAPM, the improvement is not substantial, consistent with Lewellen and Nagel's (2006) prediction. For example, the RMSE is reduced by just 20%.

The point of this paper is to show that the combination of conditioning variables and learning about time-varying factor loadings achieves a significant reduction in pricing errors. This argument finds a large part of its support from Panel B of Table 5, which presents the alphas from the learning model computed using CAY as conditioning variable. The comparison with Panel A of the same table suggests that learning improves the pricing performance of conditional models. All pricing errors, except for the small-growth portfolio, are smaller in Panel B. This impression is confirmed by the fact that the RMSE is reduced by about 23% and the CPE drops by about 45%. Also important, these two statistics are no longer significantly different from zero. Even more striking, there is a substantial improvement in the pricing performance in Table 4 with respect to both models, which do not use conditioning variables. As an example, the decrease in the CPE relative to the unconditional CAPM is a remarkable 75%.

As discussed in Section 2, the conditional asset pricing literature has proposed other conditioning variables besides CAY to instrument betas. We would like to show that the reduction in mispricing achieved by combining learning with conditional information is not

specific to the choice of CAY. Among the "usual suspects," and excluding CAY, we have selected the excess return on the market, the return on HML, and the term spread. Interacting the lags of these variables with the market return generates three additional factors in the OLS tests of CAPM. Panel A of Table 6 reports the alphas from this version of conditional CAPM. Overall, the evidence suggests that these variables are not by themselves effective in reducing CAPM pricing errors. Indeed, the RMSE and CPE are only slightly smaller than the ones in Panel A of Table 4. However, when these variables are used as state variables for beta in the Kalman filter, the performance of the resulting Learning-CAPM significantly improves relative to the models in Table 4 (see Panel B of Table 6). With respect to Panel A of that table, the RMSE is reduced by about 31% and the CPE by about 59%. Also, with respect to Panel B of Table 4, the reduction is sensible. Therefore, we infer that the ability of learning to reduce pricing errors is not peculiar to the use of CAY as conditioning variable.

The evidence in the previous tables suggests that combining CAY with the three other conditioning variables should further improve the performance of the Learning-CAPM. This specification is considered in Table 7. Panel A of the table reports the pricing errors for the OLS conditional model. The results indicate that this conditional version of CAPM does not reduce pricing errors more than the specification with CAY alone. The p-values for the statistics RMSE and CPE state the rejection of the asset pricing model.

Instead, Panel B of Table 7, where the four conditioning variables are incorporated into the Kalman filter for beta, tells a completely different story. In general, the pricing errors are smaller in absolute value than in Panel A, the main exception being, again, the small-growth portfolio. In particular, the premium to the value portfolios is substantially smaller than with the previous models (see also Figure 2). The alpha of the small-value portfolio is only marginally statistically significant, and it drops to about 5% in annual terms. The decrease in the aggregate pricing error relative to the unconditional CAPM and the conditional models without learning is substantial. For example, the RMSE and CPE are respectively 56% and 30% of the corresponding levels for the unconditional CAPM in Table 4. Finally, these two

statistics are not significantly different from zero, suggesting that the asset pricing model is not rejected by the data. As stated before, given that the standard error of the alphas for this model and the bootstrapped standard deviation of these two statistics are not significantly larger than that for the unconditional CAPM, which is instead rejected, we are not inclined to impute the failure to reject to the lack of statistical power with respect to the tests of other CAPM specifications.[9]

Finally, it is interesting to compare the performance of our theoretically motivated learning model to that of Fama and French's (1993) empirical model, which is known to perform well on this set of test assets. Table 8 contains the alphas from the three-factor model. The Fama-French model produces smaller pricing errors for value stocks than our learning specification in Panel B of Table 7. It does worse, however, in pricing growth portfolios. In terms of aggregate mispricing, the Fama-French specification yields a RMSE that is only 19% smaller than the learning model of Table 7. The difference in the CPE is larger, because the three-factor model reduces the alphas of value portfolios, which have comparatively lower return volatility and therefore receive more weight. Incidentally, the rejection of the Fama-French model in spite of the smallest aggregate mispricing is imputed to the fact that the three factors absorb most of the residual variance and therefore increase the power of the tests.

In conclusion, the evidence suggests that incorporating learning into conditional models pushes the pricing performance of CAPM substantially closer to that of the best multifactor models.

4. Related Literature

In terms of modeling strategy, our work has drawn inspiration from Timmermann (1993) and Lewellen and Shanken (2002). In particular, Lewellen and Shanken (2002) show that

[9]For example, the bootstrapped standard deviation of the CPE statistic for the unconditional CAPM in Panel A of Table 4 is 0.034, while for the learning model in Panel B of Table 7, it is 0.038. The two standard deviations are therefore very close.

learning about the unobservable means of the dividends generates both predictability and excess variance in returns. While placing our paper in this line of research, we innovate by focusing on uncertainty about second moments. The advantage of this choice is to capture the impact of the long-run behavior of risk for today's expected returns. This element is missed if one focuses only on the long-run mean of the dividend process. In the present formulation of the paper, investors eventually learn B^i, which is an unknown constant. So, as in Lewellen and Shanken, the component of the value premium that originates from learning is expected to disappear asymptotically. Because β_t^i varies over time, learning the long-run mean of B^i can potentially take a long time. One could modify the model to allow time-variation in the long-run mean of the factor loading B^i as well. In this case, investors would never fully learn about the long-run mean and there would always be a non-negligible learning component in the alphas.

Rationality implies that conditional expectations evolve according to Bayes' rule, which naturally leads to modeling the evolution of investors' expectations via the Kalman filter. Multivariate GARCH models are an econometric alternative to Kalman-filtered estimates of beta. For example, Engle, Bollerslev, and Wooldridge (1988) and Engle, Lilien, and Robins (1987) test a CAPM models where the time-variation in beta is modeled as a (multivariate) GARCH. Engle and Lee (1999) adopt a model of volatility that is closely related to ours. They model the volatility of the market return as an autoregressive process that is reverting back to a slowly time-varying mean. Instead of modeling the time variation of volatilities, we model the time variation of betas. Schwert and Seguin (1990) also model the time variation of covariances in a GARCH framework and test the CAPM with time-varying betas on the universe of size-sorted portfolios. The main econometric difference in our approach is that it allows idiosyncratic movements in beta.

When second moments are constant, their estimation can be arbitrarily improved by increasing the frequency of observations (Merton (1980)). When second moments are time-varying, increasing the frequency of observations can lead to improved efficiency even when the wrong structural model of second moments is imposed (Nelson (1992) and Diebold,

Andersen, Bollerslev, and Labys (2003)). The driving force for our results is the evolution of the expectation of beta's long-run mean. An increase in the frequency of the data does not help in making better inference about the long-run mean of beta. Only a longer timeseries can improve investors' inferences about B^i.

5. Conclusions

In deriving factor pricing models, the existing literature has proceeded in one of two ways: either by assuming that second moments and risk premia are constant to derive unconditional restrictions, or by modeling the evolution of conditional moments as constant functions of state variables. We extend the latter approach by introducing unobserved, long-run movements in beta. This timevariation affects investors' expectation of the risk factor loading. Eventually, a long history of returns does matter in determining the required return on characteristic-sorted portfolios. Under a normality assumption, it follows naturally from Bayes' rule that betas have to be estimated via the Kalman Filter.

We perform time-series asset pricing tests on the size and B/M sorted portfolios. The introduction of learning into standard conditional CAPM models, by estimating betas with the Kalman filter, reduces pricing errors substantially. Whereas standard conditional CAPM formulations are rejected in the time-series, our learning augmented specifications are not rejected.

We have made some simplifying assumptions that could be relaxed in future research. Throughout the paper, we have assumed that betas evolve as linear, autoregressive processes with homoskedastic innovations. Nonlinear Bayesian estimation methods such as Markov-Chain Monte-Carlo would allow us to relax the assumption of linearity and constant variances of innovations. We have also simplified our approach by assuming only one priced factor. It is straightforward to extend our methodology to a setting with multiple factors. One appealing feature of our approach, however, is to show that a conditional one-factor model can go a long way in explaining stock returns, once learning about systematic risk is introduced.

Appendix 1: Derivation of the Learning-CAPM

In this appendix, we present a framework that derives Equation (1) from the basic principles of no-arbitrage. We denote excess returns to asset i by R^i_{t+1}. Ross (1976a and 1976b) shows—in what has become to be known as the fundamental theorem of asset pricing—that the absence of arbitrage implies the existence of a strictly positive pricing kernel M_{t+1} such that

$$E_t\left[M_{t+1}R^i_{t+1}\right] = 0. \tag{A-1}$$

Most asset pricing models can be derived from Equation (A-1) with an appropriate specification of the stochastic discount factor. We assume that unexpected returns depend linearly on surprises to the stochastic discount factor:

$$R^i_{t+1} - E_t\left[R^i_{t+1}\right] = -b^i_{t+1}\left(M_{t+1} - E_t\left[M_{t+1}\right]\right) + \varepsilon^i_{t+1}, \tag{A-2}$$

where ε^i_{t+1} denotes idiosyncratic risk, which is a random variable that is independently distributed of M_{t+1}, while b^i_{t+1} is the loading of asset returns on the pricing kernel. The assumption that return surprises depend linearly on innovations to the discount factor is standard. It holds naturally in continuous time and is also assumed in the derivation of the APT. An implicit assumption in (A-2) is that innovations to the pricing kernel depend on only one source of risk. We make this assumption as we focus on pricing models with a single risk factor in our empirical application. However, we could have assumed that shocks to M_{t+1} are subject to shocks from multiple sources. Everything that follows could easily be extended to such a setting.

We denote the value-weighted excess return on the market by R^M_{t+1} and assume that the idiosyncratic risk in factor loadings b^i_{t+1} averages out crosssectionally. The value-weighted cross-sectional average of factor loadings b^i_{t+1} is thus known at time t and we denote it by $\bar{b}_t$. We further assume that the value-weighted idiosyncratic risk ε^i_{t+1} averages to zero across assets. The value-weighted average of Equation (A-2) then implies the following unexpected

return to the market portfolio:

$$R_{t+1}^{M} - E_t\left[R_{t+1}^{M}\right] = -\bar{b}_t\left(M_{t+1} - E_t\left[M_{t+1}\right]\right). \tag{A-3}$$

Surprises in the market return thus depend linearly on surprises in the pricing kernel. We have not restricted the sign of $\bar{b}_t$, but we would expect it to be positive in normal times: a surprisingly high realization of the stochastic discount factor leads to stronger discounting and thus lower returns.

One can replace equation (A-3) into equation (A-2) to obtain an expression of stock returns as a function of the market return:

$$R_{t+1}^{i} - E_t\left[R_{t+1}^{i}\right] = \beta_{t+1}^{i}\left(R_{t+1}^{M} - E_t\left[R_{t+1}^{M}\right]\right) + \varepsilon_{t+1}^{i}, \tag{A-4}$$

where $\beta_{t+1}^{i} = b_{t+1}^{i}/\bar{b}_t$ is the loading on the market factor for time $t+1$ returns on asset i.

Now one can derive expressions for $E_t\left[R_{t+1}^{i}\right]$ and $E_t\left[R_{t+1}^{M}\right]$ by replacing Equations (A-3) and (A-2) into the pricing equation (A-1). In particular, one obtains

$$E_t\left[R_{t+1}^{i}\right] = \beta_{t+1|t}^{ie} E_t\left[R_{t+1}^{M}\right], \tag{A-5}$$

which corresponds to equation (1) in the text. Notice that

$$\beta_{t+1|t}^{ie} = E_t\left[\beta_{t+1}^{i}\right] = cov_t\left(R_{t+1}^{i}, R_{t+1}^{M}\right)/var_t\left(R_{t+1}^{M}\right), \tag{A-6}$$

which can be easily verified using Equation (A-4) and the assumption that time $t+1$ shocks to factor loadings are idiosyncratic.

Finally, by substituting Equation (A-5) into (A-4), one obtains the following expression for excess returns on individual assets:

$$R_{t+1}^{i} = \beta_{t+1}^{i} R_{t+1}^{M} + \left(\beta_{t+1|t}^{ie} - \beta_{t+1}^{i}\right) E_t\left[R_{t+1}^{M}\right] + \varepsilon_{t+1}^{i}. \tag{A-7}$$

Hence, stock returns are determined by three components. The first component is the market return R^M_{t+1} times the systematic risk β^i_{t+1}: assets with a higher beta have more systematic risk, and investors are compensated for that risk with higher returns. The second term is the surprise in factor loadings: a positive shock to the factor loading of asset i between t and $t+1$ requires expected returns to increase and hence prices to fall, which explains the negative dependence of R^i_{t+1} on $\beta^i_{t+1} - \beta^{ie}_{t+1|t}$ surprises to asset betas. The third determinant of returns is the idiosyncratic return ε^i_{t+1}.

By denoting

$$\eta^i_{t+1} = \left(\beta^{ie}_{t+1|t} - \beta^i_{t+1}\right) E_t\left[R^M_{t+1}\right] + \varepsilon^i_{t+1}, \tag{A-8}$$

in Equation (A-7), one obtains Equation (3) in the text.

Appendix 2: Derivation of the Kalman-Filter

This section gives the details about the Kalman filter. Let's introduce the following notation:

$$\xi^i_t = \begin{pmatrix} B^i \\ \beta^i_t \end{pmatrix} \quad \tilde{F}^i = \begin{pmatrix} 1 & 0 \\ 1 - F^i & F^i \end{pmatrix} \quad H_t = \begin{pmatrix} 0 \\ R^M_t \end{pmatrix} \quad U^i_{t+1} = \begin{pmatrix} 0 \\ u^i_{t+1} \end{pmatrix} \quad \Phi^i_t = \begin{pmatrix} \mathbf{0}' \\ \phi^{i\prime} \end{pmatrix},$$

where $\mathbf{0}$ is a vector of zeros of the same dimension as y_t. Then the system of Equations (7)–(9) can be written as:

$$\xi^i_{t+1} = \tilde{F}\xi^i_t + \Phi^i y_t + U^i_{t+1} \quad \forall i \tag{A-9}$$

$$R^i_t = H'_t \xi^i_t + \eta^i_t \quad \forall i. \tag{A-10}$$

Furthermore, denote the variance-covariance matrix of the forecast error as follows:

$$\Gamma^i_{t+1|t} = E_t\left[\left(\xi^i_{t+1} - E_t\left[\xi^i_{t+1}\right]\right)\left(\xi^i_{t+1} - E_t\left[\xi^i_{t+1}\right]\right)'\right]. \tag{A-11}$$

With this notation, the Kalman filter from Hamilton (1994) can be directly applied:

$$E_t\left[\xi_{t+1}^i\right] = \tilde{F}^i E_t\left[\xi_t^i\right] + \kappa_t^i\left(R_t^i - H_t' E_t\left[\xi_t^i\right]\right) \tag{A-12}$$

$$\kappa_t^i = \tilde{F}^i \Gamma_{t|t-1}^i H_t\left(H_t' \Gamma_{t|t-1}^i H_t + \sigma_\eta^{i2}\right)^{-1}. \tag{A-13}$$

Now, using the notation introduced in Section 1, we obtain:

$$\begin{aligned}\begin{bmatrix} B_t^{ie} \\ \beta_{t+1|t}^{ie} \end{bmatrix} &= \begin{bmatrix} E_t\left[B^i\right] \\ E_t\left[\beta_{t+1}^i\right] \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 1-F^i & F^i \end{bmatrix}\begin{bmatrix} B_{t-1}^{ie} \\ \beta_{t|t-1}^{ie} \end{bmatrix} + \begin{bmatrix} 0 \\ \phi^{i\prime} y_t \end{bmatrix} + \kappa_t^i\left(R_t^i - E_{t-1}\left[R_t^i\right]\right),\end{aligned} \tag{A-14}$$

where

$$\kappa_t^i = \begin{bmatrix} 1 & 0 \\ 1-F^i & F^i \end{bmatrix}\Gamma_{t|t-1}^i\begin{bmatrix} 0 \\ R_t^M \end{bmatrix}\left(\begin{bmatrix} 0 \\ R_t^M \end{bmatrix}'\Gamma_{t|t-1}^i\begin{bmatrix} 0 \\ R_t^M \end{bmatrix} + \sigma_\eta^{i2}\right)^{-1}. \tag{A-15}$$

Writing each updating equation separately yields Equations (4) and (5) where:

$$K_t^i = \frac{\gamma_{t|t-1}^{i(1,2)}}{\gamma_{t|t-1}^{i(2,2)}\left(R_t^M\right)^2 + \sigma_\eta^{i2}} R_t^M \qquad k_t^i = \frac{\left(1-F^i\right)\gamma_{t|t-1}^{i(1,2)} + \gamma_{t|t-1}^{i(2,2)} F^i}{\gamma_{t|t-1}^{i(2,2)}\left(R_t^M\right)^2 + \sigma_\eta^{i2}} R_t^M \qquad \kappa_t^i = \begin{pmatrix} K_t^i & k_t^i \end{pmatrix}', \tag{A-16}$$

and $\gamma_{t|t-1}^{i(q,r)}$ are the $(q\text{-th}, r\text{-th})$ element of the matrix $\Gamma_{t|t-1}^i$ which evolves as:

$$\Gamma_{t+1|t}^i = \left(\tilde{F}^i - \kappa_t^i H_t^{i\prime}\right)\Gamma_{t|t-1}^i\left(\tilde{F}^{i\prime} - H_t^i \kappa_t^{i\prime}\right) + \sigma_\eta^2 \kappa_t^i \kappa_t^{i\prime} + \begin{pmatrix} 0 & 0 \\ 0 & \sigma_u^2 \end{pmatrix}. \tag{A-17}$$

Appendix 3: Covariance between η_{t+1} and R^M_{t+1}

This section proves that the conditional covariance between η_{t+1} and R^M_{t+1} in Equation (3) is zero. Let us recall Equation (3):

$$R^i_{t+1} = \beta^i_{t+1} R^M_{t+1} + \eta^i_{t+1}, \tag{A-18}$$

where

$$\eta^i_{t+1} = \left[\left(\beta^{ie}_{t+1|t} - \beta^i_{t+1}\right) E\left[R^M_{t+1}\right] + \varepsilon^i_{t+1}\right]. \tag{A-19}$$

Next, we provide the derivation of the conditional covariance between η_{t+1} and R^M_{t+1}:

$$\begin{aligned}
& Cov_t\left(R^M_{t+1}, \eta^i_{t+1}\right) \\
= \; & Cov_t\left(R^M_{t+1}, \left(\beta^{ie}_{t+1|t} - \beta^i_{t+1}\right) E_t\left[R^M_{t+1}\right] + \varepsilon^i_{t+1}\right) \\
= \; & Cov_t\left(R^M_{t+1}, -\beta^i_{t+1} E_t\left[R^M_{t+1}\right]\right) \\
= \; & Cov_t\left(R^M_{t+1}, -\left(\left(1 - F^i\right) B^i + F^i \beta^i_t + \phi^{i\prime} y_t + u^i_{t+1}\right)\right) E_t\left[R^M_{t+1}\right] \\
= \; & Cov_t\left(R^M_{t+1}, -u^i_{t+1}\right) E_t\left[R^M_{t+1}\right] \\
= \; & 0.
\end{aligned}$$

We can take the second step because we are conditioning on time t information. The final step depends on our assumption of idiosyncratic innovations in β^i_{t+1}. In other words, we assume that the covariance between factor loadings and market risk premia occurs through information known at time t. To be concrete, we assume that u^i_{t+1} in Equation (2) is uncorrelated with R^M_{t+1}. The correlation between factor loadings and the market risk premium comes through the variables y_t, which are known at time t and therefore do not matter for $Cov_t\left(R^M_{t+1}, \eta^i_{t+1}\right)$.

Appendix 4: Alpha Decomposition

This section proves the decomposition in Equation (6).

Denote the unconditional OLS beta by β^i_{OLS} and the unconditional OLS alpha estimated on the same sample by α^i_{OLS}. Then:

$$\begin{aligned}
\alpha^i_{OLS} &= E\left[R^i_{t+1}\right] - \beta^i_{OLS} E\left[R^M_{t+1}\right] \\
&= E\left[\beta^i_{t+1} R^M_{t+1}\right] - \beta^i_{OLS} E\left[R^M_{t+1}\right] \\
&= E\left[\beta^i_{t+1}\right] E\left[R^M_{t+1}\right] + Cov\left[\beta^i_{t+1}, R^M_{t+1}\right] - \beta^i_{OLS} E\left[R^M_{t+1}\right] \\
&= E\left[\beta^i_{t+1} - \beta^i_{OLS}\right] E\left[R^M_{t+1}\right] + Cov\left[\beta^i_{t+1}, R^M_{t+1}\right] \\
&\quad \underbrace{E\left[\beta^i_{t+1} - \beta^i_{OLS}\right] E\left[R^M_{t+1}\right]}_{\text{Learning component}} + \underbrace{Cov\left[\beta^i_{t+1}, R^M_{t+1}\right]}_{\text{Covariance component}}.
\end{aligned}$$

Appendix 5: Plim $\hat{\alpha}_i = 0$ under Learning-CAPM

In this section, we prove that, under the null of the Learning-CAPM, the sample average of the alphas defined in Equation (10) converges to zero. For the law of large numbers, the sample average converges to the unconditional expectation of α^i_{t+1}:

$$\begin{aligned}
& E\left[\alpha^i_{t+1}\right] \\
= \; & E\left[R^i_{t+1}\right] - E\left[\beta^{i\ kf}_{t+1|t} R^M_{t+1}\right] \\
= \; & E\left[\beta^i_{t+1}\right] E\left[R^M_{t+1}\right] + Cov\left[\beta^i_{t+1}, R^M_{t+1}\right] \\
& -E\left[\beta^{i\ kf}_{t+1|t}\right] E\left[R^M_{t+1}\right] - Cov\left[\beta^{i\ kf}_{t+1|t}, R^M_{t+1}\right].
\end{aligned}$$

The Kalman filter estimate of β_{t+1}^i is equal to the conditional expectation of β_{t+1}^i plus orthogonal sampling error, which can be ignored. Hence, we obtain

$$\begin{aligned}
& E\left[\alpha_{t+1}^i\right] \\
= \; & E\left[\beta_{t+1}^i\right] E\left[R_{t+1}^M\right] + Cov\left[\beta_{t+1}^i, R_{t+1}^M\right] \\
& -E\left[E_t\left[\beta_{t+1}^i\right]\right] E\left[R_{t+1}^M\right] - Cov\left[E_t\left[\beta_{t+1}^i\right], R_{t+1}^M\right] \\
= \; & Cov\left[\beta_{t+1}^i, R_{t+1}^M\right] - Cov\left[E_t\left[\beta_{t+1}^i\right], R_{t+1}^M\right],
\end{aligned}$$

where we can take the last step thanks to the law of iterated expectations. Finally, one can use the expression for β_{t+1}^i in Equation (2) and the fact that u_{t+1}^i is an idiosyncratic shock to prove that $E\left[\alpha_{t+1}^i\right] = 0$.

The argument above proves that the definition of α_{t+1}^i in equation (6) is the correct way to go about testing our Learning-CAPM.

References

Ang, A., and J. Chen, 2007, "CAPM Over the Long-Run: 1926-2001," *Journal of Empirical Finance*, 14, 1-40.

Barberis, N., and A. Shleifer, 2003, "Style Investing," *Journal of Financial Economics*, 68, 161-199.

Bollerslev, T., R. Engle, and J. Wooldridge, 1988, "A capital asset pricing model with time varying covariances," *Journal of Political Economy*, 96, 116-131.

Breeden, D., 1979, "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265-296.

Campbell, J., 1996, "Understanding Risk and Return," *Journal of Political Economy*, 104, 298-345.

Campbell, J., and T. Vuolteenaho, 2004, "Bad Beta, Good Beta," *American Economic Review*, 94, 1249-75.

Chan, Y., and N. Chen, 1991, "Structural and return characteristics of small and large firms," *Journal of Finance*, 46, 1467-1484.

Chen, N., R. Roll, and S. Ross, 1986, "Economic Forces and the Stock Market," *Journal of Business*, 59, 383-403.

Cochrane, J., 2001, *Asset Pricing*, Princeton University Press, Princeton, NJ.

Davis, J., E. Fama, and K. French, 2000, "Characteristics, Covariances, and Average Returns: 1929-1997," *Journal of Finance*, 55, 389-406.

Diebold, F., T. Andersen, T. Bollerslev, and P. Labys, 2003, "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579-626.

Engle, R., T. Bollerslev and J. Wooldridge, 1988, "A Capital Asset Pricing Model with Time Varying Covariances," *Journal of Political Economy*, 96, 116-131.

Engle, R., and G. Lee, 1999, "A Permanent and Transitory Component Model of Stock Return Volatility," in: R. Engle and H. White (ed.), Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger, Oxford: Oxford University Press, 475-497.

Engle, R., D. Lilien, and R. Robins, 1987, "Estimation of Time Varying Risk Premia in the Term Structure: the ARCH-M Model," *Econometrica*, 55, 391-407.

Estrella, A. and G. Hardouvelis, 1991, "The Term Structure as a Predictor of Real Economic Activity," *Journal of Finance*, 46, 555-76.

Fama, E., and K. French, 1992, "The Cross-section of Expected Stock Returns," *Journal of Finance*, 47, 427-465.

Fama, E., and K. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3-56.

Fama, E., and K. French, 2006, "The Value Premium and the CAPM," *Journal of Finance*, 61, 2163-2185.

Fama, E. and J. MacBeth, 1973, "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607-636.

Ferson, W., and C. Harvey, 1999, "Conditioning variables and the cross-section of stock returns," *Journal of Finance*, 54, 1325-1360.

Ferson, W., S. Kandel, and R. Stambaugh, 1987, " Tests of Asset Pricing with Time-Varying Expected Risk Premiums and Market Betas," *Journal of Finance*, 42, 201-220.

Franzoni, F., 2002, "Where is beta going? The Riskiness of Value and Small Stocks," Ph.d. dissertation, MIT.

Franzoni, F., 2009, "The changing nature of market risk," working paper. Swiss Finance Institute.

Gibbons, M., S. A. Ross, and J. Shanken, 1989, "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57, 1121-52.

Hamilton, J. D., 1994, *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hansen, L., R. Jagannathan, 1997, "Assessing Specification Errors in Stochastic Discount Factor Models," *Econometrica*, 55, 587-614.

Harvey, C., 1987, "The Real Term Structure and Consumption Growth," *Journal of Financial Economics* 22, 305-334.

Harvey, C., 1989, "Time-Varying Conditional Covariances in Tests of Asset Pricing Models," *Journal of Financial Economics*, 24, 289-317.

Jagannathan, R., and Z. Wang, 1996, "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance*, 51, 3-53.

Jegadeesh, N., and S. Titman, 1993, "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance*, 48, 65-91.

Jostova, G., and A. Philipov, 2005, "Bayesian Analysis of Stochastic Betas," *Journal of Financial and Quantitative Analysis,* 40, 747-778.

Koopman S. J., and J. Durbin, 2001, *Time Series Analysis by State Space Methods*, Oxford University Press.

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

Lewellen, J., and S. Nagel, 2006, "The Conditional CAPM Does Not Explain Asset-Pricing Anomalies," *Journal of Financial Economics*, 82, 289-314.

Lewellen, J., S. Nagel, and J. Shanken, 2007, "A Skeptical Appraisal of Asset Pricing Tests," manuscript.

Lewellen, J., and J. Shanken, 2002, "Learning, Asset-pricing Tests, and Market Efficiency," *Journal of Finance 57*, 1113-1145.

Lucas, R., 1978, "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1429-1446.

Merton, R. C., 1973 "An Intertemporal Asset Pricing Model," *Econometrica*, 41, 867-887.

Merton, R. C., 1980, "On Estimating the Expected Return on the Market: an Exploratory Investigation," *Journal of Financial Economics*, 8, 323-361.

Nelson, D., 1992, "Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model," *Journal of Econometrics*, 52, 61-90.

Ross, S., 1976a, "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341-360.

Ross, S., 1976b, "Risk, Return, and Arbitrage," in I. Friend and J. Bicksler (eds.), *Risk and Return in Finance, Volume 1*, Ballinger, Cambridge, 189-218.

Schwert, G. W., and P. Seguin, 1990, "Heteroskedasticity in Stock Returns," *Journal of Finance*, 45, 1129-1155.

Timmermann, A., 1993, "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Price," *Quarterly Journal of Economics*, 108, 1135-1145.

Table 1: Parameter Estimates. For the twenty-five size- and B/M-sorted portfolios, the table reports the maximum-likelihood parameter estimates for the parameters in system (7). Except Panel A, each panel corresponds to a different set of state variables in the state equation of the system. The estimated parameters are the autoregression coefficient F, the variance of the error in the observable equation (σ_η^2), and the variance of the error in the state equation 1 (σ_u^2). P-values are given in brackets. The Kalman filter is estimated on all available data between 1926:Q3 and 2004:Q4. CAY is available between 1951:Q4 and 2003:Q2. Portfolio returns and the other state variables (HML, term, and the market return) are available on the whole sample.

	small growth		small value		large growth		large value	
	Panel A: no cond. variables							
σ_η^2	97.45	[0.00]	51.17	[0.00]	7.11	[0.00]	26.27	[0.00]
σ_u^2	0.58	[0.06]	0.66	[0.01]	0.01	[0.00]	0.42	[0.00]
F	-0.09	[0.57]	0.20	[0.15]	-0.26	[0.48]	0.35	[0.00]
	Panel B: CAY							
σ_η^2	95.66	[0.00]	51.85	[0.00]	7.60	[0.00]	26.54	[0.00]
σ_u^2	0.62	[0.07]	0.65	[0.01]	0.00	[0.00]	0.41	[0.00]
F	-0.02	[0.91]	0.16	[0.25]	0.80	[0.00]	0.32	[0.00]
ϕ^{CAY}	-8.39	[0.43]	6.10	[0.45]	-0.10	[0.80]	5.21	[0.38]
	Panel C: HML, TERM, and MKT							
σ_η^2	98.33	[0.00]	53.00	[0.00]	7.79	[0.00]	26.93	[0.00]
σ_u^2	0.55	[0.08]	0.58	[0.00]	0.00	[0.49]	0.40	[0.00]
F	-0.14	[0.38]	-0.17	[0.20]	-0.18	[0.65]	0.20	[0.09]
ϕ^{HML}	-0.01	[0.42]	-0.02	[0.05]	0.00	[0.59]	-0.01	[0.29]
ϕ^{term}	0.11	[0.26]	0.17	[0.02]	-0.05	[0.00]	0.11	[0.02]
ϕ^{mkt}	0.00	[0.53]	0.02	[0.00]	0.00	[0.53]	0.01	[0.14]
	Panel D: CAY, HML, TERM, and MKT							
σ_η^2	96.81	[0.00]	53.11	[0.00]	7.79	[0.00]	27.02	[0.00]
σ_u^2	0.57	[0.09]	0.57	[0.00]	0.00	[0.39]	0.39	[0.00]
F	-0.09	[0.57]	0.02	[0.87]	-0.18	[0.71]	0.25	[0.01]
ϕ^{CAY}	-12.96	[0.25]	1.57	[0.86]	0.31	[0.88]	2.39	[0.70]
ϕ^{HML}	-0.01	[0.62]	-0.01	[0.28]	0.00	[0.83]	-0.01	[0.33]
ϕ^{term}	0.14	[0.16]	0.17	[0.03]	-0.05	[0.01]	0.10	[0.03]
ϕ^{mkt}	0.00	[0.87]	-0.01	[0.18]	0.00	[0.57]	0.00	[0.93]

Table 2: Summary Statistics for Betas. Panel A reports the estimates of beta and their standard errors from time-series regressions of portfolio excess returns on the value-weighted market index. The test assets are the twenty-five size- and B/M-sorted portfolios. For the same assets, Panel B reports the mean of the Kalman-filtered estimates of beta and their standard deviations, when no conditioning variables are included in the state equations. Panels C to E report means and standard deviations of Kalman-filtered estimates of beta when conditioning variables are included in the state equations. The sets of conditioning variables are given in the table. Returns are quarterly compounded returns. The sample period is 1963:Q3 to 2004:Q4.

	B/M quintile									
	1	2	3	4	5	1	2	3	4	5
Size q.	Panel A: Unconditional CAPM									
	OLS Betas					St. Error of Beta				
1	1.65	1.40	1.21	1.12	1.19	0.08	0.07	0.06	0.06	0.07
2	1.55	1.27	1.13	1.05	1.10	0.06	0.05	0.05	0.05	0.06
3	1.44	1.17	1.01	0.96	1.00	0.05	0.04	0.04	0.05	0.06
4	1.31	1.09	0.97	0.95	1.01	0.04	0.04	0.04	0.04	0.05
5	1.03	0.92	0.77	0.75	0.81	0.03	0.03	0.03	0.04	0.05
	Panel B: Unconditional Learning									
	Mean of Beta					St. Dev. of Beta				
1	1.65	1.62	1.45	1.39	1.51	0.04	0.09	0.16	0.11	0.11
2	1.38	1.29	1.23	1.26	1.41	0.07	0.13	0.20	0.10	0.19
3	1.32	1.19	1.15	1.20	1.31	0.15	0.06	0.15	0.04	0.23
4	1.19	1.08	1.10	1.16	1.43	0.17	0.08	0.03	0.11	0.14
5	0.97	0.91	0.90	0.93	1.23	0.01	0.03	0.05	0.17	0.15
	Panel C: Learning cond. on CAY									
	Mean of Beta					St. Dev. of Beta				
1	1.65	1.62	1.46	1.40	1.51	0.11	0.09	0.15	0.10	0.12
2	1.38	1.30	1.24	1.27	1.42	0.13	0.11	0.20	0.14	0.21
3	1.30	1.20	1.16	1.19	1.32	0.18	0.08	0.16	0.17	0.24
4	1.15	1.08	1.10	1.18	1.43	0.20	0.08	0.12	0.15	0.20
5	0.97	0.92	0.90	0.96	1.24	0.02	0.09	0.09	0.20	0.16
	Panel D: Learning cond. on HML, TERM, and MKT									
	Mean of Beta					St. Dev. of Beta				
1	1.74	1.96	1.63	1.56	1.55	0.32	0.66	0.30	0.31	0.26
2	1.34	1.34	1.25	1.31	1.48	0.11	0.15	0.20	0.19	0.26
3	1.31	1.19	1.16	1.21	1.39	0.15	0.06	0.16	0.16	0.41
4	1.15	1.09	1.10	1.20	1.51	0.16	0.08	0.13	0.14	0.20
5	0.97	0.90	0.91	0.96	0.92	0.05	0.02	0.08	0.18	0.28
	Panel E: Learning cond. on CAY, HML, TERM, and MKT									
	Mean of Beta					St. dev. of Beta				
1	1.65	1.62	1.50	1.44	1.52	0.19	0.20	0.22	0.23	0.21
2	1.35	1.31	1.73	1.28	1.47	0.16	0.15	0.56	0.18	0.26
3	1.31	1.19	1.16	1.17	1.35	0.18	0.06	0.18	0.18	0.26
4	1.15	1.08	1.09	1.26	1.49	0.20	0.08	0.36	0.38	0.23
5	0.97	0.92	0.91	0.98	1.26	0.05	0.09	0.10	0.20	0.18

Table 3: Alpha Components. The table reports estimates of the unconditional alpha for the twenty-five size- and B/M-sorted portfolios, as well as estimates of the Learning and Covariance components of alpha that are identified in equation (6). Panel A reports the intercept from OLS regressions of portfolio excess return on the market excess return. In Panel B, the learning component is the difference between Kalman-filtered betas and OLS betas (from Panels E and A of Table 2, respectively) times the sample mean of the excess market return. In Panel C, the covariance component is the sample covariance between the Kalman-filtered beta and the excess return on the market. The frequency is quarterly and the sample period is 1963:Q3 to 2004:Q4.

	B/M quintile				
	1	2	3	4	5
Size q.	Panel A: Unconditional Alphas				
1	-1.42	0.61	1.02	1.78	2.05
2	-0.93	0.25	1.22	1.47	1.73
3	-0.76	0.57	0.75	1.27	1.71
4	-0.26	-0.04	0.82	1.25	1.24
5	-0.23	0.08	0.33	0.59	0.54
	Panel B: Learning Component				
1	0.00	0.34	0.39	0.43	0.51
2	-0.27	0.04	0.18	0.35	0.51
3	-0.22	0.04	0.23	0.36	0.49
4	-0.25	0.00	0.19	0.35	0.65
5	-0.08	-0.01	0.20	0.33	0.66
	Panel C: Covariance Component				
1	-0.25	-0.01	-0.06	0.04	0.11
2	-0.31	-0.04	0.04	0.26	0.25
3	-0.40	0.05	0.15	0.34	0.23
4	-0.43	0.01	0.26	0.29	0.26
5	0.02	0.21	0.17	0.33	0.08

Table 4: Alphas: No Conditioning Variables. Panel A reports the intercept from OLS time-series regressions of portfolio excess returns on the value-weighted market index. The test assets are the twenty-five size- and B/M-sorted portfolios. For the same assets, Panel B reports the alphas for the learning model without conditioning variables. For each portfolio, alpha is computed as the portfolio times series mean of the time t excess portfolio return minus the time t product of the Kalman filter estimate of beta and the excess return on the market. t-statistics are given in parentheses below each estimate. For the learning models the standard errors are computed from the time-series standard deviation of time t pricing error. Below each estimate, t-statistics are given in parentheses. At the bottom of each panel the table reports three statistics: the square root of the mean-squared pricing error over the twenty-five portfolios (RMSE); the composite pricing error (CPE), which is a quadratic form in the vector of the twenty-five portfolio pricing errors, where the weighting matrix is a diagonal matrix with estimated return variances on the main diagonal; and the bootstrapped p-value for these two statistics. Returns are quarterly compounded returns. The sample period is 1963:Q3 to 2004:Q4.

	B/M quintile									
	1	2	3	4	5	1	2	3	4	5
Size q.	Panel A: OLS model					Panel B: learning model				
1	-1.42	0.61	1.02	1.78	2.05	-1.43	0.21	0.80	1.46	1.63
	(-1.97)	(0.99)	(1.82)	(3.17)	(3.17)	(-2.02)	(0.33)	(1.37)	(2.52)	(2.43)
2	-0.93	0.25	1.22	1.47	1.73	-0.59	0.31	1.09	1.13	1.24
	(-1.77)	(0.54)	(2.92)	(3.27)	(3.21)	(-1.12)	(0.65)	(2.57)	(2.44)	(2.18)
3	-0.76	0.57	0.75	1.27	1.71	-0.33	0.54	0.51	0.92	1.26
	(-1.75)	(1.66)	(2.04)	(3.09)	(3.35)	(-0.78)	(1.58)	(1.36)	(2.14)	(2.43)
4	-0.26	-0.04	0.82	1.25	1.24	0.22	-0.03	0.61	0.84	0.62
	(-0.74)	(-0.13)	(2.56)	(3.61)	(2.69)	(0.62)	(-0.11)	(1.88)	(2.31)	(1.21)
5	-0.23	0.08	0.33	0.59	0.54	-0.15	0.07	0.13	0.14	-0.01
	(-0.95)	(0.37)	(1.17)	(1.90)	(1.34)	(-0.64)	(0.31)	(0.44)	(0.42)	(-0.03)
	RMSE		1.076			RMSE		0.814		
	(p-value)		(0.008)			(p-value)		(0.041)		
	CPE		0.197			CPE		0.108		
	(p-value)		(0.009)			(p-value)		(0.057)		

Table 5: Alphas: Conditioning on CAY. Panel A reports the intercept from OLS time-series regressions of portfolio excess returns on the value-weighted market index and the market index scaled by lagged CAY. The test assets are the twenty-five size- and B/M-sorted portfolios. For the same assets, Panel B reports the alphas for the learning model with CAY as conditioning variable. For each portfolio, alpha is computed as the portfolio times-series mean of the time t excess portfolio return minus the time t product of the Kalman filter estimate of beta and the excess return on the market. For the learning models the standard errors are computed from the time-series standard deviation of time the t pricing error. Below each estimate, t-statistics are given in parentheses. At the bottom of each panel the table reports three statistics: the square root of the mean-squared pricing error over the twenty-five portfolios (RMSE); the composite pricing error (CPE), which is a quadratic form in the vector of the twenty-five portfolio pricing errors, where the weighting matrix is a diagonal matrix with estimated return variances on the main diagonal; and the bootstrapped p-value for these two statistics. Returns are quarterly compounded returns. The sample period is 1963:Q3 to 2004:Q4.

	B/M quintile									
	1	2	3	4	5	1	2	3	4	5
Size q.	Panel A: OLS model					Panel B: learning model				
1	-0.92	0.76	0.90	1.76	1.89	-1.18	0.27	0.69	1.31	1.43
	(-1.20)	(1.14)	(1.48)	(2.90)	(2.70)	(-1.66)	(0.44)	(1.18)	(2.30)	(2.18)
2	-0.67	0.17	1.04	1.15	1.35	-0.35	0.25	1.00	0.86	0.97
	(-1.19)	(0.33)	(2.31)	(2.39)	(2.34)	(-0.67)	(0.53)	(2.38)	(1.94)	(1.76)
3	-0.51	0.43	0.51	0.82	1.39	-0.14	0.48	0.38	0.57	0.99
	(-1.09)	(1.16)	(1.28)	(1.91)	(2.55)	(-0.33)	(1.41)	(1.03)	(1.43)	(1.94)
4	0.21	-0.19	0.43	0.86	0.72	0.42	-0.05	0.37	0.60	0.33
	(0.58)	(-0.56)	(1.30)	(2.38)	(1.48)	(1.25)	(-0.15)	(1.21)	(1.73)	(0.68)
5	-0.24	-0.09	0.23	0.20	0.26	-0.17	-0.11	-0.04	-0.06	-0.20
	(-0.91)	(-0.36)	(0.75)	(0.62)	(0.60)	(-0.68)	(-0.53)	(-0.13)	(-0.19)	(-0.44)
	RMSE		0.864			RMSE		0.668		
	(p-value)		(0.034)			(p-value)		(0.121)		
	CPE		0.127			CPE		0.070		
	(p-value)		(0.035)			(p-value)		(0.147)		

Table 6: Alphas: Conditioning on HML, TERM, and MKT. Panel A reports the intercept from OLS time-series regressions of portfolio excess returns on the value-weighted market index, the market index scaled by the lagged return on HML, the market index scaled by the lagged term spread, and the market index scaled by its lagged return. The test assets are the twenty-five size- and B/M-sorted portfolios. For the same assets, Panel B reports the alphas for the learning model with HML, TERM, and the lagged market index as conditioning variables. For each portfolio, alpha is computed as the portfolio times-series mean of the time t excess portfolio return minus the time t product of the Kalman-filtered estimate of beta and the excess return on the market. For the learning models the standard errors are computed from the time-series standard deviation of the time t pricing error. Below each estimate, t-statistics are given in parentheses. At the bottom of each panel the table reports three statistics: the square root of the mean-squared pricing error over the twenty-five portfolios (RMSE); the composite pricing error (CPE), which is a quadratic form in the vector of the twenty-five portfolio pricing errors, where the weighting matrix is a diagonal matrix with estimated return variances on the main diagonal; and the bootstrapped p-value for these two statistics. Returns are quarterly compounded returns. The sample period is 1963:Q3 to 2004:Q4.

	B/M quintile									
	1	2	3	4	5	1	2	3	4	5
Size q.	Panel A: OLS model					Panel B: learning model				
1	-1.44	0.57	0.94	1.75	1.94	-1.56	0.01	0.40	1.02	1.32
	(-1.97)	(0.92)	(1.69)	(3.14)	(3.06)	(-2.18)	(0.01)	(0.69)	(1.75)	(2.02)
2	-0.84	0.23	1.19	1.43	1.55	-0.66	0.12	0.91	0.93	0.87
	(-1.57)	(0.48)	(2.84)	(3.17)	(2.91)	(-1.24)	(0.26)	(2.16)	(2.03)	(1.57)
3	-0.62	0.58	0.78	1.19	1.55	-0.33	0.06	0.41	0.79	0.95
	(-1.42)	(1.66)	(2.11)	(2.92)	(3.01)	(-0.78)	(0.12)	(1.09)	(1.93)	(1.85)
4	-0.17	-0.05	0.71	1.18	1.12	0.32	-0.10	0.50	0.69	0.35
	(-0.49)	(-0.14)	(2.23)	(3.38)	(2.40)	(0.94)	(-0.31)	(1.63)	(1.94)	(0.70)
5	-0.23	-0.03	0.24	0.55	0.41	-0.08	0.10	0.17	-0.02	-0.25
	(-0.94)	(-0.12)	(0.83)	(1.74)	(1.01)	(-0.35)	(0.44)	(0.56)	(-0.05)	(-0.55)
	RMSE		1.012			RMSE		0.667		
	(p-value)		(0.009)			(p-value)		(0.133)		
	CPE		0.174			CPE		0.069		
	(p-value)		(0.009)			(p-value)		(0.165)		

Table 7: Alphas: Conditioning on CAY, HML, TERM, and MKT. Panel A reports the intercept from OLS time-series regressions of portfolio excess returns on the value-weighted market index, the market index scaled by lagged CAY, the market index scaled by the lagged return on HML, the market index scaled by the lagged term spread, and the market index scaled by its lagged return. The test assets are the twenty-five size- and B/M-sorted portfolios. For the same assets, Panel B reports the alphas for the learning model with CAY, HML, TERM, and the lagged market index as conditioning variables. For each portfolio, alpha is computed as the portfolio times-series mean of the time t excess portfolio return minus the time t product of the Kalman-filtered estimate of beta and the excess return on the market. For the learning models, the standard errors are computed from the time-series standard deviation of the time t pricing error. Below each estimate, t-statistics are given in parentheses. At the bottom of each panel the table reports three statistics: the square root of the mean-squared pricing error over the twenty-five portfolios (RMSE); the composite pricing error (CPE), which is a quadratic form in the vector of the twenty-five portfolio pricing errors, where the weighting matrix is a diagonal matrix with estimated return variances on the main diagonal; and the bootstrapped p-value for these two statistics. Returns are quarterly compounded returns. The sample period is 1963:Q3 to 2004:Q4.

	B/M quintile									
	1	2	3	4	5	1	2	3	4	5
Size q.	Panel A: OLS model					Panel B: learning model				
1	-0.93	0.76	0.89	1.77	1.88	-1.26	0.16	0.46	1.05	1.28
	(-1.21)	(1.14)	(1.48)	(2.96)	(2.77)	(-1.76)	(0.25)	(0.79)	(1.79)	(1.97)
2	-0.63	0.17	1.05	1.16	1.28	-0.40	0.15	1.10	0.78	0.76
	(-1.10)	(0.33)	(2.33)	(2.41)	(2.24)	(-0.75)	(0.33)	(1.42)	(1.74)	(1.39)
3	-0.44	0.46	0.54	0.82	1.33	-0.17	0.45	0.32	0.56	0.81
	(-0.95)	(1.23)	(1.37)	(1.92)	(2.42)	(-0.39)	(1.34)	(0.89)	(1.41)	(1.59)
4	0.24	-0.19	0.40	0.85	0.67	0.47	-0.08	0.30	0.55	0.10
	(0.66)	(-0.54)	(1.20)	(2.34)	(1.38)	(1.38)	(-0.24)	(0.72)	(1.12)	(0.19)
5	-0.25	-0.14	0.18	0.20	0.19	-0.09	-0.10	-0.06	-0.16	-0.32
	(-0.93)	(-0.59)	(0.58)	(0.63)	(0.43)	(-0.38)	(-0.48)	(-0.23)	(-0.49)	(-0.70)
	RMSE		0.852			RMSE		0.607		
	(p-value)		(0.032)			(p-value)		(0.205)		
	CPE		0.123			CPE		0.059		
	(p-value)		(0.030)			(p-value)		(0.253)		

Table 8: Alphas: Three-Factor Model. The table reports the intercept from OLS time-series regressions of portfolio excess returns on the three Fama and French (1993) factors: the excess return on the value-weighted market index, HML, and SMB. The test assets are the twenty-five size- and B/M-sorted portfolios. Below each estimate, t-statistics are given in parentheses. At the bottom, the table also reports: the square root of the mean-squared pricing error over the twenty-five portfolios (RMSE); the composite pricing error (CPE), which is a quadratic form in the vector of the twenty-five portfolio pricing errors, where the weighting matrix is a diagonal matrix with estimated return variances on the main diagonal; and the bootstrapped p-value for these two statistics. Returns are quarterly compounded returns. The sample period is 1963:Q3 to 2004:Q4.

	B/M quintile				
Size q.	1	2	3	4	5
1	-1.62	-0.14	0.14	0.77	0.59
	(-4.52)	(-0.54)	(0.55)	(3.03)	(2.19)
2	-0.93	-0.38	0.42	0.36	0.27
	(-3.44)	(-1.65)	(1.87)	(1.61)	(1.23)
3	-0.51	0.09	-0.06	0.21	0.33
	(-2.08)	(0.37)	(-0.26)	(0.83)	(1.23)
4	0.24	-0.48	0.15	0.43	0.08
	(1.00)	(-1.70)	(0.60)	(1.71)	(0.28)
5	0.31	0.01	0.01	-0.06	-0.31
	(1.68)	(0.05)	(0.06)	(-0.29)	(-1.00)
	RMSE		0.494		
	(p-value)		(0.000)		
	CPE		0.033		
	(p-value)		(0.017)		

Figure 1: Kalman-Filtered Betas. For selected portfolios and for two specifications of the Kalman filter, the graphs report the Kalman-filtered series of the level of beta (solid line) and of the long-run mean of beta (dashed line). The selected portfolios are the small growth (11), the small value (15), the large growth (51), and the large value (55). The two specifications are the one where no conditioning variables are included in the state equations (BETA and B series) and the one with all conditioning variables (CAY, HML, term, and market return) included in the state equation (BETAALL and BALL series). The sample on which the series are estimated is between 1926:Q3 and 2004:Q4. The series are graphed between 1931:Q1 and 2004:Q4.

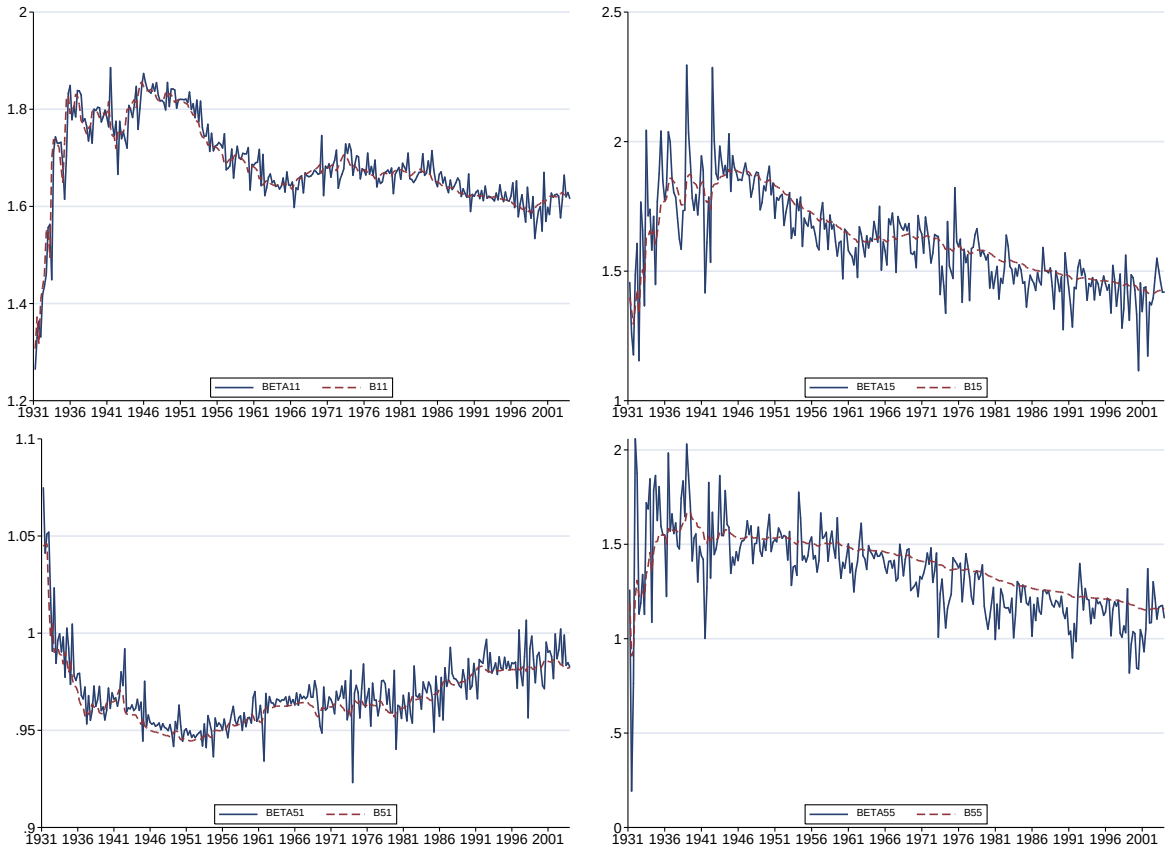

Figure 1: (continued)

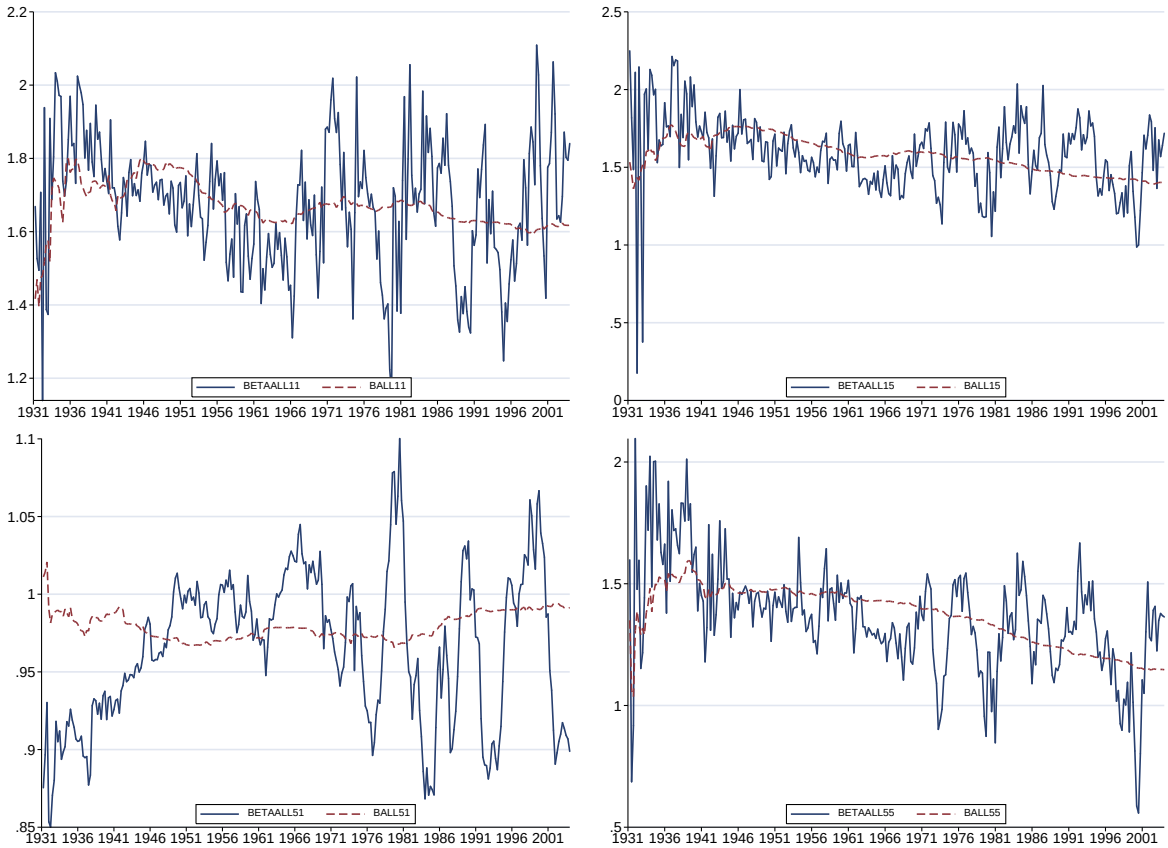

Figure 2: Portfolio Pricing Errors. The figure contains graphs of the pricing error for the twenty-five size- and B/M-sorted portfolios generated by the different asset pricing models considered in Tables 4 to 7. The dashed lines depict a two-standard-error band around zero.

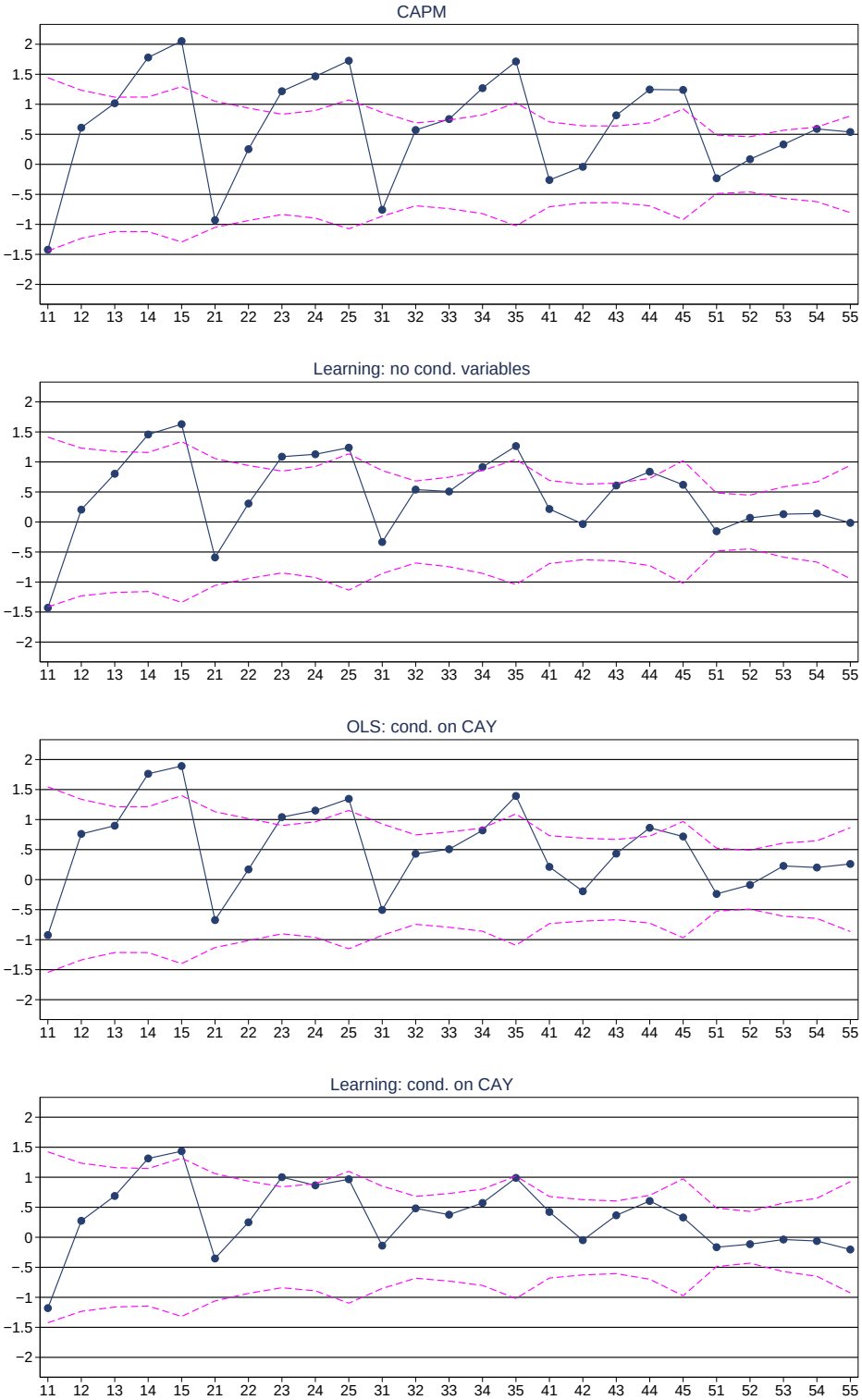
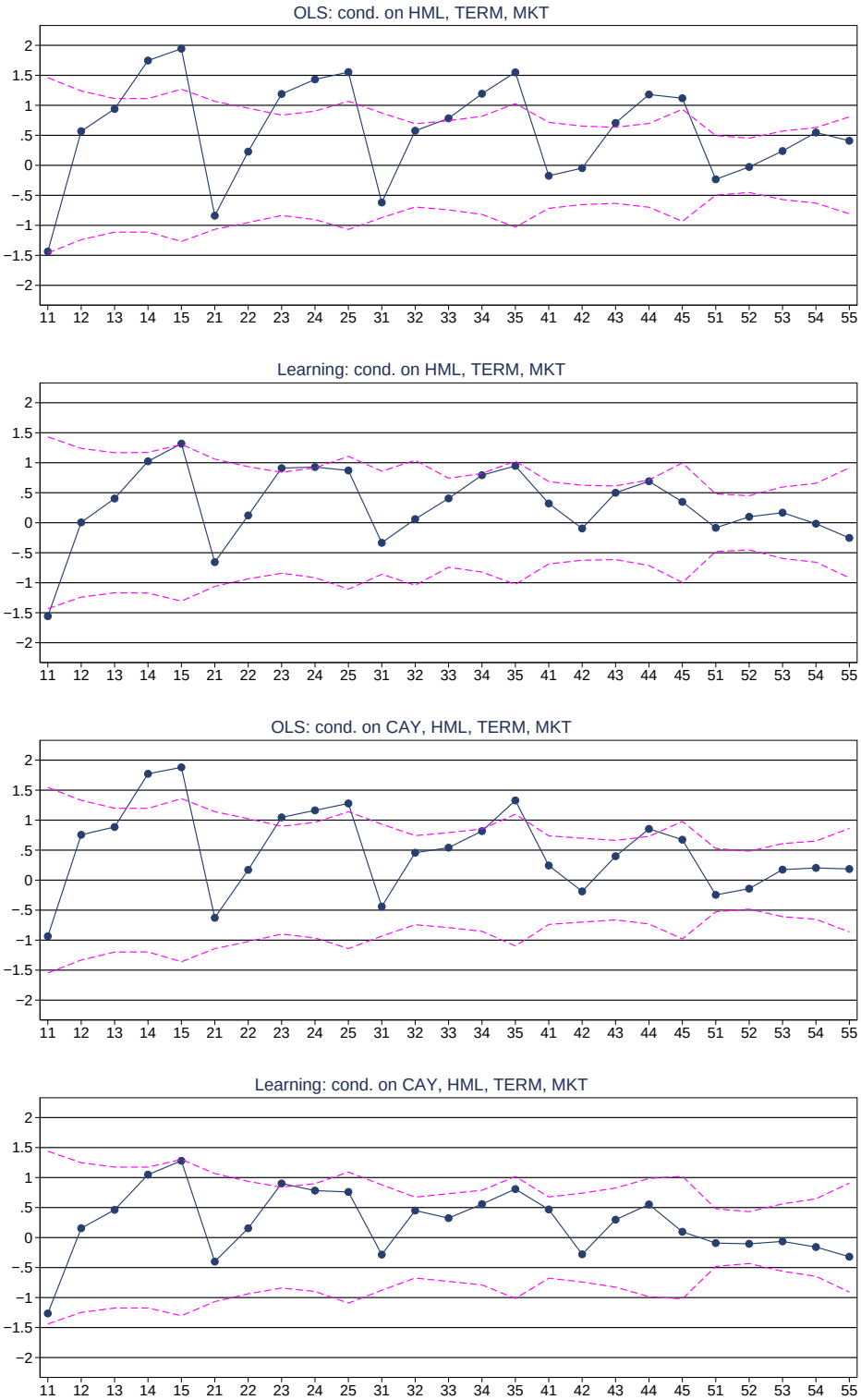
Figure 2: (continued)