Market Beta Dynamics and Portfolio Efficiency

Eric Ghysels
University of North Carolina, Department of Economics, Chapel Hill, NC 27599 and Department of Finance Kenan-Flagler Business School, eghysels@unc.edu, www.unc.edu/ eghysels

Eric Jacquier
Finance Department, HEC Montréal, CIREQ, and CIRANO, eric.jacquier@hec.ca, www.hec.ca/pages/eric.jacquier

This paper introduces a new estimation for the dynamics of betas. It combines two previously separate approaches in the literature, data-driven filters and parametric methods. Namely, we show how to estimate the parametric beta dynamics by instrumental variables combined with block-sampling - but not overlapping window filters - of data-driven betas. Instrumental variables are needed because of the measurement errors in empirical betas. We find that, while betas are very strongly autocorrelated, neither aggregate nor firm-specific variables explain much of their quarterly variation. We then compare block-samplers and overlapping window filters using a criterion of economic significance. Namely, we track the out-of-sample performance of portfolios optimized subject to target beta constraints. For target betas of zero, the case of many hedge funds, we show that estimation error results in systematic overshooting of the target beta. These portfolios benefit from the use of medium to long term estimation windows of daily returns.

Key words: Betas, Systematic Risk, Instrumental Variable Estimators, Asset Pricing

History: First Draft: June 2002. This version August 2006.

1. Introduction

The empirical limitations of the CAPM have directed research toward models of conditional expected excess returns where expectations are conditional on the investor's, intrinsically hidden, time varying information set, see e.g., Merton (1973), Jagannathan and Wang (1996), Campbell (1996). This makes the models very hard to test as one needs proxies for the information set to capture properly the dynamics of factor premia and betas, see e.g., Cochrane (2005, p. 145). Expected returns typically involve products of a factor risk premium by the asset beta with respect to the factor, The empirical literature has so far focused more on conditional time varying factor premia than betas. However, time-varying risk premia, now well accepted, can not alone address the deficiencies of basic models, see for example Lewellen and Nagel (2006).

There is no consensus on the nature of the time variation of betas. Optimal investment in Berk et al. (1999) links predictability in betas to changes in growth options. Conditional betas in Santos and Veronesi (2004) depend on firm characteristics and state variables driving the opportunity set. The structural links in these models are too complex for direct estimation. In Hamada (1972) and Rubinstein (1973), levered equity betas rise with financial leverage. But they can move in the opposite direction in even simple models of growth options, e.g., Jacquier et al. (2004). The empirical evidence is also ambiguous. Braun et al. (1996) incorporate time varying betas with leverage effects in a GARCH. They find little leverage effect or even time variation. Bekaert and Wu (2000) find weak evidence of leverage in betas. Jacquier et al. (2004) conclude that growth options effects dominate financial leverage, but their cross-sectional results characterize more the long-run variation in betas than those over shorter horizons below a year.

This lack of clear characterization of the time variation of betas is unfortunate because of mounting evidence that mis-specified beta dynamics can seriously affect asset pricing tests, e.g., Wang (2003). For portfolio optimization, large covariance matrices are often modeled with parsimonious factor structures, a strategy only as good as its imbedded forecasts of betas. There are two main empirical approaches to modeling the dynamics of conditional betas. One can use purely data-driven filters such as the rolling sample estimates in Fama and MacBeth (1973) and many after. Or, one can impose parametric relationships between betas and proxies for the state of the economy, as Shanken (1990), Harvey (1991), and Ferson and Harvey (1993, 1999). This second approach has been argued to be more powerful as it ties directly to theory, but it is vulnerable to mis-specifications. Indeed, Ghysels (1998) shows that several well-known parametric time-varying beta models are so badly mis-specified that constant beta models outperform them.

This paper combines the robustness of data-driven filters with the power of parametric methods to estimate β dynamics. Namely, we estimate nesting models $\beta_t = \alpha\beta_{t-1} + \gamma X_{t-1} + \upsilon_t$, where past beta and economic variables determine the current beta.[1] Naive estimation suffers from biases due

[1] The analysis can extend to non-linear specifications, we concentrate here on linear dynamics.

to errors in the variables as one must use an estimate of the true β_{t-1}. We design an instrumental variables (IV) estimator of α, γ, the dynamics of the true unobserved β.[2] This allows us to incorporate into the robust data-driven filters the added precision of the parametric approach, without its potential mis-specification or errors in the variables. We show that not using the IV estimator leads to large downward biases in the estimate of the persistence of β_t. Properties of IV estimators are only asymptotic, so we simulate the small-sample behavior of our IV estimator. We show that it has good small sample properties, capturing well the true beta dynamics.

With this new estimator, we can reassess the empirical specification of beta dynamics, here at quarterly horizons. First, we find that true betas exhibit far higher autocorrelation than previously estimated. Indeed, one would expect the β's of large firms to change slowly as projects are added or concluded. Economic theory also implies that betas should vary slowly. For example, in Gomes et al. (2003) betas are linked to productivity shocks, themselves often calibrated with quarterly autocorrelations above 0.9. Counter to intuition and theory, the empirical evidence was that betas had low persistence. Our IV estimator corrects for this, reconciling evidence and theory. Second, and less encouraging, we find that neither aggregate nor firm-specific variables explain much of the time-variation of quarterly betas. This is in contrast with theoretical models that link time variation in betas to economic variables, e.g., Gomes et al.(2003) and Santos and Veronesi (2004), and parameterizations that link beta dynamics to macro and firm-specific variables as Ferson and Harvey (1999) and Petkova and Zhang (2005).[3]

As true betas are latent, comparing the usefulness of models for beta dynamics is not easy. Hedge funds often need to neutralize factor betas. Pension funds optimize portfolios subject to target betas. They all need good forecasts of betas. To evaluate the economic value of competing out-of-sample forecasts of β, we build optimized portfolios subject to target betas, using these forecasts.

[2] An alternate, computationally intensive approach would be a Markov Chain Monte Carlo algorithm to simulate the latent β_t. It requires additional distributional assumptions, with their associated potential mis-specifications. See Jostova and Philipov (2005) and Ang et al. (2006).

[3] This also relates to Lewellen and Nagel (2003) who model variations in betas with data-driven estimators. But the objectives differ. They want to test the conditional CAPM without having to specify conditioning information. So they do not link betas to the economy.

We document the performance of these portfolios, especially their realized betas.[4] We find that filters with short historical windows produce the worst portfolio performance. Also, of daily returns always result in performance superior to the more often used monthly returns. Finally, funds that set target betas far from the average beta of one are most vulnerable to estimation error.

The paper is organized as follows. Section 2 reviews the theoretical motivations of beta dynamics. Section 3 introduces the new instrumental variable estimator. Section 4 reports the empirical results. Section 5 ranks filters via their out-of-sample portfolio performance. Section 6 concludes.

2. Conditional Beta Dynamics: Theoretical Implications

Modern asset pricing models have implication on the dynamics of conditional betas. Typically, the stochastic discount factor is a deterministic function of latent state variables Y_t, and factors F_t. In turn, the conditional expected excess return of asset i is written as a linear projection:

$$E_t[r_{t+1}^i|Y_t] = \sum_{j=1}^{F} \beta_{jt}^i(Y_t)\lambda_{jt}(Y_t),$$

where λ_{jt} is the premium associated with risk factor j.[5] The β_t's are then *deterministic* functions of the state variable, by virtue of their definition. Namely for factor k and the vector of state variables Y_t, $\beta_{k,t}^i = Cov[r_{t+1}^i, F_{k,t+1}|Y_t]/Var[F_{k,t+1}|Y_t]$.

Purely data-driven forecasts of β's, even somewhat parameterized as in GARCH, do not exploit the implications of asset pricing models on the dynamics of betas. So, they may lack precision. See, .e.g, the lack of sharp results in Braun et al. (1996) and Bekaert and Wu (1999). On the other hand, strict parameterizations of betas as deterministic functions of state variables are subject to model misspecification, as shown in Ghysels (1998).

There are many state variable specifications. For example, in Brennan et al. (2004) and Brandt and Chapman (2005), betas are functions of random state variables driving the investment opportunity set. In Campbell and Cochrane (1999), the surplus consumption ratio with respect to an

[4] Minimizing variance avoids confusions due to the notoriously difficult issue of mean forecast. DeMiguel et al.(2004) find that simple strategies that ignore changes in the mean do surprisingly well.

[5] A starting point is generally the fundamental pricing equation as in Harrison and Kreps (1979). See for example Assumption 3.4 and Theorem 2 in Garcia and Renault (2001). See also the technical supplement to this paper for a more detailed discussion.

external habit level is a state variable. Bansal and Yaron (2005) use predictable components of consumption and dividend growth as well as consumption volatility. For all these variations of the state variables, factors, and their dynamics, the betas inherit their dynamics from their complex deterministic link with the state variables.

While the above models are in discrete time, continuous-time models yield similar results for the dynamics of betas. The latter are possibly more relevant to our empirical approach, as one studies the properties of filters through their limit to continuous time. Consider for example Santos and Veronesi (2004), hereafter SV. In their economy, conditional betas are a complex, deterministic function of both aggregate and firm-specific state variables. As the state variables have diffusive dynamics, so does β by Itô's Lemma. That is:

$$d\beta_t^i = \mu_\beta^i(t)dt + \sigma_\beta^i(t)dB_t. \tag{1}$$

The drift and volatility functions $\mu_\beta^i(t)$ and $\sigma_\beta^i(t)$ follow from the functionals in SV and the dynamics of the state variables, but the mapping is too complex to be practically useful. See SV for the difficulties of structural estimation of β in an equilibrium model.

We can however incorporate some implications of models such as SV into a simpler reduced form approach. Starting from an Euler discretization of equation (1):

$$\beta_t^i - \beta_{t-1}^i = \mu_\beta^i(t-1) + \sigma_\beta^i(t-1)\varepsilon_t^i,$$

we will consider linear projections of β_t^i - β_{t-1}^i onto past betas, i.e. β_{t-1}^i and the state variables, taking into account the fact that neither state variables nor β's are observable.

This reduced form approach is more than just a work around the complex mapping in (1). Unlike the strict parametric approach, it is robust to mis-specifications of the underlying equilibrium model, and to the additional error introduced by the use of proxies in lieu of the unobservable state variables. Denote X_{t-1}^i the vector of these proxies, the linear projection becomes

$$E_{t-1}[\beta_t^i - \beta_{t-1}^i | \beta_{t-1}^i, X_{t-1}^i] \equiv E_{t-1}[\mu_\beta^i(t-1) | \beta_{t-1}^i, X_{t-1}^i] \equiv (\alpha_i - 1)\beta_{t-1}^i + \gamma_i X_{t-1}^i.$$

It can be viewed as a linear dynamic approximation to the theoretical specifications of betas. Consequently, we focus on linear regression models:

$$\beta_t^i - \beta_{t-1}^i = (\alpha_i - 1)\beta_{t-1}^i + \gamma X_{t-1}^i + \upsilon_t^i. \tag{2}$$

When $\alpha_i = 1$, the proxies completely capture the dynamics of betas. Equation (1) suggests this could be the case if 1) the model was correct, 2) the linear projection was a good enough approximation of the structural function, and 3) we could observe the state variables without error. So, β_{t-1}^i on the right-hand side of (2) allows for persistence not captured by the proxies. Alternatively, when $\gamma_i = 0$, the proxies do not capture any beta dynamics.

3. Estimating beta dynamics: a new approach

To estimate (2), one needs to substitute a data-driven filter for β. The true β_t is replaced with $\beta_t + \epsilon_t^J$, where ϵ_t^J is the estimation error of a candidate filter J for β. Regression (2) then becomes:

$$\beta_t - \beta_{t-1} = (\alpha - 1)(\beta_{t-1} + \varepsilon_{t-1}^J) + \gamma X_{t-1} + (\upsilon_t + \varepsilon_{t-1}^J - \varepsilon_t^J),$$

where the explanatory variable is correlated with the error. This results in, typically downward, biased estimates of α the parameter of the beta dynamics. The bias can be sizeable. An example of this bias may be found in Hsieh (1991) who estimates daily standard deviations from intra-day IBM returns, a block-sampling estimator. He then fits an AR(5) on the logarithm of the daily standard deviation estimate. The five coefficients sum up to 0.77, well below the typical values above 0.95 found by GARCH or stochastic volatility likelihood-based estimation.

The classic approach to eliminate the bias requires instrumental variables orthogonal to ϵ_{t-1}^J. The orthogonality guarantees a consistent estimation of α. The instruments must also be highly correlated with β_{t-1} itself for the procedure to have some power. So, we need instruments orthogonal to the measurement error in β_{t-1} and highly correlated with the true β_{t-1}? We will see that we can indeed find such instruments if we use block-sampling filters for β, not so if we use overlapping rolling-window estimators. We will show how to exploit the properties of block-sampling filters, to design an instrumental variables estimator of beta dynamics, i.e., of the parameters in (2). We now review the asymptotic properties of overlapping and block-sampling estimators

3.1. Properties of data-driven filters

Data-driven estimators specify a sampling frequency for the underlying returns, a data window length and a weighting scheme. Various combinations of sampling frequencies and window length can produce rolling-sample estimators where adjacent estimates of beta use overlapping blocks of data. or block-sampling where adjacent estimates contain no overlapping returns. There are no clear rules on the choice of sampling frequency and the number of lags, although many still use overlapping windows of 60 monthly returns as Fama and MacBeth.

Two types of asymptotic arguments can be used to model how discretely sampled returns approximate a continuous time process. First, as in Foster and Nelson (1996), we center on rolling sample estimators. Second, we discuss the more recent development in Barndorff-Nielsen and Shephard (2004), who concentrate on non-overlapping block-sampling estimators. Typically, the analysis initiates from a continuous time diffusion process for the vector of stock prices:

$$dS_t = S_t \mu_t dt + S_t \Omega_t^{1/2} dB_t^S, \tag{3}$$

where the drift and volatility are characterized as measurable functions of the filtration generated by the price process S_t.[6] Ω_t contain the conditional variance and covariances necessary for the computation of conditional betas. Here, the first element of S_t is the market portfolio. So, $\beta_{i,t} = \Omega_{(1i),t}/\Omega_{(ii),t}$. The asymptotic arguments then relate to the estimation of Ω_t, looking at limits to continuous time from discretely sampled data.

3.1.1. Overlapping rolling-sample estimators

In Foster and Nelson (1996), one samples data at discrete fixed interval h and denotes $S_\tau^{(h)}$ for $\tau = 0, h, 2h, \ldots$ the discretely sampled realizations S_t. Given, regularity conditions they show that the shock to the discretized process $\Delta S_\tau^{(h)} \equiv S_\tau^{(h)} - S_{\tau-h}^{(h)}$ is a local martingale difference with an almost surely finite conditional covariance $h\Omega_\tau^{(h)}$, i.e., $\Omega_\tau^{(h)}$ is measurable. Then they discuss rolling sample estimators of Ω, denoted $\hat{\Omega}_{(ij)\tau}^{(h)}$, that are asymptotically normally distributed. Such an estimator with l lags can be written as:

$$\hat{\Omega}_{(ij)\tau}^{(h)} = \sum_{t=\tau-lh}^{\tau} w_{(ij)\tau}^{(h)} [\Delta S_{it}^{(h)} - h\hat{\mu}_{it}^{(h)}][\Delta S_{jt}^{(h)} - h\hat{\mu}_{jt}^{(h)}], \tag{4}$$

[6] See our technical supplement and Nelson and Foster (1996) for further details on the steps in this section.

where $\hat{\mu}^{(h)}$ estimates $\mu^{(h)}$, typically also using a rolling sample. For an equal-weighting scheme, $w^{(h)}_{(ij)t} = [lh]^{-1}$. Foster and Nelson (1996) show that, symptotically, $h^{-1/4}[\hat{\beta}^{(h)}_{i\tau} - \beta^{(h)}_{i\tau}]$ converges in distribution to a standard normal with mean zero and some covariance matrix. That covariance is a function of nuisance parameters involving higher moments of the underlying process S_t.

The asymptotic properties of the above estimator very much depends on the choice of the weighting scheme. In turn the theoretically optimal, in mean squared error, weighting schemes given by Foster and Nelson also rely on the above nuisance parameters and are not easily applied.[7] Here we forego this "theoretical" optimality and use equal weights for the purpose of comparing data-driven estimators.

3.1.2. Non-overlapping block-sampling estimators The other class of estimators often used is the block-sampling estimator. For example, Merton (1980) and French et al. (1987) use monthly windows of daily returns to compute monthly variances. Hsieh (1991) uses daily windows of intraday returns to compute daily variances. Note that the asymptotics in Foster and Nelson is about the estimation of Ω_t, the instantaneous or "spot" covariance matrix. The focus in Barndorff-Nielsen and Shephard (2004), hereafter BNS, is slightly different, and so are their resulting estimators and their asymptotic properties. Instead of the spot covariance matrix, they analyze the ability of block-sampling estimators to estimate $\Lambda_{[t-1,t]} \equiv \int_{t-1}^{t} \Omega_\tau d\tau$, denoted the 'integrated' covariance matrix. They specifically consider **non**-overlapping blocks of data, and compute over the interval $[t-1, t]$:

$$\tilde{\Lambda}^h_{(ij)[t-1,t]} = \sum_{k=1}^{1/h} [\Delta_h S_{i(t+kh)} - h\hat{\mu}^{(h)}_{i[t-1,t]}][\Delta_h S_{j(t+kh)} - h\hat{\mu}^{(h)}_{j[t-1,t]}]', \tag{5}$$

where $\hat{\mu}^{(h)}_{[t-1,t]}$ estimates $\int_{t-1}^{t} \mu_\tau$. The emphasis is on non-overlapping data blocks and estimators that sum inner products of the return vector $\Delta_h S$. We denote $\tilde{\beta}^h_{i[t-1,t]}$ the conditional betas based on the integrated covariance estimate $\tilde{\Lambda}^h_{[t-1,t]}$ in (5). In contrast, $\hat{\beta}^h_{it}$, based on the spot estimate $\hat{\Omega}^h_t$ from (4) correspond to 'spot' betas. There are distinctions. For example, the $\hat{\beta}$s based on (4) are

[7] They show for a large class of continuous path stochastic volatility models the optimal weighting scheme is exponentially declining and provide formulas for the optimal decay rate and lag length of various estimation procedures.

at conditional on the data window used in the rolling sample estimator.[8] In contrast the integrated conditional betas from (5) are conditional on the data sampled over the data block $[\tau-1,\tau]$.

There is an important difference between the asymptotic behaviors of betas based on (4) and (5). BNS consider a "fill-in" asymptotics, where one samples more frequently within the fixed block $[t-1,t]$. This is different from the continuous record asymptotic analysis of Foster and Nelson. BN-S, theorem 2, show that as $h\to\infty$:

$$\sqrt{h}\, vech(\tilde{\Lambda}^h_{(ij)[t-1,t]} - \Lambda_{t-1,t}) \to N(0,\Pi_t),$$

where the covariance matrix Π_t varies across sampling block. Hence the resulting asymptotic distribution is a mixture of normals. This is different from the continuous record asymptotics of Foster and Nelson, who obtain (1) a pure gaussian limit, and (2) a convergence rate of only $h^{1/4}$, not $\sqrt{h}$. Specifically, Π_t involves the quarticity, that is, the fourth moments of returns.

Under BNS, we may, with mild regularity conditions, further assume that Π_t is independent across non-overlapping blocks. Hence, the sampling error of estimated betas is a mixed Gaussian distribution independent across non-overlapping blocks of data. While the most general analysis in BNS does not result in independence of conditional fourth moments across non-overlapping blocks, this is a mild additional assumption with which we proceed to the estimation of beta dynamics.

3.2. IV estimation of β dynamics with the block-sampler

We therefore substitute the block-sampling estimator $\tilde{\beta}^h_{[t-1,t]}$ based on the covariance estimate (5), for β in (2). We, in effect, approximate beta with an integrated rather than a spot beta. Since empirical betas should be moving slowly through time , e.g., Braun et al. (1996), the difference between a spot and integrated beta are minor is not an issue here. This will be confirmed by our simulation results.[9]

Since the block-sampling method results in independent estimators of β_t from block to block, we will use as instruments for β_{t-1} in (2) simple lags of $\tilde{\beta}^h_{[t-1,t]}$. The instruments will be highly

[8] Strictly speaking, they are estimates of a spot beta process which is a measurable function of the entire past of stock returns. As distant pasts don't matter, rolling sample estimators do better with downward sloping weights.

[9] For intuition, in the extreme case of constant spot betas, the integrated equals the spot times the length of the block. All betas in the regressions are therefore simply re-scaled.

correlated with β_{t-1} through the persistence of the true β_t, yet uncorrelated with estimation error because of the independence of the estimator across blocks. This exploits the results of BNS as since in the previous section. In contrast, overlapping estimators in (2) would 1) introduce further autocorrelation in the regression error and 2) preclude the use of lagged estimates as instruments.[10]

4. Empirical Results on Beta Dynamics

After describing the data, we report on the autoregressive beta dynamics estimated by IV. A small-sample Monte Carlo study compares OLS and IV estimators.

4.1. The data

We compute the daily and monthly returns of 25 value-weighted industry portfolios formed on the basis of 2 to 3 digit SIC codes, from July 1962 to December 2004. Such portfolios are widely used in the empirical literature. We estimate β's with respect to the weighted CRSP index. We will illustrate a number of the many practical variations of the window length, the sampling frequency of underlying returns, and the frequency of re-computation of betas. We denote $\hat{\beta}_{x,y,t}$ the estimator, with calendar span x, sampling frequency of the returns y, and time of computation t. As the firm specific characteristics are only available quarterly, we recompute betas quarterly, i.e., t is a quarterly index.

Absent the firm-specific variables, we could re-estimate β more frequently, for example monthly. However the instrumental variables approach, while it can corrects for biases, can not change the inherent lack of power of some implementations. With only 22 daily returns per month, the resulting $\hat{beta}$ would be even more variable than $\hat{\beta}_{q,d}$, resulting in even more imprecise estimation for regression 2. This could in principle be addressed by the use of intra-daily returns, as in Andersen et al. (2005). However, intra-daily returns are available for relatively short data spans. Moreover, intra-daily data would increase the potential problem of non-synchronous trading in the estimation of β as in SW and of model error in the asymptotics, e.g., Bandi and Russell (200?) and Ait-Sahalia (200?). So, to estimate β, we will sample returns no more frequently than daily.

[10] One could, in principle, use a lag of the estimator long enough to eliminate correlation with the estimation error. This lag would in turn render the procedure useless for lack of power.

Table 1 Descriptive statistics on the 25 industry portfolios

We use daily returns from July 1962 to December 2004. "# stocks" shows the smallest and the average number of stocks in a portfolio over the period. The next two columns show the time-series average and the first-order (quarterly) autocorrelation of $\hat{\beta}_{q,d,t}$, the quarterly estimate of β from a quarterly window of daily returns. The last three columns show the absolute value of quarterly changes in β's, estimated from a quarterly window of daily returns (q,d), and five-year windows of daily (5y,d) and monthly returns (5y,m).

Industry	SIC code	# stocks		$\overline{\hat{\beta}_{q,d}}$	ρ	$\lvert\Delta\hat{\beta}_{q,d}\rvert$	$\lvert\Delta\hat{\beta}_{5y,d}\rvert$	$\lvert\Delta\hat{\beta}_{5y,m}\rvert$
Mining & Minerals	10 12 14	61	62	0.74	0.65	0.28	0.028	0.040
Oil and Gas	13 29 46	117	221	0.91	0.70	0.16	0.021	0.034
Food wholesale	1 2 7 9 20 21	129	149	0.83	0.78	0.12	0.015	0.023
Food retail	54	28	43	0.70	0.56	0.15	0.016	0.031
Construction	15-17 24 32 52	85	148	1.06	0.64	0.13	0.015	0.025
Textile	22 23 31	87	132	0.87	0.58	0.15	0.019	0.034
Basic (metal,paper)	26 33	139	136	1.01	0.56	0.13	0.013	0.019
Durable goods	25 50 55 57	55	180	1.07	0.63	0.14	0.017	0.027
Printing	27	23	74	0.81	0.50	0.12	0.012	0.023
Chemicals	28	129	249	1.02	0.58	0.10	0.010	0.018
Rubber & Plastic	30	32	57	0.92	0.52	0.16	0.018	0.031
Metal	34	80	117	0.88	0.42	0.12	0.015	0.018
Machines	35	142	309	1.25	0.39	0.20	0.016	0.025
Electronics	36	152	360	1.31	0.66	0.16	0.015	0.021
Transport	37 40-42 44 45 47	192	218	1.13	0.60	0.13	0.012	0.022
Instruments	38	47	221	1.22	0.56	0.16	0.015	0.026
Misc. Manufacture	39	24	61	0.97	0.58	0.20	0.021	0.036
Communication	48	20	93	0.87	0.67	0.18	0.016	0.025
Utilities	49	117	186	0.56	0.53	0.13	0.014	0.021
Trade	51 53 56 59 72	93	248	1.03	0.56	0.14	0.015	0.027
Entertainment	58 78 79	41	132	1.17	0.60	0.20	0.018	0.031
Finance	60-64	45	468	0.97	0.71	0.12	0.016	0.026
Real Estate	65	33	58	0.83	0.78	0.20	0.028	0.042
Holdings Companies	67	59	554	0.84	0.73	0.08	0.011	0.016
Services	70 73 76 80-83 86 87 89	25	510	1.22	0.59	0.19	0.019	0.027
Average		78	199	0.97	0.60	0.15	0.017	0.027

Table 1 shows the composition of the portfolios and some descriptive statistics. The column $\overline{\hat{\beta}_{q,d}}$ reports the average of $\hat{\beta}_{q,d,t}$ for the period. It shows that, although the industry portfolios do not explicitly sort on β, they do exhibit a good range of average betas, from 0.56 for utilities to 1.31 for the electronics industry.

To proxy for firm-specific state variables as in SV and others, we use the market value of equity (MV), the ratio of debt to equity (DTE), and the ratio of book to market value (BTM). We obtain quarterly values from COMPUSTAT from 1970Q4, until 2004Q1. These variables are used extensively in the literature. DTE measures financial leverage, an increase of which should increase the levered equity beta, see Hamada (1972) and Rubinstein (1973). BTM is similar to the reciprocal of the ratio of book value of assets to total firm value (A/V), see Smith and Watts (1992). It is a measure of operating leverage, linked to past and possibly future performance, i.e., a proxy for growth opportunities. Growth opportunities have higher betas than assets in place. If large negative returns are linked to the loss of growth opportunities, the un-levered equity beta should decline after large losses. So, variations in BTM could track variations in β, see e.g., Berk et al. (1999) and Jacquier et al. (2004).

We also collect monthly values, until december 2004 of five series commonly used to proxy for the global state variables. The level of interest rates is represented by the lagged value of the 30 day US Treasury Bill return. The term premium is the difference between US long and intermediate term yields. The default premium is the difference between the Moody Baa and the US long term yields. The dividend yield is the average $S\&P500$ income return over the past 12 months, lagged by one month. The market return is the $S\&P500$ total return lagged by one month. The government and S&P500 series are extracted from the Ibbotson database. We compute quarterly values for these five quantities. For the bonds and the dividend yield, we use the latest monthly value as of the start of the quarter. For the market, we use the quarterly return lagged by one quarter.

4.2. Description of the β dynamics

Before proceeding with the formal estimation of equation (2), we document the salient differences between major competing block-sampling and rolling estimators. Given our quarterly re-estimation frequency, the only estimator to which the IV procedure can be applied is $\hat{\beta}_{q,d,t}$, the block-sampling estimator which uses non-overlapping quarterly windows of daily returns. In contrast with this short window of daily returns, the rolling estimator most commonly used in the finance literature is $\hat{\beta}_{5y,m,t}$, based upon 5 years of monthly returns. Re-estimated quarterly, it does not qualify as a block-sampling estimator. We also introduce $\hat{\beta}_{5y,d,t}$, which should be more precise according to both asymptotics presented before. It is however rarely used. The last three columns in Table 1 show the average of $|\beta_t - \beta_{t-1}|$, for these estimators. The block-sampling estimator $\hat{\beta}_{q,d,t}$ has the highest variability, 0.15 on average, from 0.08 for the Holding companies to 0.28 for Mining and Minerals. The variability of $\hat{\beta}_{5y,d,t}$ is about a tenth of that, 0.017 on average. The variability of the commonly used $\hat{\beta}_{5y,m,t}$ averages 0.027. Although much lower than for the quarterly block-sampler, it is still 60% higher than for its daily-returns based counterpart.[11]

Figure 1 show how very different these estimators are. First consider the effect of window length: The left-side plots show the quarterly estimates of the estimators based upon daily returns, for three industries. The quarterly window $\hat{\beta}_{q,d}$ displays enormous variability. The five-year window $\hat{\beta}_{5y,d}$ removes all the high frequency variability but displays a strong lag, a known and undesirable feature of one-sided filters. While a lot of the variability of $\hat{\beta}_{q,d}$ may owe to estimation error, $\hat{\beta}_{5y,d}$ may over-smoothe and delay a lot of information on the true beta. A 5 year window could be a very costly way to reduce the estimation error in $\hat{\beta}_{q,d}$. As a compromise between these extreme window lengths, we introduce $\hat{\beta}_{13m,d}$. With asymptotic arguments as in Foster and Nelson (1996), one can show that $\hat{\beta}_{13m,d}$ and $\hat{\beta}_{5y,m}$ have similar asymptotic MSE, see Andreou and Ghysels (2006). Indeed, this compromise candidate $\hat{\beta}_{13m,d}$, dashed line in the left plots, has a far smaller lag than the five-year window estimator. It seems to preserve the core time-variation in $\hat{\beta}_{q,d}$, while removing the high-frequency shocks likely due to estimation noise.

Consider now the effect of sampling frequency. The right plots in Figure 1 show $\hat{\beta}_{5y,m}$ - solid line, and $\hat{\beta}_{5y,d}$ - dotted line, for the same industries and on the same scale as the left plots. Their time-series patterns differ to a surprising extent. The two betas are often apart sometimes by large amounts, as for Oil and Gas, and Durable Goods. So, in addition to a difference in variability, seen in Table 1, these two sampling frequencies produce estimates far apart for long periods.

Although daily returns have been available since 1962 and $\hat{\beta}_{5y,d}$ is preferable to $\hat{\beta}_{5y,m}$ by all asymptotic arguments, it is still hardly used. This may be because of the perceived effects of non-trading on the estimation, discussed in Scholes and Williams (1977). We computed a Scholes-Williams, hereafter SW, version of $\hat{\beta}_{5y,d}$ with a five-day window. Namely, one regresses the portfolio

[11] One may worry that the industry portfolios do not produce enough time-series variability in β, MV, DTE, and BTM. So we also formed 27 portfolios sorted annually along three levels of MV, DTE, and BTM. The results were similar and can be found in the technical appendix to the paper.

returns on two lags, two leads, and the contemporaneous market returns. The SW beta is the sum of these five coefficients. It is the dashed lines on the right plots of Figure 1. One sees that it vindicates neither of the other two estimators. It sometimes agrees with $\beta_{5y,m}$, sometimes with $\beta_{5y,d}$, sometimes traces a path of its own. It is indeed more often than not above the original $\hat{\beta}_{5y,d}$, perhaps correcting a downward bias. However, while the SW beta may reduce a bias due to non-trading, it is variable since it requires more regressors than the standard market model. The SW estimator seems to offer no clear benefit.

4.3. Estimation of the β dynamics

We now estimate the dynamics of β using the quarterly block-sampling, aka integrated beta: $\hat{\beta}_{q,d,t}$. Namely we implement the regression

$$\hat{\beta}_{q,d,t} - \hat{\beta}_{q,d,t-1} = (\alpha - 1)\hat{\beta}_{q,d,t-1} + \gamma X_t + \upsilon_t, \tag{6}$$

by instrumental variables.

The empirical analysis focuses on three specifications of regression (6). First, *Model 1* is a pure AR(1) where X_t contains only an intercept. Then we consider models, where $\alpha = 1$, to examine the time series relation between the time variation of $\hat{\beta}_{q,d}$ and the firm-specific and macroeconomic variables. In, Model 2 the X_t contains only firm specific variables. Model 3 represents the case where X_t contains only the economy-wide variables. We also consider the nesting combination of Models 2 and 3, and of Models 1, 2, and 3.

The instrumental variables are the second lag of beta, and the X variables themselves. The characteristics, BTM, DTE, and Log(MV) are scaled by their time series mean and variance, so γ represents the change in β related to a one-standard deviation change in the variable. Table 2 reports estimates of the above models for the 25 industry portfolios. The industry portfolios are not sorted by DTE, BTM and MV. On may worry that potential high correlations between the three characteristics may affect the precisions of estimation. We computed these correlations. The averages of the 25 correlations is 0.71 between BTM and DTE, -0.67 between BTM and MV, -0.45 between DTE and BTM, for the 27 sorted portfolios.

Table 2 IV estimation of β dynamics for 25 industries

The table reports the time series IV estimation β dynamics as in equation (6): $\hat{\beta}_{q,d,t} = \alpha\hat{\beta}_{q,d,t-1} + \gamma X_{t-1} + \upsilon_t$. Model 1 is a pure AR(1) where γX_{t-1} is an intercept. Otherwise $\alpha = 1$, and X_{t-1} is only firm characteristics (Model 2) or only macro variables (Model 3). We also estimate the nesting models 2+3 and 1+2+3. The instrumental variables are the second lag of beta and the X variables. The R-squares are adjusted for degrees of freedom. The characteristics, BTM, DTE, and Log(MV) are scaled by their time series mean and variance. For each explanatory variable, the two rows report the 15^{th}, 50^{th}, 85^{th} percentiles of the 25 estimates (top row) and IV t-statistics (bottom row). The 85^{th} percentile of the z score is 1.04. The calendar span is from 1970 to 2004, when firm characteristics are available.

	Model 1			Model 2+3			Model 1+2+3		
$\hat{\beta}_{qd,t-1}$	0.81	0.93	0.98				0.67	0.79	0.96
	7	8	11				3.0	4.3	5.5
		Model 2							
BTM	-0.03	-0.01	0.01	-0.04	0.00	0.02	-0.06	-0.02	0.01
	-0.9	-0.2	0.3	-1.1	-0.1	0.4	-1.1	-0.4	0.4
DTE	-0.01	0.1	0.02	-0.02	0.01	0.05	-0.02	0.01	0.05
	-0.4	0.2	0.08	-0.4	0.4	1.4	-0.5	0.3	1.1
Log(MV)	-0.01	0.00	0.02	0.00	0.02	0.04	-0.01	-0.01	0.04
	-0.2	0.0	0.7	-1.1	0.5	1.1	-0.3	0.3	1.2
$\bar{R}^2$	0.0	0.0	0.01						
	Model 3								
Tbill	-0.90	-0.41	0.15	-1.52	-0.47	0.11	-1.44	-0.62	0.34
	-0.7	-0.4	0.1	-1.4	-0.3	0.1	-1.4	-0.4	0.3
TP	-4.32	-2.63	0.79	-6.24	-3.63	0.62	-5.59	-2.77	3.40
	-1.0	-0.5	0.2	-1.5	-0.7	0.1	-1.4	-0.6	0.6
DP	-0.94	0.87	2.86	-0.77	1.03	3.01	-1.60	1.12	3.84
	-0.2	0.3	1.0	-0.2	0.4	1.1	-0.8	0.3	1.1
DY	-0.77	-0.24	0.45	-1.46	0.40	3.48	-0.84	1.42	3.81
	-0.4	-0.1	0.3	-0.4	0.2	1.1	-0.3	0.4	1.2
$R_{m,t-1}$	-0.22	0.04	0.22	-0.26	-0.01	0.25	-0.28	-0.04	0.18
	-1.3	0.2	1.0	-1.3	-0.1	1.1	-1.5	-0.2	0.8
$\bar{R}^2$	.01	.01	.02	.01	.03	.04	.03	.08	.11

The results for Model 1 are in Table 2. The magnitude of the lagged β coefficient is the same, on average 0.93. Again none of the information variables can predict the time variation of β. In fact, a large number of the t-values for the 15^{th} and 85^{th} percentiles have magnitude near 1.04, the 15% or 85% cutoff point of a z score.

The last three columns show that, even collectively, the information variables do not explain more than about 10% of the lagged coefficient α. The latter falls to an average of 0.79 when the variables are added to the estimation.

The results for the 27 sorted portfolios are in Table 2. For each explanatory variable, a first row reports the 15^{th}, 50^{th}, and 85^{th} percentiles of the 25 point estimates, and just below, of the 25 t-statistics. We also report the adjusted R-squares. The first three columns contain the results for each model estimated alone. The results for Model 1, show that the median autocorrelation of $\hat{\beta}_{q,d}$ is 0.93, with 70% above 0.81. This supports the view that β is time varying and highly persistent. Below are the estimates of γ for Models 2. None of the firm characteristics explain the time-series variability of β. The R-squares of Model 2 are close to zero. Model 3 is hardly more successful than Model 2, with no significant coefficients and essentially zero R-squares. There is no evidence of significance of any macro variable. Yet, there seems to be a systematic relationship for two economic variables. The Term and Default Premium coefficients are very often negative and positive respectively.

The three central columns report the results on the estimation of Models 2 and 3 together. As expected, given the individual results, their R-squares are still extremely low. Again none of the variables are significant. The complete model, 1+2+3, is in the last three columns. The R-squares jump to 12% on average due to the introduction of the lagged β. The estimate of α is still very high, on average 0.73, even in the presence of eight additional variables. The information variables explain away little of the lagged value of β, about only 20%.

4.4. IV versus OLS: Rejoinder

At this stage one may wonder whether the classical OLS estimation would yield very different results from the IV procedure proposed in this paper. We now show that not correcting for measurement errors in estimates of beta leads to very misleading and biased results regarding the dynamics. Hence, the use of the IV estimator is crucial. Table 3 reports the OLS and the IV estimation of α for Models 1 and Models 1+2+3.

The differences are very large. For Model 1, parameter estimates go from an average across industries of .59 for OLS to .90 for IV. For Model 3, the average change is from .40 to .80. We do not report here on the other parameter estimates (γ for the economic variables), as they don't differ much across the two estimators. The IV procedure, which corrects for the errors in measuring

Table 3 OLS versus IV Estimation of β Dynamics

The table reports the estimates $\hat{\alpha}_{OLS}$ and $\hat{\alpha}_{IV}$, of the regression: $\hat{\beta}_{q,d,t} - \hat{\beta}_{q,d,t-1} = (\alpha - 1)\hat{\beta}_{q,d,t-1} + \gamma X_{t-1} + \upsilon_t$. Model 1 is an AR(1) where γX_{t-1} is an intercept. Model 1+2+3 is the nesting model with all the proxy variables.

	Model 1		Model 1+2+3	
Industry	$\hat{\alpha}_{OLS}$	$\hat{\alpha}_{IV}$	$\hat{\alpha}_{OLS}$	$\hat{\alpha}_{IV}$
Mining & Minerals	0.66	0.93	0.55	0.92
Oil and Gas	0.70	0.90	0.49	0.82
Food wholesale	0.78	0.93	0.54	0.74
Food retail	0.60	0.87	0.52	0.78
Construction	0.65	0.78	0.45	0.53
Textile	0.58	0.94	0.44	1.00
Basic (metal,paper)	0.58	0.83	0.41	0.79
Durable goods	0.64	0.79	0.58	0.65
Printing	0.49	0.64	0.48	0.68
Chemicals	0.58	0.83	0.45	0.70
Rubber & Plastic	0.51	0.85	0.42	0.82
Metal	0.42	0.93	0.30	0.76
Machines	0.38	0.78	0.30	0.53
Electronics	0.66	0.96	0.37	0.94
Transport	0.60	0.87	0.37	0.65
Instruments	0.56	1.04	0.45	0.79
Misc. Manufacture	0.57	0.93	0.56	0.79
Communication	0.67	1.00	0.37	0.73
Utilities	0.53	0.93	0.43	1.00
Trade	0.57	0.95	0.42	0.90
Entertainment	0.60	1.02	0.42	1.05
Finance	0.71	0.90	0.51	0.82
Real Estate	0.78	0.98	0.63	0.97
Holding Companies	0.73	0.98	0.47	0.86
Services	0.59	0.87	0.38	0.77
Average	0.61	0.90	0.45	0.80

beta, shows that betas are very much persistent, far more as would be suggested by the biased OLS estimates. The downward bias is not surprising, as typically measurement errors tend to reduce persistence measures.

There is discussion in the econometric literature about the small sample properties of the IV estimators. We address this issue with a simulation experiment. This will allow us to assess the behavior of the IV estimator for the relevant sample sizes. Consider that the quarterly $\hat{\beta}$ estimates the true β with an error ϵ_t.[12] That is:

$$\hat{\beta}_t = \beta_t + \epsilon_t, \quad \epsilon_t \sim (0, \sigma_z).$$

[12] As noted in section 3.1.2, the asymptotic distribution of the measurement error is a mixture of Gaussian random variables. We use a normal error for simplicity in this Monte Carlo experiment.

We then specify alternate processes for the quarterly true β. In the first specification:

$$\beta_t = 0.05 + 0.95\beta_{t-1} + v_t.$$

This models a strong autocorrelation and an unconditional mean of $0.05/(1-0.95) = 1$. We only need to calibrate reasonable magnitudes for σ_v and σ_ϵ. For σ_ϵ^2, we use the average variance of $\hat{\beta}_{OLS}$ for a regression with a quarter of daily returns, about 0.011. One can argue that this may be a lower bound on the variance of the estimation error, so our results a conservative with respect to the effect of errors in the variables. On the other hand, the realized variance of $\hat{\beta}_{q,d,q}$ is on average 0.048 for the sorted portfolios. So, we use $0.048 - 0.011 = 0.037$ as the variance of the true β_t, which in turn implies $\sigma_v = 0.06$.

We simulate 3000 dynamic regressions estimating α, 0.95 here, by OLS and IV, with sample size 170. The results are in Panel a of Table 4. The IV and the OLS estimates are on average very close to what we reported for the actual data in Table 3. The IV shows nearly no bias with a sampling mean of 0.93 while the OLS is severely biased downward with a mean of 0.65. The IV RMSE is also drastically lower than the OLS. To get an idea of the amount of the correction needed, consider testing at the 5% level of significance, the null hypothesis that $\alpha = 0.95$. The OLS-based test incorrectly rejects the null 99% of the time! The IV-based test rejects the null about 3% of the time. Some of the time the IV-estimator may produce an estimate of α larger than 1. So, we also show results on the truncated IV estimator, $\hat{\alpha}_{IV2} = Min(\hat{\alpha}_{IV}, 0.99)$.

We also consider the case where the quarterly β_t is driven by a persistent macro-onomic variable x_t. In addition to documenting the effectiveness of the the IV estimator, we want to make sure that, in small sample, the coefficient of the economic variable is correctly estimated. Namely, consider:

$$\begin{aligned} \beta_t &= 0.05 + 0.95\beta_{t-1} + x_t + v_t \\ x_t &= 0.95\delta x_{t-1} + a_t. \\ \hat{\beta}_t &= \beta_t + \epsilon_t \end{aligned} \tag{7}$$

Again, we set the mean of β equal to 1, as well as a strong autocorrelation for the economic variable. The true β is observed with error, $\sigma_\epsilon^2 = 0.011$ as above. v_t can be seen to reflect the fact

Table 4 Sampling Properties of OLS and IV Estimators of β Dynamics

The table reports the sampling behavior of OLS and IV estimators of the regressions:

$$\hat{\beta}_t = \alpha_0 + \alpha_1 \hat{\beta}_{t-1} + \gamma x_t + \upsilon_t, \quad t = 1, \ldots, 170,$$

where, $\alpha_0 = 0.05$ and $\alpha_1 = 0.95$. $\hat{\beta}_t = \beta_t + \epsilon_t$, where $\sigma^2(\hat{\beta}) = 0.048$, and the variance of the estimation error is $\sigma^2_\epsilon = 0.011$. Panels a and b show results for $\gamma = 0$ and $\gamma = 1$. When present, the economic variable x_t follows a zero-mean AR(1) with autocorrelation 0.95, and σ_x, σ_v set so that half of the variance of β_t comes from x_t and half from the AR(1) process. We simulate 3000 draws from the small sample distribution of $\hat{\alpha}$ and $\hat{\gamma}$, OLS and IV. IV2 is the truncated IV estimator: $\hat{\alpha}_{1,IV2} = Min(0.99, \hat{\alpha}_{1,IV})$. RMSE is the root mean squared error of estimation of the true parameter 0.95. $t(\alpha_1)$ is the t-statistic under the null $\alpha_1 = 0.95$. The test size is based on the 5^{th} percentile of the distribution of a student-t with 169 degrees of freedom, a cut-off point of -1.65. The t-statistic in column IV2 is based solely on the draws that are not truncated.

Panel a: True $\beta_t = 0.05 + 0.95\beta_{t-1} + \upsilon_t$

		OLS	IV	IV2
$\hat{\alpha}_1$	Mean	0.65	0.93	0.92
	5%	0.47	0.79	0.79
	95%	0.81	1.03	0.99
	RMSE	0.32	0.08	0.08
$t(\alpha_1)$	Mean	-5.01	-0.22	-0.36
	Size	0.99	0.027	0.031

Panel b: True $\beta_t = 0.05 + 0.95\beta_{t-1} + x_t + \upsilon_t$

	No x_t in regression		x_t in regression			
	$\hat{\alpha}_{1,OLS}$	$\hat{\alpha}_{1,IV}$	$\hat{\alpha}_{1,OLS}$	$\hat{\gamma}_{OLS}$	$\hat{\alpha}_{1,IV}$	$\hat{\gamma}_{1,IV}$
Mean	0.65	0.93	0.44	0.59	0.80	0.27
5%	0.45	0.79	0.24	0.25	0.52	0.04
50%	0.66	0.94	0.45	0.57	0.81	0.25
95%	0.81	1.04	0.65	0.96	1.01	0.59
RMSE	0.32	0.08	0.52	0.47	0.22	0.75

that 1) a linear process for $(\beta|x_t)$ is only an approximation of a typically more complex structural relationship, and 2) x_t is only a proxy for the relevant economic variable. Note that these errors can not be identified separately, so we are only concerned with the variance of their sum for the purpose of simulation.[13]. The variance of β_t in (7) is still 0.037, but it now comes the AR(1) and the variable x_t. We attribute 50% of the variance to each component. Therefore, we set $\sigma^2_x = \sigma^2_v = 0.0185(1 - 0.95^2)$.

For simulated data from (7), we estimate the model on $\hat{\beta}$, first without, then with, the variable

[13] All these models imply non-identifiable ARMA representations for $\hat{\beta}_T$

x_t on the right-hand-side of the regression. The results are in Panel b of Table 4. The first two columns show the sampling properties of the OLS and IV estimators when the economic variable is not include in the regression. The purpose is to show that our calibration leads to results nearly identical to the experiment with no underlying driving variable. The last four columns show the estimation of α_1 and γ. The results are clear. First the IV estimator still dominates the OLS if the purpose is to estimate properly the dynamics of the true β_t in excess of the economic variable. However, this comes at a cost, namely the estimator, for this sample size underestimates the impact of the driving variable, more than the OLS estimator. Neither estimator can estimate α and γ jointly without downward bias.

The IV regression without the economic variable yields the best estimator of the intrinsic dynamics of the true β conditional on the economic variable. The OLS estimator without lagged betas yields the best estimator of the coefficients of the economic variables. When both economic variables and lagged betas are included, the IV estimator is best for the lagged beta coefficient but not for the economic variables. For our data, we recall that the OLS estimates for the economic and firm-specific variables were similar to the IV estimators. So, we can conclude that the lack of significance of the economic variable is not due to the use of the IV estimator. On the other hand, the simulation results show that the slight decrease in $\hat{\alpha}_{1,IV}$ from the estimation of model 1 and model 1+2+3, may be due to a slight downward bias of $\hat{\alpha}_{1,IV}$ in the presence of x_t.

5. Optimal portfolios

The empirical analysis so far showed that only pure time series models adequately describe beta dynamics. For the purpose of estimating the quarterly dynamics of β_t, we were restricted to block-samplers with quarterly windows. We now turn to a slightly different question: Given that pure time series models for quarterly beta dynamics appear to dominate those driven by economic variables, what is the best filter to forecast market betas. Here, we are no longer constrained to block-sampling filters, we can also use overlapping filters introduced above, for example $\hat{\beta}_{5y,m}$, the classic estimator based on 5 years of monthly returns or $\hat{\beta}_{5y,d}$, based on 5 years of daily returns.

Hence, we turn to the question of the best filtering scheme for betas. If betas were observable, one could rely on the usual regression diagnostics, such as out-of-sample predictability etc. As they are not observable, we will rely on economic criteria directly related to the characterization of conditional betas.

To assess the economic value of various filters of β, we use their forecasts of β_t to optimize portfolios. Recall that a market model regression of returns on market returns serves as a decomposition of the fraction of variance associated to the market return and to the residuals. We construct portfolios with ex-ante minimal residual variance subject to $\beta = 0$. Success will be measured by the realized betas of these portfolios. They should also have very low variance as setting $\beta = 0$ removes the systematic risk. Finally their systematic risk should be a very small fraction of their total risk. This will allow us to assess the economic relevance of any differences between these estimators. We use a residual covariance matrix consistent with $\hat{\beta}$, i.e., the covariance matrix of $R - R_m\hat{\beta}$.[14]

We compare the purely data-driven rolling estimators, $\hat{\beta}_{q,d,t-1}$, $\hat{\beta}_{13m,d,t-1}$, $\hat{\beta}_{5y,m,t-1}$, and $\hat{\beta}_{5y,d,t-1}$, where $t-1$, indicates that the value available at the end of quarter $t-1$ is used to optimize and invest in quarter t. Portfolios are re-optimized quarterly. We then summarize the performance of the optimized portfolios. We also consider, as benchmarks, some strategies which are not feasible. First, $\hat{\beta}_{q,d,t}$, of course not available at $t-1$, represents a perfect foresight of the future estimate of β in quarter t. Second, we report the results of a strategy based on the constant β to assess the economic relevance of the time variation in β modeled by the competing filters.

To gauge the effectiveness of the $\beta = 0$ constraint, we also include the global minimization of the total covariance matrix. We use Ω_t, the perfect knowledge of the future estimate of Ω for the quarter of investment, and Ω_{t-1}, the previous quarter estimate. Ω_{t-1} is computed from daily returns for quarter $t-1$, the same data as used to compute $\hat{\beta}_{q,d,t-1}$. The portfolios are re-balanced quarterly. Results are computed from 1971 to 2004, the period for which all estimators are available.

[14] We also used two alternative, constrained residual covariance matrices. First, we used a diagonal residual covariance matrix. Second, we estimated the residual covariance matrix, constraining the correlations to be equal. The results were about the same as shown here. Also, to concentrate on the comparison of the alternative filters, we did not use potentially beneficial cross-sectional techniques, such as shrinkage.

5.1. Empirical results portfolio efficiency

The results are in Table 5. We conduct two separate experiments, optimizing on the universe of the 25 industries. Panel A includes the portfolios optimized with target betas of zero. The first five columns document the realized β of the optimized portfolios. We first consider the bias, how close to the target were we on average? The bias resulting from $\hat{\beta}_{q,d,t-1}$ is 0.21, twice that for $\hat{\beta}_{13m,d}$, three times as large as for $\hat{\beta}_{5y,d}$, which is the best *simple* estimator on that criterion, with a bias of 0.07. The most popular estimator in Finance, $\hat{\beta}_{5y,m}$ does worse than its daily counterpart $\beta_{5y,d}$ with a bias of 0.2.

Portfolio managers are also affected by the variability of realized quarterly betas. How closely can one track the target from quarter to quarter? The second column in Table 5 shows the standard deviations of quarterly realized portfolio betas over the 151 quarters of the experiment. These standard deviations are about 20% smaller for the shorter windows (1 quarter and 13 months) than for the 5-year windows. The third column shows the Root Mean Squared error around the desired $\beta = 0$, combining bias and variance. The quarterly window and the 5-year window of monthly returns do markedly worse than the 13-month and the 5 year window of daily returns. The best overall estimator is $\hat{\beta}_{13m,d}$. In the next two columns showing the 10^{th} and 90^{th} percentiles of the 151 realized portfolio betas, the upward bias is apparent. The last column document the fraction of total risk which was systematic for these portfolios. Consistent with their better tracking ability, $\hat{\beta}_{5y,d}$ and $\hat{\beta}_{13m,d}$ produced the lowest amount of systematic risk. Finally, according to asymptotic MSE filtering theory, $\hat{\beta}_{13m,d}$ should be about as efficient as $\hat{\beta}_{5y,m}$, it does in fact better.[15]

The following row shows that the realized β of the strategy Ω_{t-1} is 0.34, and the RMSE is 0.40. So, while an unconstrained variance minimization does result in a somewhat low beta, it does not reduce factor exposure as effectively as an explicit zero-beta target combined with strategy.

It is interesting to compare the above strategies with a constant β, which Ghysels (1998) found to dominate parametric beta estimates. Its full-sample look-ahead advantage insures a nearly zero

[15] The technical supplement reports similar results for the 27 portfolios sorted three-ways on the basis of Size, Debt equity, and Market to book ratios.

Table 5 Out-of-sample performance of competing estimators of β: Optimizing on 25 industry sectors

The first column shows the optimization strategy, following the nomenclature in the paper. Strategies with a β minimize residual variance subject to a zero beta. Strategies, Ω_t, Ω_{t-1}, simply minimize variance. β_{cst} uses a constant beta and covariance matrix computed from the entire sample. The first three columns report the mean, standard deviation and RMSE relative to the target, of the realized beta. The next two columns show the top and bottom 10% of the realized β's over the 151 quarters from 1971Q1 to 2004Q1. These are the quarters for which all the strategies can be implemented. The last three columns document the realized portfolio variability. σ is the total standard deviations, σ_ϵ the unsystematic standard deviation from the market model, and Frac. is the systematic variance as a fraction of total variance, all expressed in percentage terms.

			β			variability		
Strategy	Mean	Stdev.	RMSE	10%	90%	σ	σ_ϵ	Frac.
Panel a: Target β is 0.								
$\hat{\beta}_{q,d,t-1}$	0.21	0.21	0.30	-0.04	0.50	11.7	10.9	13
$\hat{\beta}_{13m,d,t-1}$	0.10	0.22	0.24	-0.15	0.38	12.4	11.9	8
$\hat{\beta}_{5y,d,t-1}$	0.07	0.26	0.27	-0.28	0.39	13.4	12.8	9
$\hat{\beta}_{5y,m,t-1}$	0.20	0.25	0.32	-0.08	0.52	12.8	12.1	11
Ω_{t-1}	0.34	0.20	0.40	0.10	0.61	10.5	8.9	28
$\hat{\beta}_{q,d,t}$	0.00	0.00	0.00	0.00	0.00	6.2	6.2	0
$\hat{\beta}_{cst}$	-0.01	0.32	0.32	-0.39	0.42	15.1	14.4	9
Ω_t	0.18	0.08	0.20	0.08	0.29	5.6	5.1	18
Panel b: Target β is 1.								
$\hat{\beta}_{q,d,t-1}$	1.00	0.017		0.978	1.021	13.0	0.9	99
$\hat{\beta}_{13m,d,t-1}$	1.00	0.017		0.977	1.021	13.0	0.9	99
$\hat{\beta}_{5y,d,t-1}$	1.00	0.018		0.979	1.025	13.0	0.9	99
$\hat{\beta}_{5y,m,t-1}$	1.01	0.032		0.979	1.043	13.2	0.9	99
$\hat{\beta}_{q,d,t}$	1.00	0.000		1.000	1.000	13.0	0.5	100
$\hat{\beta}_{cst}$	1.00	0.037		0.967	1.044	13.1	1.2	99

bias. But its variability (0.32 standard deviation) results in a performance comparable to that of the two worst estimators. The constant beta is dominated by the top two time-varying estimators. One can conclude that the *data-driven* beta estimators outperform the constant beta specification.

We now verify whether constraining β to zero effectively helped reduce the variance of the portfolios. The column σ reports the annualized realized portfolio standard deviations. The perfect β foresight portfolio has a standard deviation of 6.2%, approached by none of the feasible β strategies which can lower standard deviations to 11.7% at best. For comparison, the index standard deviation for the period is 15% per year. The perfect variance foresight strategy Ω_t yields a 5.6% standard deviation. The naive variance strategy reaches a 10.5% standard deviation. The constrained $\beta = 0$ strategies approach but do not beat the naive variance strategy.

The $\beta = 0$ strategies may not beat the simple variance minimization simply because the global minimum variance portfolio has a positive, albeit low, beta. Yet, there may be, for very large covariance matrices, benefits from constraining the portfolio to have zero beta, together with a reduction in the number of free parameters in the residual covariance matrix. We simply do not observe this effect for these relatively small portfolios.

The salient fact in table Panel A of Table 5 is that all the estimators systematically "overshoot" the target beta. This result is very important especially for many hedge fund managers. It may be surprising that unbiased estimates lead to systematic realized biases. The reason may be as follows. To set a zero beta, the optimization gives large weights to the lowest betas and small weights to the highest betas. This is because the target of zero lies far below the average beta in the investment universe, one. Now, the smaller (larger) estimated betas are likely to contain negative (positive) estimation errors. As in Fama Macbeth, the smallest (largest) sorted betas likely have negative (positive) estimation error. The optimization then loads up on these underestimated betas and this produces a positive realized beta the next period.

We can verify this by repeating the optimization with a target $\beta = 1$. Here there should be no realize bias. Panel B in Table 5 shows the results. The first column shows that indeed all there is no realized bias anymore. Further the realized standard deviation of the portfolio beta is small, in the order of 0.02 for all the strategies. These optimized portfolios are far more diversified than those in Panel A, therefore more effectively reducing the variability of the portfolio beta. 80% of the time, realized quarterly betas are between 0.98 and 1.03. With a target beta of 1, this good tracking is confirmed by the realized fraction in systematic risk: it approaches 100% for all. The most mediocre performance comes from $\hat{\beta}_{5y,m}$. But the filters all result in very satisfactory tracking of beta. Of course, the variance reduction is limited if one imposes a beta of 1, only 13% standard deviation compared to the 15% for the index.

To summarize hedge funds that neutralize risk, very much like our zero beta portfolios, are very sensitive to estimation error in betas. For them the estimation error in beta turns into systematic positive exposure to risk factors, and it is crucial to choose the best possible estimator of beta. On the other hand, thaditional funds that set target betas in the vicinity of the market beta effectively diversify away estimation error.

5.2. Sensitivity of $\hat{\beta}_{5y,m}$ to predictability in returns

With predictable returns, for example via the firm-specific and global information set variables, the regression of total returns on factors (here the market return) may yield biased estimates of betas. Ideally, one should regress shocks on returns to shocks on factor. A large literature claims some returns predictability even at the monthly horizon, e.g., Ferson and Harvey (1991). Estimators of beta based on daily returns are not affected as predictability is a minute fraction of variance at the daily horizon. For example, the daily standard deviation of the S&P 500 is 17 times its mean over the past 70 years.

The poor performance of the estimator based on monthly returns $\hat{\beta}_{5y,m}$ could possibly be due to a sensitivity to predictability in returns. So we perform a robustness check. We regress the monthly market return on the monthly information variables. The R-square is 8%, the t-statistics of Tbill, TP, DP, DY, $R_{m,t-1}$, are -3.1, 1.5, 4.1, 1.7, -0.2. The residual of this regression estimates the shock to the index return. We run the same regression for the 25 portfolio monthly returns, with R-squares between 5% and 12%. These residuals are the shocks to the portfolio returns. We compute new $\hat{\beta}_{5y,m}$s using these shocks in returns. They are nearly identical to the original. The 25 correlations between their first differences range between 0.96 and 0.98.

6. Conclusions and directions for future research

Modern asset pricing models often imply that the dynamics of betas is linked to global and firm specific state variables. These structural links are in general too complex for direct estimation. The extent empirical literature either uses robust but possibly imprecise pure data-driven filters, or strict parameterizations not always directly related to a structural model and sensitive to misspecifations.

We introduce a linear dynamic model that complements the robustness of filtered betas with the implications on betas of most intertemporal models, without incurring the sensitivity to misspecification of strict parameterizations. We show how to estimate the model without bias by instrumental variables, and that it can be done if the first step β filter is a block sampler, that is, with non-overlapping windows. As the IV estimator is justified asymptotally, we confirm its good properties with small sample simulations. We find that quarterly betas have strong autocorrelation on the order of 0.95. In contrast, the use of standard methods results in much lower autocorrelation around 0.6. Also, the typical proxies for aggregate and firm-specific state variables do not explain much of the time series variation of portfolio quarterly betas.

One can not use overlapping long-window filters to estimate the dynamics of β, but they could still effectively predict future βs. Conversely, a short-window block-sampler, while a required first-step for the estimation of the beta dynamics, may not predict beta precisely. We compare the out-of-sample predictive performance of a number of block-sampling and rolling sample beta filters, ranking them along an economic criterion. Namely, we optimize portfolios subject to target beta constraints, and track the realized performance and betas of these portfolios.

We find that daily returns produce uniformly better beta filters than monthly returns. Namely, the resulting portfolios track the beta target more closely. Also, long windows of 13 months, even five years, do better than the short quarterly window. We use two beta targets; $\beta = 1$ represents a fund with average systematic risk, $\beta = 0$ is akin to a hedge fund neutralizing factor risk.[16] For a target of 1, the realized betas track the target very well for all filters. These portfolio diversify estimation error well. In contrast, zero-beta portfolios systematically overshoot the target $\beta = 0$. This is because they load up on underestimated betas and go short on overestimated betas.

Our results raise questions and suggest possible directions for future research. The inability of most proxies to explain the dynamics of beta seems to contradict the implications of some asset pricing models. But most models are not horizon specific. In contrast, Jacquier et al. (2004) find a

[16] Although the portfolios in our experiment are still fully invested.

cross-sectional link between betas and some growth proxies. Their cross-sectional analysis, which precludes the use of global proxies, relates to long-term links, while our current time series analysis addresses short-run changes. Our method applied on longer horizons could yield differing evidence on the impact of state variables on the dynamics of betas. We included, for parsimony, only a subset of the variables often used as proxies for growth opportunities. Possibly, without degenerating into data mining, alternate proxies might be tested.

A possible search for a better functional form, again without degenerating into data mining, could be of interest. For example, both negative book value following successive negative earnings, and high book-to-market ratio resulting from fallen share prices, can point at few growth prospects in the firm. This could result in a a non-linear relationship between betas and book-to-market. Non-linear conditional beta models have been considered, see e.g., Bansal et al.(1993). The IV procedure introduced here is not restricted to a linear regression setting. One could consider more general models such as $\beta_t - \beta_{t-1} = g_1(\beta_{t-1}) + g_2(X_{t-1}) + v_t$, for some functions g. See, for example Hausman et al. (1995) on the treatment of measurement errors in nonlinear models.

For comparability with the literature, our optimal portfolios are built from, already large, sector portfolios. So our results may overstate the precision of estimation of beta in a stock-based portfolio. The systematic overshoot of beta targets, due to estimation error in individual betas, could be far more dramatic with stock-based portfolios. Finally, we concentrate here on equal-weighted filters for betas. We also studied non-equal weighted filters with no obviously clear advantage over the filters shown here. While these preliminary results were not very optimistic on the ability of alternative filters to improve upon them, it is possible that some declining-weight filters with long windows should also produce good performance.

Finally, the IV estimator that we proposed in this paper may be used for estimation of beta dynamics in a variety of other contexts. For example, Ang et al. (2006) study downside risk via 'negative' beta portfolios and study the time series dynamics of such betas. Our IV estimator applies to Ang et al. (2006) and can correct the bias in the estimated autocorrelations.

Acknowledgments

The authors thank CIRANO and IFM2 for financial support. The paper has benefitted from discussions with René Garcia, Alan Marcus, Eric Renault, Sheridan Titman, and the participants of the CIRANO and HEC workshops.

References

At-Sahalia, Y., P. Mykland, and L. Zhang. 2005. How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* **18** 351–416.

Andersen, T.G, T. Bollerslev, F.X. Diebold, P. Labys, P. 2003. Modeling and Forecasting Realized Volatility. *Econometrica* **71** 529–626.

Andersen, T.G., T. Bollerslev, F. X. Diebold, J. Wu. 2006. Realized Beta: Persistence and Predictability", in T. Fomby and D. Terrell (eds.) *Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series - Part B, Volume 20.* 1–40.

Andreou, E., E. Ghysels. 2006. Sampling frequency and window length trade-offs in data-driven volatility estimation: appraising the accuracy of asymptotic approximations. in T. Fomby and D. Terrell (eds.) *Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series, Part A Volume 20.* 155–182.

Ang, A., J. Chen, Y. Xing. 2006. Downside Risk *Review of Financial Studies,* forthcoming.

Bandi, F.M., and J.R. Russell. 2004. Microstructure noise, realized variance, and optimal sampling. U. of Chicago working paper.

Bansal, R., D. A. Hsieh, S. Viswanathan. 1993. No arbitrage and arbitrage pricing: A new approach. *Journal of Finance* **48** 1231–1262.

Barndorff-Nielsen, O., N. Shephard. 2004. Econometric analysis of realized covariation: high frequency covariance, regression and correlation in financial economics. *Econometrica* **72** 885–925.

Bekaert, G., G. Wu. 2000. Asymmetric volatility and risk in equity markets. *Review of Financial Studies* **13** 1–42.

Berk, J., Green, R., V. Naik. 1999. Optimal Investment, Growth Options, and Security Returns. *Journal of Finance* **54** 1553–1607.

Brennan, M.J., Wang, A.S., Y. Xia. 2004. Intertemporal capital asset pricing and Fama and French three-factor model. *Journal of Finance* **59** 17431776.

Brandt, M., D. Chapman. 2006. Linear Approximations and Tests of Conditional Pricing Models. Working Paper, Boston College and Duke.

Braun, P., D. Nelson, A. Sunier. 1995. Good news, bad news, volatility and betas. *Journal of Finance* **50** 1575–1604.

Brennan, M.J., E.S. Schwartz, R. Lagnado. 1997. Strategic asset Allocation. *Journal of Economic Dynamics and Control* **21** 1377–1403.

Campbell, J. 1996. Understanding risk and return. *Journal of Political Economy* **104** 298–345.

Campbell, J., J. Cochrane. 1999. By force of habit: a Consumption based explanation of aggregate stock market behavior. *Journal of Political Economy* **107** 205–251.

Cochrane, J. 2005. *Asset Pricing* - Revised Edition, Princeton University Press.

DeMiguel, A.-V., L. Garlappi, R. Uppal. 2004. How inefficient are simple asset-allocation strategies? Discussion Paper LBS.

Fama, E.F., J.D. MacBeth. 1973. Risk, return and equilibrium: empirical tests. *Journal of Political Economy* **81** 607–636.

Ferson, W.E., C.R. Harvey. 1991. The variation of economic risk premiums. *Journal of Political Economy* **99** 385–415.

Ferson, W.E., C.R. Harvey. 1993. The risk and predictability of international equity returns. *Review of Financial Studies* **6** 107–131.

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

French, K.R., G.W. Schwert, R.F. Stambaugh (1987), "Expected stock returns and volatility. *Journal of Financial Economics* **19** 3–29.

Garcia, R., E. Renault. 2001. Latent Variable Models for Stochastic Discount Factors. in J. Cvitanic, E. Jouini and M. Musiela (eds.) *Handbooks in Mathematical Finance: Topics in Option Pricing, Interest Rates and Risk Management.* Cambridge University Press.

Ghysels, E. 1998. On Stable factor structures in the pricing of risk: do time-varying betas help or hurt? *Journal of Finance* **53** 549–573.

Gomes, J., L. Kogan, L. Zhang. 2003. Equilibrium Cross Section of Returns. *Journal of Political Economy* **111** 693–732.

Hamada, R.S. 1972. The effects of the firm's capital structure on the systematic risk of common stocks. *Journal of Finance* **27** 435–452.

Harrison, J. M., D. Kreps. 1979. Martingale and Arbitrage in Multiperiod Securities Markets", *Journal of Economic Theory* **20** 381–408.

Harvey, C.R. 1989. Time-varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics* **24** 289–317.

Hsieh, D.. 1991. Chaos and Nonlinear Dynamics: Application to Financial Markets. *Journal of Finance* **46** 1839–1877.

Jacquier, E., Titman, S., Yalcin, A. 2004. Predicting systematic risk: implications from growth options. HEC Working Paper.

Jagannathan, R., Z. Wang. 1996. The conditional CAPM and the cross-section of expected returns. *Journal of Finance* **51** 289–317.

Jostova, G., A. Philipov. 2005. Bayesian analysis of stochastic betas. *Journal of Financial and Quantitative Analysis* **40(4)** 747–778.

Lewellen, J., S. Nagel. 2006. The Conditional CAPM Does Not Explain Asset-Pricing Anomalies. *Journal of Financial Economics* (forthcoming).

Merton, R.C. 1973. An intertemporal capital asset pricing model. *Econometrica* **41** 867–887.

Merton, R.C. 1980. On estimating expected returns on the market: an exploratory investigation. *Journal of Financial Economics* **8** 323–363.

Nelson, D.B. 1996. Asymptotic Filtering Theory for Multivariate ARCH Models. *Journal of Econometrics* **71** 1–47.

Nelson, D.B., D.P. Foster. 1996. Continuous record asymptotics for rolling sample variance estimators. *Econometrica* **64** 139–174.

Petkova, R, L. Zhang. 2005. Is Value Riskier Than Growth? *Journal of Financial Economics*,**78(1)**, 187–202.

Rubinstein, M.E.. 1973. A mean-Variance Synthesis of Corporate Financial Theory. *Journal of Finance* **28** 167–181.

Santos, T., P. Veronesi. 2004. Conditional Betas. Working Paper, GSB Columbia University.

Shanken, J.. 1990. Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics* **45** 99–120.

Smith, W.S., R.L. Watts. 1992. The Investment Opportunity Set and Corporate Financing, Dividend and Compensation Policies", *Journal of Financial Economics* **32** 263–292.

Wang, K. 2003. Asset Pricing with Conditioning Information: A New Test. *Journal of Finance* **58** 161–196.

Figure 1 Competing estimators of β_t for selected industries

The left plots quarterly estimates of $\hat{\beta}_{q,d,q}$, $\hat{\beta}_{13m,d,q}$, and $\hat{\beta}_{5y,d,q}$, with windows of 1 quarter (q), 13 months (13m), and 5 years (5y) of daily returns (d).The right plots show $\hat{\beta}_{5y,d,t}$, its Scholes-Williams version, and $\hat{\beta}_{5y,m,t}$, that uses 5 years of monthly returns. See Table1 for SIC codes.