On the (Im)Possibility of Estimating Expected Return from Risk-Neutral Variance*

Valery Polkovnichenko‡

March 2019

Abstract

Linear equations for expected return on a stock in terms of risk-neutral variance of return form a continuum with indeterminate coefficients. The equations can be identified by setting arbitrary slope or intercept. Risk-neutral variance is a sufficient statistic for expected return under the additional strong restrictions on the cross-section and time-variation of certain second moments of returns. Empirical tests strongly reject an integral component of these restrictions as well as their direct implication of stock-specific constant intercepts, casting doubt on general feasibility of estimating expected returns solely from risk-neutral variances. For some stocks, risk-neutral variance determines upper or lower bound on expected return, independently of the risk aversion in the underlying economy. Combining moments under the risk-neutral and physical distribution, rather than relying on one type exclusively, appears to be a promising path forward.

Keywords: cross-section of expected returns, risk-neutral distribution

*First draft: May 2018. The most recent version of the paper is available from the author’s SSRN page.

‡Board of Governors of the Federal Reserve System, Division of Research and Statistics. The analysis and conclusions set forth are those of the author and do not indicate concurrence by other members of the research staff or the Board of Governors. I am grateful to brown bag seminar participants in the Risk Analysis section of the Division of Research and Statistics for helpful comments and suggestions. I also thank Irisa Lee for excellent research assistance. All errors and omissions remain my own responsibility. Contact: Federal Reserve Board of Governors, Division of Research and Statistics, Washington DC 20551. E-mail: valery.y.polkovnichenko@frb.gov; Phone: (+1)(202) 452-3618.
**Introduction**

Estimating expected return on a stock is a notoriously difficult problem. Traditional methods, based on no-arbitrage equations, require specification of a stochastic discount factor (SDF) and estimation of the conditional covariance of return with SDF using historic data. In a recent paper, Martin and Wagner (2018) propose to circumvent both of these hurdles by developing a methodology for estimating expected return on a stock using only risk-neutral variance of return (henceforth RN or Q variance). They derive a linear equation for expected excess stock return over the index in terms of RN variance of return in excess of its average for the index constituents. The advantages of this methodology are that such linear equations appear to have known coefficients, thus eliminating the need for estimating conditional covariances under the physical probability $P$, and that RN variances can be computed from equity option prices without specifying SDF.

In this paper, I present general analysis of linear equations expressing expected return through RN variance and delineate conditions when the latter is a sufficient statistic for the former. I empirically test the implications of these conditions as well as one of their integral components. I show that potential advantages of these equations are elusive: their coefficients are indeterminate, neither known nor constant, while the assumptions required for their identification do not appear to hold in the data. I conclude that it is generally impossible to estimate expected returns of stocks solely from RN variances of returns.

I show that linear equations expressing expected return on a stock through RN variance (in excess of their respective index averages) form a continuum with indeterminate coefficients. To fix ideas, let $E^P_t\{R_{i,t+1}\}$ be the conditional expected return on stock $i$ for horizon from $t$ to $t+1$ under the physical probability $P$ and $SVIX^2_{i,t} \equiv Var^Q_t \left[ \frac{R_{i,t+1}}{R_{f,t+1}} \right]$ be the RN variance of return scaled by the risk free return $R_{f,t+1}$. Let $I$ be an arbitrary index portfolio of stocks and $\overline{SVIX}^2_t$ denote the weighted average of RN variances for this portfolio. Then, linear equations for expected return on a stock in excess of the index can be written either with common constant slope or with common constant intercept as follows:

$$\frac{E^P_t\{R_{i,t+1}\} - E^P_t\{R_{I,t+1}\}}{R_{f,t+1}} = c_{i,t}(\theta) + \frac{\theta}{2} \left( SVIX^2_{i,t} - \overline{SVIX}^2_t \right), \quad \text{or} \quad (1)$$

$$\frac{E^P_t\{R_{i,t+1}\} - E^P_t\{R_{I,t+1}\}}{R_{f,t+1}} = c + \frac{\theta_i(c)}{2} \left( SVIX^2_{i,t} - \overline{SVIX}^2_t \right), \quad (2)$$
where $\theta \neq 0$ and $c$ are arbitrary real numbers and $c_{i,t}(\theta)$ and $\theta_{i,t}(c)$ are time-varying stock-specific functions of these parameters. There are no theoretically predicted values for slope $\theta/2$ in equation (1) or intercept $c$ in equation (2) and there are infinitely many slope-intercept combinations feasible in each equation.

The indeterminacy of coefficients in equations (1) and (2) is a generic property, independent from the derivation method, and arises from the projection of the full (well-identified) two-factor equation onto the reduced (indeterminate) one-factor equation. As a result, when the coefficients in one-factor equations are estimated as free parameters there are no predictions about their values because they are not identified jointly. Identification of the equations is possible either by fixing arbitrary common slope $\theta/2$ in (1) or intercept $c$ in (2) while estimating the other coefficient as a stock-specific constant. In that case, the only substantiative restriction is on time-variation of the estimated coefficient, which requires additional strong assumptions on the return moments. These assumptions turn out to be unconventional on an introspective basis and they also do not appear to hold empirically.

When arbitrary common constant slope $\theta/2$ is assumed for identification of equation (1), the intercepts $c_{i,t}(\theta)$ are time-varying functions of the slope. If the intercepts are restricted to constant fixed effects $c_i(\theta)$, I show that this assumption can at most hold for one value of the slope.\(^1\) In that case, RN variance is a sufficient statistic for expected return when certain excess conditional covariances are collinear (over time) with excess RN variances and the collinearity coefficient must be the same for all stocks. This constraint guarantees that the only source of stock-specific time-variation in expected returns is restricted to RN variances. On introspection, given a large number of equations with a single unknown parameter, such constraint appears formidable. Moreover, direct empirical tests of constant stock-specific intercepts are rejected (for $\theta = 1$) for sub-portfolios of the S&P 500 index constituents.

To continue theoretical analysis of conditions for RN variance sufficiency, I express them through $P$-moments of stock returns. I use the expected return equations from Martin (2017), derived from no-arbitrage identities, and the approximations for stochastic discount factor (SDF) from Kadan and Tang (2017). I show that stock returns must satisfy a linear cross-sectional constraint on

\(^1\)Otherwise the method is impractical because, as shown in section 2.1, constant intercepts assumption for more than one value of the slope implies constant excess RN variances and constant expected excess returns for all stocks.
conditional $P$-variance (in excess of the index weighted average) and conditional $P$-beta (in excess of one). The coefficients in this constraint are not free parameters but depend on the slope chosen in the expected return equation and risk preferences of the representative investor. As a consequence, the cross-sectional component of the constraint is unlikely to hold for $\theta = 1$ when the relative risk aversion substantially exceeds one (above log utility).

This constraint is a $P$-moments counterpart to the auxiliary assumptions imposed in Martin and Wagner (2018) in terms of RN moments and it appears unlikely to hold generically for stocks with low-to-moderate idiosyncratic risk as they conjecture. The constraint requires “precise amount” of idiosyncratic risk as a function of systematic market variance and preference parameters of the representative investor. It imposes linear cross-sectional structure on the variances of returns and, in the time dimension, the constraint requires a stark one-factor representation of the idiosyncratic conditional (excess) variances. Thus, idiosyncratic components of excess variances of all stocks must be perfectly correlated over time in order to satisfy these conditions while their common driving factor is shown to be a function of conditional market variance. Asset pricing models typically disregard idiosyncratic variance and treat it as unspecified residual due to its irrelevance for expected returns. In contrast, the conditions for RN variance sufficiency impose intricate structure on the second moments of returns and restrict cross sectional and time-series behavior of idiosyncratic components.

The general sufficiency constraint depends on the slope chosen in the expected return equation, but the constraint on one-factor structure of idiosyncratic variances does not. The latter constraint can be independently tested and I present such tests in this paper. The tests require conventional assumptions that conditional variances and covariances can be estimated from historical data and that beta with respect to a broad market index is co-monotone with the true beta with respect to the market portfolio. The tests strongly reject one-factor representation, which is consistent with the rejection of the direct tests of constant intercepts. Because the variance test is independent from the slope, these results put in doubt general feasibility of suppressing time-variation of the intercepts for any value of the slope.

While deriving sufficiency conditions I also show that, for some stocks, RN variances can be used to construct lower or upper bounds on expected returns. These bounds do not depend on the risk aversion in the economy which distinguishes them from the bounds derived in Kadan and Tang.
The bounds require conditional $P$-beta and conditional $P$-variance to be on the “opposite sides” (i.e., below and above) of their respective weighted averages in the index.

The second case, equation (2), with common constant intercept $c$ has additional challenges which negate some of the intended advantages of this method. In that case, the general equations have stock-specific time-varying slopes $\theta_{i,t}(c)$ and restricting them to stock-specific constants $\theta_i(c)$ also requires strong sufficient conditions, similar to those in the identification method with common slopes. Moreover, the equations are correctly specified only within groups of stocks with approximately equal slopes $\theta_i(c)$. Such subsets of stocks are unobservable and their constituents may change over time. Because there is no single value of the slope which applies to the entire cross section, estimating the equations in practice is challenging. Even though an equation applicable to a group of stocks with approximately equal slopes expresses their expected returns only through RN variances, the identification of suitable stocks with a common slope for estimation requires inferences about conditional $P$-moments, precisely those which this methodology attempts to circumvent. On the other hand, imposing a common slope and common intercept for the entire cross-section of stocks mis-specifies the equations.

Overall, my results do not support the case for estimating expected return exclusively from RN variance, but they suggest potential benefits of integrating $Q$-moments with traditional methods. Throughout the paper I rely on the no-arbitrage identities originally proposed in Martin (2017) which offer a potentially better way to estimate expected returns. They combine information from RN variances and certain $P$-covariances and have constant and well-identified coefficients. However, their estimation does require assumptions about SDF as well as traditional use of historic information to estimate conditional $P$-covariances. Martin and Wagner (2018) show that forecasting returns is somewhat more accurate with RN variances than with other traditional variables. This suggests that, in some cases, the benefits of including RN variances may outweigh the specification error introduced from forcing common slope and intercept. Therefore, the methodology which does not introduce specification error and appropriately combines RN variances with $P$-covariances, even if the latter are estimated imprecisely, may prove advantageous over the estimation which relies on them in isolation. Detailed empirical analysis of this interesting venue is outside the scope of this paper and is left to future research.

The related literature includes several recent papers. Martin (2017) estimates a lower bound
on the expected return of the market index portfolio using RN variance computed from the prices of index options. Kadan and Tang (2017) explore analogous bounds for individual stocks under the constraints imposed on the risk aversion of the representative agent. Using the approximations derived by Kadan and Tang (2017) I derive different expected return bounds, applicable to certain stocks, which are independent of the risk aversion in the underlying economy. The expected return equations discussed in this paper are derived by Martin and Wagner (2018, 2016). Both versions obtain the same expected return formulas by using two different derivation methods and henceforth I refer to them as MW, specifying the version only when it is necessary for clarity.

The contribution of my paper is in extending theoretical analysis of the expected return equations and their identification conditions, while the empirical tests are focused on the assumptions underpinning the theory.

1 Expected returns and RN variances

1.1 Notation and preliminary derivations

Time is discrete and there is a well defined probability space for all relevant random variables. There is a state variable (vector) $X_t$ which follows a Markov stochastic process and drives the conditional distributions of returns on traded assets. Denote $P_t \equiv P(\cdot \mid X_t)$ as the (physical) conditional probability distribution of state $X_{t+1}$ at some future date $t+1$ with the corresponding conditional expectation denoted as $E^P\{\cdot \mid X_t\} \equiv E^P_t\{\cdot\}$. Period length from $t$ to $t+1$ is left unspecified and can be chosen in the empirical application. There are $N$ traded stocks with stochastic (gross) returns between $t$ and $t+1$ denoted as $R_{i,t+1}$. There is also a traded index portfolio of stocks with $N \times 1$ vector of weights $w_t = (w_{1,t}, \ldots, w_{N,t})^\top$ and $1^\top w_t = 1$, where $1$ is $N \times 1$ vector of ones. The index return is given by $R_{I,t+1} = w_t^\top R_{t+1}$, where $R_{t+1} = (R_{1,t+1}, \ldots, R_{N,t+1})^\top$ is $N \times 1$ vector of stock returns. In addition to the risky assets, assume there exists a risk-free security which deterministically pays one unit of account at $t+1$ with return denoted as $R_{f,t+1}$.

In the absence of arbitrage there exists a random variable $m_{t+1} > 0$ (a.s.) such that

\begin{align}
E^P_t\{m_{t+1} R_{i,t+1}\} &= 1, \ i = 1, \ldots, N \\
E^P_t\{m_{t+1}\} &= \frac{1}{R_{f,t+1}}
\end{align}

\textsuperscript{2}Byström (2018) calculates RN variances from CDS contracts and uses MW formula, so the results developed here apply to the related part of the theory underlying his work.
where, without loss of generality, I normalize $m_t = 1$. The variable $m_{t+1}$ is referred to as the pricing kernel or stochastic discount factor (SDF). It can also be used by invoking Radon-Nikodym theorem to define a new probability measure $Q_t$ which is absolutely continuous with respect to physical probability $P_t$ such that:

$$\frac{dQ_t}{dP_t} = R_{f,t+1} m_{t+1}$$

The measure $Q_t$ is referred to as the risk-neutral (RN) probability and I denote the expectation with respect to it by superscript $Q$ (e.g. $E^Q_t\{\cdot\}$) to distinguish it from the expectation taken under the physical probability $P$ (e.g. $E^P_t\{\cdot\}$). Other moments, such as variances and covariances, under the two measures are also distinguished by similar superscripts. I refer to $Q$, $P$ and $m$ without time subscripts where it does not cause confusion.

From the definitions of $Q$ and $m$ it follows that:

$$E^P_t\{m_{t+1} R_{i,t+1}\} = \frac{1}{R_{f,t+1}} E^Q_t\{R_{i,t+1}\} = 1, \quad i = 1, \ldots, N$$

Define $R_{g,t+1}$ as follows:

$$R_{g,t+1} \equiv \frac{1}{m_{t+1}}$$

where $R_{g,t+1}$ is a return because its price is equal to one by construction. Although $R_{g,t+1}$ may not be in the span of $R_{i,t+1}$'s, this possibility has no effect on any derivations to follow.\(^3\)

The following derivations appear in Martin (2017) and are subsequently used as the starting point to derive MW formula for expected return on a stock. Consider two securities $i$ and $j$ (which could be individual stocks or arbitrary portfolios such as index $I$) and write for their returns:

$$E^Q_t\{R_{i,t+1} R_{j,t+1}\} = E^Q_t\{R_{i,t+1}\} E^Q_t\{R_{j,t+1}\} + Cov^Q_t(R_{i,t+1}, R_{j,t+1})$$

$$= R^2_{f,t+1} + Cov^Q_t(R_{i,t+1}, R_{j,t+1})$$

where the second equality follows from the property of $Q$ in (6). Consider now the same expectation after the change of measure:

$$\frac{1}{R_{f,t+1}} E^Q_t\{R_{i,t+1} R_{j,t+1}\} = E^P_t\{m_{t+1} R_{i,t+1} R_{j,t+1}\}$$

$$= E^P_t\{m_{t+1} R_{j,t+1}\} E^P_t\{R_{i,t+1}\} + Cov^P_t(m_{t+1} R_{j,t+1}, R_{i,t+1})$$

$$= E^P_t\{R_{i,t+1}\} + Cov^P_t(m_{t+1} R_{j,t+1}, R_{i,t+1})$$

\(^3\)This return has similar properties to the growth-optimal portfolio return used in MW derivations. It is without loss of generality which return to use here since it later drops out from the formulas. I use the reciprocal of SDF because it does not require additional definitions and in complete markets the two returns coincide.
Use (8) and (9) to obtain:

\[ R_{f,t+1} + \frac{1}{R_{f,t+1}} \text{Cov}^Q_t(R_{i,t+1}, R_{j,t+1}) = E_t^P \{ R_{i,t+1} \} + \text{Cov}^P_t(m_{t+1} R_{j,t+1}, R_{i,t+1}) \]  

(10)

Setting \( j = i \) in the above gives:

\[ R_{f,t+1} + \frac{1}{R_{f,t+1}} \text{Var}^Q_t(R_{i,t+1}) = E_t^P \{ R_{i,t+1} \} + \text{Cov}^P_t(m_{t+1} R_{i,t+1}, R_{i,t+1}) \]  

(11)

Following MW notation, define additional quantities as follows:

\[
\text{SVIX}_{i,t}^2 = \text{Var}^Q_t \left[ \frac{R_{i,t+1}}{R_{f,t+1}} \right] 
\]

(12)

\[
\overline{\text{SVIX}}^2_t = \sum_{i=1}^{N} w_{i,t} \text{SVIX}_{i,t}^2 
\]

(13)

MW also utilize another result from Martin (2017) with regards to the bound on expected return on the market portfolio. Using this bound, the index risk premium can be substituted into the expression for expected excess return on a stock over the index to find the excess return of the stock over the risk free rate. For brevity, I focus on the expected return in excess of the index.

1.2 Expected returns from RN variances

In this section I reproduce the derivation from the published version of MW (2018) and Appendix A follows the earlier working paper version MW (2016) with a slightly different derivation to obtain the same equations. To proceed, start with equation (10) and set \( j = g \) to derive:

\[ E_t^P \left\{ \frac{R_{i,t+1}}{R_{f,t+1}} \right\} - 1 = \text{Cov}^Q_t \left[ \frac{R_{i,t+1}}{R_{f,t+1}}, \frac{R_{g,t+1}}{R_{f,t+1}} \right] \]  

(14)

Consider the following decomposition of return \( i \) using a linear projection under the \( Q \)-measure:

\[
\frac{R_{i,t+1}}{R_{f,t+1}} = \alpha_{i,t}^Q + \beta_{i,t}^Q \frac{R_{g,t+1}}{R_{f,t+1}} + u_{i,t+1} 
\]

(15)

\[ E_t^Q \{ u_{i,t+1} \} = 0 \]  

(16)

\[ E_t^Q \left\{ \frac{R_{g,t+1}}{R_{f,t+1}} u_{i,t+1} \right\} = 0 \]  

(17)

\[ \beta_{i,t}^Q \equiv \frac{\text{Cov}^Q_t \left[ \frac{R_{i,t+1}}{R_{f,t+1}}, \frac{R_{g,t+1}}{R_{f,t+1}} \right]}{\text{Var}^Q_t \left[ \frac{R_{g,t+1}}{R_{f,t+1}} \right]} \]  

(18)
where $\alpha_{i,t}^Q$ is a conditional constant. Using the definition of RN beta in (18) rewrite (14) as follows:

$$
E_t^P \left\{ \frac{R_{i,t+1}}{R_{f,t+1}} \right\} - 1 = \beta_{i,t}^Q Var_t^Q \left[ \frac{R_{g,t+1}}{R_{f,t+1}} \right]
$$

(19)

Equation (15) implies the following variance decomposition:

$$
Var_t^Q \left[ \frac{R_{i,t+1}}{R_{f,t+1}} \right] = \left( \beta_{i,t}^Q \right)^2 Var_t^Q \left[ \frac{R_{g,t+1}}{R_{f,t+1}} \right] + Var_t^Q \left[ u_{i,t+1} \right]
$$

(20)

The next step is to linearize $\left( \beta_{i,t}^Q \right)^2$ at $\beta_{i,t}^Q \approx 1$ for the special case of the equations. MW (2018, in fn. 5 and in the Internet Appendix A) note that coefficients in the expected returns equation depend on the choice of linearization location but they do not investigate this property or its implications further. I follow their more general derivation here which can be specialized by setting $\theta = 1$ to obtain the equations from the main text. To proceed, note that for any real number $\theta \neq 0$ there is a linearization near $\beta_{i,t}^Q = \frac{1}{\theta}$:

$$
\left( \beta_{i,t}^Q \right)^2 = \frac{2}{\theta} \beta_{i,t}^Q - \frac{1}{\theta^2} + \left( \beta_{i,t}^Q - \frac{1}{\theta} \right)^2 \approx \frac{2}{\theta} \beta_{i,t}^Q - \frac{1}{\theta^2}
$$

(21)

with the approximation error equal to $\left( \beta_{i,t}^Q - \frac{1}{\theta} \right)^2$. Then substitute the approximation into (20) to obtain:

$$
Var_t^Q \left[ \frac{R_{i,t+1}}{R_{f,t+1}} \right] \approx \left( \frac{2}{\theta} \beta_{i,t}^Q - \frac{1}{\theta^2} \right) Var_t^Q \left[ \frac{R_{g,t+1}}{R_{f,t+1}} \right] + Var_t^Q \left[ u_{i,t+1} \right]
$$

(22)

Now use (22) to substitute out the right-hand side in (19) to obtain:

$$
E_t^P \left\{ \frac{R_{i,t+1}}{R_{f,t+1}} \right\} - 1 \approx \frac{\theta}{2} Var_t^Q \left[ \frac{R_{i,t+1}}{R_{f,t+1}} \right] + \frac{1}{2\theta} Var_t^Q \left[ \frac{R_{g,t+1}}{R_{f,t+1}} \right] - \frac{\theta}{2} Var_t^Q \left[ u_{i,t+1} \right]
$$

(23)

Then multiply by $w_{i,t}$ to sum up and obtain expected excess return of the index portfolio:

$$
\frac{E_t^P \{ R_{i,t+1} \} - R_{f,t+1}}{R_{f,t+1}} \approx \frac{\theta}{2} \sum_{i=1}^{N} w_{i,t} Var_t^Q \left[ \frac{R_{i,t+1}}{R_{f,t+1}} \right] + \frac{1}{2\theta} Var_t^Q \left[ \frac{R_{g,t+1}}{R_{f,t+1}} \right] - \frac{\theta}{2} \sum_{i=1}^{N} w_{j,t} Var_t^Q \left[ u_{j,t+1} \right]
$$

(24)

Next, subtract (24) from (23) to eliminate the (unobservable) RN variance of $R_{g,t+1}$:

$$
\frac{E_t^P \{ R_{i,t+1} \} - E_t^P \{ R_{i,t+1} \}}{R_{f,t+1}} \approx \frac{\theta}{2} \left( Var_t^Q \left[ \frac{R_{i,t+1}}{R_{f,t+1}} \right] - \sum_{j=1}^{N} w_{j,t} Var_t^Q \left[ \frac{R_{j,t+1}}{R_{f,t+1}} \right] \right)
$$

$$
- \frac{\theta}{2} \left( Var_t^Q \left[ u_{i,t+1} \right] - \sum_{j=1}^{N} w_{j,t} Var_t^Q \left[ u_{j,t+1} \right] \right)
$$

(25)

---

4 Equations (23), (24) and (27) are analogous to equations (IA.A.2), (IA.A.3), and (IA.A.4), respectively, in the Internet Appendix A of MW (2018) (with $c = 2/\theta$ and $d = 1$). I reproduce their derivations here with a minor change of notation and in Appendix A I follow a slightly different derivation from MW (2016) to show similar properties.
To derive the expected return formula amenable to estimation as a panel regression, MW assume the following additive decomposition of the residual RN variance from the linear projection:

\[ \text{Var}_t^Q [u_{i,t+1}] = \phi_i + \psi_t \]  \hspace{1cm} (26)

This implies stock fixed effects representation in (25) by defining \( \alpha_i = -\frac{\theta}{2} \left( \phi_i - \sum_{j=1}^{N} w_{i,t} \phi_j \right) \) with the property that \( \sum_{i=1}^{N} w_{i,t} \alpha_i = 0 \). Now apply the definitions from (12) and (13) to write (25) as follows:

\[ \left( \frac{E_t^P \{ R_{i,t+1} \} - E_t^P \{ R_{f,t+1} \}}{R_{f,t+1}} \right) \approx \frac{\alpha_i(\theta)}{2} + \frac{\theta}{2} \left( \text{SVIX}_{i,t+1}^2 - \text{SVIX}_{t}^2 \right) \]  \hspace{1cm} (27)

The coefficients in (27) are not identified jointly, the equation holds for arbitrary value of linearization location parameter \( \theta \neq 0 \) and appropriately defined intercept \( \alpha_i(\theta) \). Although the derivation here follows specific method from MW (2018), the next section shows that equations like (27) and (25) represent a continuum of linear projections of expected returns onto RN variances and are not identified in general, regardless of the method used to derive them. MW choose \( \theta = 1 \) (slope of \( \frac{1}{2} \) or, equivalently, \( c = 2 \) in fn. 5 of their paper) and motivate it by relatively small approximation error from linearization near \( \beta_{i,t}^Q \approx 1 \). I show in section 2.1 that approximation is irrelevant in this context: the equations of the type in (27) are unidentified in general and are exact for arbitrary slope. Moreover, the analysis of sufficient conditions for constant intercepts, presented in section 2.3.1, suggests that they are unlikely to be satisfied for \( \theta = 1 \) and reasonable values of risk aversion typically used in asset pricing models.

## 2 Continuum of linear equations

To better understand the properties of equation (27), I use an alternative formula linking expected returns with RN variances. The formula follows from the no-arbitrage identity (11). The alternative equation contains both RN and \( P \)-measure moments and has similar structure to (27) but it is obtained without auxiliary assumptions. It reveals the origins of indeterminacy of coefficients and allows to explore the assumptions required to suppress time-variation in the coefficients.

To proceed, multiply (11) by \( w_{i,t} \) and sum up across stocks in the index to obtain:

\[ R_{f,t+1} + \frac{1}{R_{f,t+1}} \sum_{i=1}^{N} w_{i,t} \text{Var}_t^Q (R_{i,t+1}) = E_t^P \{ R_{f,t+1} \} + \sum_{i=1}^{N} w_{i,t} \text{Cov}_t^P (m_{t+1} R_{i,t+1}, R_{i,t+1}) \]  \hspace{1cm} (28)
Subtract (28) from (11) and rearrange using previous definitions to obtain:

\[
\frac{E^P_t \{ R_{i,t+1} \} - E^P_t \{ R_{i,t+1} \}}{R_{f,t+1}} = \alpha_{i,t} + \left( \text{SVIX}^2_{i,t} - \text{SVIX}^2_t \right),
\]

(29)

where I define the new term for (negative of) excess covariance

\[
\alpha_{i,t} \equiv -\frac{1}{R_{f,t+1}} \left( \text{Cov}^P_t (m_{t+1} R_{i,t+1}, R_{i,t+1}) - \sum_{j=1}^N w_{j,t} \text{Cov}^P_t (m_{t+1} R_{j,t+1}, R_{j,t+1}) \right).
\]

(30)

Note that \( \sum_{i=1}^N w_{i,t} \alpha_{i,t} = 0 \) by construction.

Equation (29) is similar to (27) in that both equations connect expected returns with RN variances in excess of their cross-sectional weighted averages but with two important differences. First, \( \alpha_{i,t} \) is asset-time-dependent excess covariance while \( \alpha_i(\theta) \) is assumed to be only asset-dependent. The constraints required for stock fixed effects can be inferred from comparing these equations and they are analyzed in section 2.3. Second, the coefficient \( \theta_2 \) in (27) is not identified while that in (29) is equal to unity. This lack of identification is analyzed next.

2.1 Projection and identification with common slope

Equation (29) suggests that indeterminacy of the coefficients in (27) is the consequence of suppressing stock-specific time variation of expected return from sources other than RN variance. Note that equation (29) has two factors: the excess RN variance and \( \alpha_{i,t} \), the negative of excess P-covariance. On the other hand, equation (27) restricts the expected returns to be a one-factor equation with excess RN variance and a constant intercept. A priori, there is no theoretical reason for excess RN variances to be the only driver of stock-specific time-variation in expected returns, displacing excess P-covariances \( \alpha_{i,t} \). As shown in section 2.2, time-variation of returns not captured by the RN variances is substantial and assumption of constant intercepts in (27) is not supported empirically. Overall, the case to omit P-covariances and retain RN variances appears to rest primarily on the ability to compute model-free estimates of the latter from option prices.

The intuition for loss of identification can be understood from the projection geometry shown on Figure 1. The figure shows expected excess stock return over the index on the vertical Z axis as a linear function of two factors: excess RN variance and (the negative of) excess covariance \( \alpha_{i,t} \) on the X and Y axes, respectively. Consider the left panel first. A stock is labeled by a small blue circle and is located on the gray plane which represents full two-factor equation (29) and has unit slopes.
Figure 1: Projecting two-factor equation onto one-factor equation

Gray plane represents full two-factor equation (29) with unit slopes on both axes. Blue plane contains projections (lines) corresponding to one factor equations (27) with arbitrary slopes. **Left panel**: Arbitrary red lines lying within the gray plane are drawn through a stock marked by a small blue circle in the gray plane. The lines’ projections on the blue plane are shown as black solid lines and a stock’s projection onto a blue plane is connected to its original location by the dashed blue line. **Right panel**: Arbitrary parallel red lines are lying within the gray plane. The red lines are projected onto the blue plane as black solid lines with arbitrarily slope. A stock is marked by a small blue circle and is connected by the dashed blue line to its projection on the blue plane.

along both axes. The vertical (blue) plane contains one-factor equations (lines lying within the blue plane) expressing expected return only in terms of excess RN variance. Arbitrary (red) lines lying withing the gray plane can be drawn through the stock, as shown by the two examples in the figure. The lines’ projections are marked as solid black lines on the blue plane, each corresponding to an instance of equation (27) and expressing expected return only in terms of excess RN variance. Since infinitely many lines within a gray plane can go through the same point, there are infinitely many projection lines with their corresponding slope-intercept combinations on the blue plane.

This basic geometry explains how there could be infinitely many equations (27) for expected returns with arbitrary slope common across stocks. This is illustrated in the right panel on Figure 1. A stock is again labeled by the small blue circle within the gray plane and it is connected to its projection on the blue plane by the dashed blue line. One can draw arbitrary parallel red lines within a gray plane, project them onto the blue plane (shown as the black lines within blue plane), and obtain one-factor equations (27) with arbitrary common slope and appropriately defined intercepts.

This argument is based only on the fact that infinitely many lines can be drawn through a
point on a plane and produce infinitely many projections onto another non-orthogonal plane. Thus, coefficient indeterminacy in equation (27) is generic as it results from the projection of the full two-factor equation onto one-factor equation, and emerges independently of the method used to derive one-factor equation.\(^5\)

To obtain the projected equations algebraically, define a time-varying function \(c_{i,t}(\theta)\) on real-valued \(\theta \neq 0\) as follows:

\[
c_{i,t}(\theta) \equiv \alpha_{i,t} - \left(\frac{\theta}{2} - 1\right) \left(\text{SVIX}^2_{i,t} - \text{SVIX}^2_t\right)
\]

(31)

Substituting \(\alpha_{i,t}\) from (31) into (29) gives:

\[
\frac{E_P^r\{R_{i,t+1}\} - E_P^r\{R_{f,t+1}\}}{R_{f,t+1}} = c_{i,t}(\theta) + \frac{\theta}{2} \left(\text{SVIX}^2_{i,t} - \text{SVIX}^2_t\right)
\]

(32)

for arbitrary slope coefficient \(\frac{\theta}{2}\) and suitably defined intercept \(c_{i,t}(\theta)\). Note that neither the slope nor the intercept in the above equation are constrained in any way by the theory beyond the equality in (31) and there are infinitely many slope-intercept pairs which satisfy this condition. Thus, the projected equations constitute a continuum which can be parameterized by a common slope \(\theta\) or by a common intercept \(c\) if their roles are reversed in (31).

Also note that equation (32) is exact and contains no approximation error. It is analogous to the approximate equation (25) obtained using linearization but, regardless of the method, the coefficients in both equations are indeterminate. Linearization error affects only the intercept and in equation (32) the intercept \(c_{i,t}(\theta)\) adjusts to make the equation precise. As shown in section 2.3, some form of variance linearization in beta is required in the sufficient conditions for constant intercepts.

There are two ways to identify projected equations: either impose arbitrary constant value of the slope \(\theta\) or the intercept \(c\) in (31) and leave the other coefficient to be a function of the chosen parameter. In general, the unconstrained coefficient is stock-specific and time-varying so additional assumptions are required for estimation as a panel regression with fixed effects or with constant stock-specific slopes. These cases are considered in greater detail in the subsequent sections.

It is only feasible (practical) to assume stock fixed effects in (32) for at most one value of the slope, otherwise the equation would collapse to a trivial case with constant expected excess returns.

\(^5\)As shown in Appendix A, an alternative methodology from MW (2016) also results in similar indeterminacy.
Since equation (32) holds for any $\theta$, differencing it for two distinct $\theta_1$ and $\theta_2$ gives:

$$c_{i,t}(\theta_1) - c_{i,t}(\theta_2) = -\frac{\theta_1 - \theta_2}{2} \left( \text{SVIX}^2_{i,t} - \overline{\text{SVIX}}^2_t \right)$$  \hspace{1cm} (33)

If both $c_{i,t}(\theta_1)$ and $c_{i,t}(\theta_2)$ are assumed to be stock-specific constants, then the excess RN variances on the right hand side also must be constant over time and so are the expected excess returns. Figure 1 (right panel) helps to understand the intuition for this observation. Constant intercepts require that, over time, stocks can only move in parallel along the red lines on the gray plane (like beads threaded on the wire). This a very strong requirement and it means that excess covariance $\alpha_{i,t}$ and the excess $\text{SVIX}^2_{i,t}$ must be collinear (over time) for each stock with the same slope coefficient as required in (31) and with constant stock-specific intercepts $c_i(\theta)$. This guarantees that a stock’s projection moves along the black line with constant intercept, i.e. the RN variance is the only source of stock-specific time variation in excess return. If this were true for any value of the slope $\theta$, i.e. in any direction of the black and red lines, then it would imply that stocks never change their location on the gray plane, an unrealistic and undesirable property. Nonetheless, the assumption of constant intercepts may hold for some unique value of $\theta$ and theoretical sufficient conditions for this case are analyzed in subsection 2.3.

For empirical estimation, MW assume common slope with stock fixed effects and estimate a panel regression using the realized excess returns of stocks over the index as follows:

$$\frac{R_{i,t+1} - R_{I,t+1}}{R_{f,t+1}} = \alpha_i + \gamma \left( \text{SVIX}^2_{i,t} - \overline{\text{SVIX}}^2_t \right)$$  \hspace{1cm} (34)

In this specification, $\gamma$ and $\alpha_i$ are not identified jointly because the projection equation is indeterminate without fixing one of its coefficients. Although the estimation is still feasible, as in the omitted variable regression, a test for any specific value of $\gamma$ is not well defined because there is no theoretical prediction about it. Note that equation (34) is the version of (29) with omitted variable of excess covariances $\alpha_{i,t}$ replaced by the fixed effect. When $\alpha_{i,t}$ factor is dropped, there is no theoretical prediction about the coefficient $\gamma$ on the excess RN variance factor.

The equation, however, can be tested by estimating the intercepts $\hat{\alpha}_i(\bar{\theta})$ for some arbitrary fixed slope $\gamma = \frac{\bar{\theta}}{2}$. While there are no predictions for the values of intercepts, they must be constant over time. This is the only substantiative restriction imposed by the equation. In the next subsection I empirically test the restriction of constant intercepts for sub-portfolios of the S&P 500 index constituents. These tests uniformly and strongly reject constant intercepts constraint for $\bar{\theta} = 1$.  

13
Importantly, the empirical tests are specific to $\bar{\theta}$ because fixed effect assumption cannot hold universally or else it implies constant expected excess returns. However, if there exist a $\bar{\theta}$ which satisfies fixed effects conditions, it is not clear how to search for it. As shown in subsection 2.3, the requirements for constant intercepts are rather demanding and do not emerge naturally from asset pricing theory. Additionally, a component of these constraints which does not depend on $\theta$ is rejected in empirical tests discussed in section 2.4, which further puts in doubt feasibility of the fixed-effects assumption.

Equation (29) can also be used to estimate expected returns. To do that, one can attempt to estimate SDF $m$ using some factors (or portfolios of test assets) and then estimate requisite $P$-covariances in $\alpha_{i,t}$ from historic data. Of course, equation (11) can also be used to estimate expected return without using the index constituents. Which equations would perform better empirically is not clear a priori but there may be a benefit from combining forward-looking RN variances with traditional approach of estimating covariances from historic data. Detailed exploration of this issue presents interesting venue for future research.

2.2 Testing for constant intercepts

In this section I present direct tests of the fixed-effect intercept assumption in the expected return equation (32). I test this restriction at portfolio level because individual stock excess returns contain significant idiosyncratic noise which weakens the tests. Note that equation (32) can be summed up with weights $w_{i,t}^p \left( \sum_{i \in p} w_{i,t}^p = 1 \right)$ for arbitrary portfolio of stocks $p$:

$$\frac{E_t^P \{ R_{p,t+1} \} - E_t^P \{ R_{I,t+1} \}}{R_{f,t+1}} = c_{p,t}(\bar{\theta}) + \frac{\theta}{2} \left( \bar{SVIX}^2_{p,t} - \bar{SVIX}^2_t \right),$$

where $\bar{SVIX}^2_{i,t}$ is defined analogously to $\bar{SVIX}^2_t$ as the weighted average of $SVIX^2_{i,t}$ of the stocks in the portfolio and $c_{p,t}(\bar{\theta})$ is the weighted average of $c_{i,t}(\theta)$. For a given $\bar{\theta}$ I use realized returns and RN variance of the portfolio $p$ and the index to compute the estimator of the intercept as follows:

$$\hat{c}_{p,t}(\bar{\theta}) = \frac{R_{p,t+1} - R_{I,t+1}}{R_{f,t+1}} - \frac{\bar{\theta}}{2} \left( \bar{SVIX}^2_{p,t} - \bar{SVIX}^2_t \right)$$

When individual stocks have constant intercepts, their portfolios should also have constant intercepts, provided that the weights in the portfolio are constant or do not vary considerably over
time. Thus for any two periods \( t \neq t' \), the constraint implies:

\[
\hat{c}_{p,t}(\bar{\theta}) = \hat{c}_{p,t'}(\bar{\theta}) = c_p(\bar{\theta})
\]

To test this restriction empirically, I construct the estimates of RN variance using equity options for stocks which were constituents of the S&P 500 index between January 1996 and December 2017. The data for volatility surfaces is obtained from Options Metrics and the data for stock returns is from the Center for Research in Security Prices (CRSP), both accessed through Wharton Research Data Services (WRDS). The test portfolios \( p = \{1, 2, 3, 4, 5\} \) are constructed by sorting stocks into quintiles based on the average SVIX\(^2\)\(_{i,t}\) over every two years to minimize rotation across portfolios. Portfolios are either value-weighted using average market capitalization during the year or equally weighed. Further technical details of sample construction and numerical computation of the RN variance measures are discussed in Appendix C.

The summary statistics for returns, RN variances and intercepts for each portfolio-maturity combination are presented in Table 1. There are on average about 477 index constituents, around 90 or more per portfolio, in a given month with available returns and RN variances. Average returns and volatility tend to increase with RN variance quintiles. The average intercepts are small relative to the average returns but the volatility of the intercepts is comparable to that of returns. High volatility of the intercepts indicates that there is considerable portfolio-specific variation in returns which is not accounted for by the variation in excess RN variance. Figure 5 shows the time series plots of value-weighted intercepts for each portfolio-maturity combination (equally-weighted plots are qualitatively very similar and are omitted for brevity). The plots confirm considerable variation of the intercepts over time. Visual examination of these series suggests that intercepts are not constant for any portfolio-maturity pair and I next subject this assumption to a formal test.

I test the restriction (37) for \( \bar{\theta} = 1 \) using monthly, quarterly, semi-annual, and annual maturity (\( h = \{30, 91, 182, 365\} \) days) for standardized options provided in the volatility surface files in Option Metrics. I test the restriction using the series of non-overlapping data for each frequency and construct pairs of intercepts at \( t \) and \( t' \) up to four periods ahead for each \( t \) (\( t < t' \leq t + 4 \)), to
estimate the regression equation\(^6\):

\[
\hat{c}_{p,t}(1) = a + b \times \hat{c}_{p,t'}(1) + e_{p,t,t'}
\]  

(38)

If the intercepts are constant, the estimates should satisfy \(a = 0\) and \(b = 1\). Table 2 shows the estimates of \(a\) and \(b\) along with their \(t\)-statistics (for equality to zero) for each portfolio \(p\) and maturity \(h\) combination with Panels A and B, respectively, showing the results for value-weighted and equally-weighted portfolios. The table also reports \(R^2\) and the number of observations for each regression. The last column shows \(p\)-value of the \(F\)-test for joint restriction \(a = 0\) and \(b = 1\).

The intercepts \(a\) are often statistically significantly different from zero, especially in the case of equally-weighted portfolios. On the other hand, the vast majority of slope coefficients \(b\) are not significantly different from zero. The \(R^2\)’s of all regressions are very low, consistent with low correlation between portfolio intercepts across different time periods and with high variation of the intercepts observed on Figure 5. The test of joint restriction \(a = 0\) and \(b = 1\) is strongly rejected with \(p\)-values equal to zero for all portfolio-horizon combinations. Overall these results do not support constant intercept constraint for \(\bar{\theta} = 1\). Although theoretically there could be another value of \(\bar{\theta}\) for which the constraints hold, the theory makes no predictions of what that value might be or how to find it. In the next section I investigate sufficient conditions for constant intercepts to show their unconventional nature.

### 2.3 Sufficient conditions for constant intercepts

Estimation of equation (27) as a panel regression requires additional assumption that intercepts are stock-specific constants. This assumption requires that excess RN variances and excess covariances in \(\alpha_{i,t}\) are collinear (over time) so that stocks remain on the same equation lines. Graphically, on Figure 1 (right panel), this implies that, over time, the movement of all relevant excess conditional moments must be along parallel red lines in the gray plane in order to maintain constant intercepts of the black projection lines on the blue plane.

To investigate the conditions for stock-specific constant intercepts I use equations (27) and (29) with the latter equation derived from identities without auxiliary assumptions. By comparing the

---

\(^6\)Thus, each intercept is used up to four times as the left hand side variable resulting in the total of \(N = 4T - 10\) observations for the regression, where \(T\) is the number of non-overlapping observations available for each frequency. Calendar months and quarters are used for monthly and quarterly frequency, semiannual observations begin in January and July while annual observations begin in January.
two equations I can back out the implied value of \( \alpha_i(\theta) \):

\[
\alpha_i(\theta) = \alpha_{i,t} + \left(1 - \frac{\theta}{2}\right) \left(SVIX_{i,t}^2 - \overline{SVIX}_t^2\right) - \frac{1}{R_{f,t+1}} \left(Cov_t^P(m_{t+1}R_{i,t+1}, R_{i,t+1}) - \sum_{j=1}^{N} w_{j,t}Cov_t^P(m_{t+1}R_{j,t+1}, R_{j,t+1})\right)
\]

(39)

Note that by construction, and consistent with additive decomposition assumption in (26), weighted \( \alpha_i(\theta) \)'s sum up to zero in the cross section at every time \( t \). However, the terms comprising \( \alpha_i(\theta) \) are generally asset-time-dependent. Given large dimension of typical empirical cross section of stocks, a system of such conditions is unlikely to hold for some fixed \( \theta \) for all securities without additional constraints on returns.

One straightforward, if impractical, way to assure that \( \alpha_i(\theta) \)'s are stock fixed effects for any \( \theta \) would be to assume that RN variances \( SVIX_{i,t}^2 \) and \( P \)-covariances in \( \alpha_{i,t} \) are each additively separable into time effects and asset-specific constant terms so that their cross-sectional deviations from the weighted mean are asset-specific constants. That is, to assume that \( SVIX_{i,t}^2 = a_i + b_t \) and \( Cov_t^P(m_{t+1}R_{i,t+1}, R_{i,t+1}) = c_i + d_t \). However, such an assumption would imply that the right hand side of (27) is a stock-specific constant and it would be inconsistent with the main purpose of the equation to estimate time-varying and forward-looking expected returns.

This is the same conclusion as in the previous section that imposing stock fixed effects in (33) for more than one value of the slope implies constant excess RN variances and constant excess expected returns. It points to an embedded tension between the assumption of constant stock-specific intercepts and the implicit promise of the formula to allow unconstrained model-free estimation of time-varying expected returns. The assumption of stock-fixed effect allows stock-specific time-variation in expected return to operate only through RN variance and restricts time-variation in the remaining terms to be common across stocks. The implications of these conditions can be best understood by explicitly linking SDF \( m \) and stock returns.

In order to unpack the terms comprising \( \alpha_i(\theta) \) I use the approximations from Kadan and Tang (2017) to derive conditions under which \( \alpha_i(\theta) \) can be treated as a fixed effect. Before using these approximations it is helpful to rewrite (39) expanding the expressions for RN variances. Note that no-arbitrage condition implies that \( E_t^P\{m_{t+1}(R_{i,t+1} - R_{f,t+1})\} = 0 \) and I can express \( SVIX_{i,t}^2 \) as
follows:

\[ SVIX_{t,t}^2 \equiv E_t^Q \left\{ \left( \frac{R_{i,t+1} - R_{f,t+1}}{R_{f,t+1}} \right)^2 \right\} = \frac{1}{R_{f,t+1}} E_t^P \left\{ m_{t+1} \left( R_{i,t+1} - R_{f,t+1} \right)^2 \right\} \]

\[ = \frac{1}{R_{f,t+1}} \left[ E_t^P \left\{ m_{t+1} R_{i,t+1}^e \right\} E_t^P \left\{ R_{i,t+1}^e \right\} + Cov_t^P \left( m_{t+1} R_{i,t+1}^e, R_{i,t+1}^e \right) \right] \]

\[ = \frac{1}{R_{f,t+1}} Cov_t^P \left( m_{t+1} R_{i,t+1}^e, R_{i,t+1}^e \right) \]

(40)

where \( R_{i,t+1}^e \equiv R_{i,t+1} - R_{f,t+1} \) denotes the excess return over the risk free rate. Substituting into (39) produces:

\[ \alpha_i(\theta) = \frac{1}{R_{f,t+1}} \left( 1 - \frac{\theta}{2} \right) \left[ Cov_t^P \left( m_{t+1} R_{i,t+1}^e, R_{i,t+1}^e \right) - \sum_{j=1}^{N} w_{j,t} Cov_t^P \left( m_{t+1} R_{j,t+1}^e, R_{j,t+1}^e \right) \right] \]

\[ - \frac{1}{R_{f,t+1}} \left[ Cov_t^P \left( m_{t+1} R_{i,t+1}, R_{i,t+1} \right) - \sum_{j=1}^{N} w_{j,t} Cov_t^P \left( m_{t+1} R_{j,t+1}, R_{j,t+1} \right) \right] \]

(41)

Without additional assumptions about SDF \( m \), it is not possible to determine the implications of these conditions for individual stocks. To advance further, I rely on some results developed by Kadan and Tang (2017). They derive approximation to the covariance term in equation (11) which is used as the starting point for MW derivations.

Kadan and Tang assume a representative agent economy with a homogeneous utility function \( u(w) \) and the relative risk aversion \( RRA(w) \equiv -w \frac{u''(w)}{u'(w)} \). They also assume that the index portfolio is the market portfolio and therefore the stochastic discount factor can be represented as:

\[ m_{t+1} = \frac{1}{\lambda_t} u' \left( R_{I,t+1} \right), \text{ where } \lambda_t = R_{f,t+1} E_t^P \{ u'(R_{I,t+1}) \}. \]

(42)

Under these assumptions (see Appendix B), they show that \( P \)-covariances in (11) can be approximated as follows:

\[ Cov_t^P \left( m_{t+1} R_{i,t+1}, R_{i,t+1} \right) \approx \frac{u'(E_t^P \{ R_{I,t+1} \})}{R_{f,t+1} E_t^P \{ u'(R_{I,t+1}) \}} \left[ Var_t^P \left( R_{i,t+1} \right) \right. \]

\[ \left. - RRA \left( E_t^P \{ R_{I,t+1} \} \right) Var_t^P \left( R_{I,t+1} \right) \beta_{i,t}^P \right] \]

(43)

where \( \beta_{i,t}^P \) is the standard beta (under the \( P \)-measure) from the linear projection of return \( i \) on the index portfolio return. I extend this approximation (in Appendix B) to another covariance term for
excess returns of the type appearing in (41):

\[
\text{Cov}_t^P(m_{t+1}R_{i,t+1}^e, R_{i,t+1}^e) \approx \frac{u'(E_t^P\{R_{I,t+1}\})}{R_{f,t+1}E_t^P\{u'(R_{I,t})\}} \left[ \text{Var}_t^P(R_{i,t+1}) - \text{RRA}(E_t^P\{R_{I,t+1}\}) \text{Var}_t^P(R_{I,t+1}) \frac{E_t^P\{R_{f,t+1}^e\}}{E_t^P\{R_{I,t+1}^e\}} \beta_{i,t}^P \right]
\]

(44)

where, as before, superscript “e” denotes the excess return over the risk free return. Using the approximation (43) I can express \( \alpha_{i,t} \) as follows:

\[
\alpha_{i,t} \approx -\frac{u'(E_t^P\{R_{I,t}\})}{R_{f,t+1}E_t^P\{u'(R_{I,t})\}} \left[ \text{Var}_t^P(R_{i,t+1}) - \sum_{j=1}^N w_{j,t} \text{Var}_t^P(R_{j,t+1}) \right. \\
\left. - \text{RRA}(E_t^P\{R_{I,t}\}) \text{Var}_t^P(R_{I,t+1}) \frac{E_t^P\{R_{f,t+1}^e\}}{E_t^P\{R_{I,t+1}^e\}} \beta_{i,t}^P \right] \tag{45}
\]

where I used the fact that the weighted average P-beta of the index stocks is equal to one. Similarly, using (44), I can express the remaining terms of (41):

\[
\left(1 - \frac{\theta}{2}\right)(\text{SVIX}_{i,t}^2 - \text{SVIX}_t^2) \approx \frac{(1 - \frac{\theta}{2}) u'(E_t^P\{R_{I,t}\})}{R_{f,t+1}E_t^P\{u'(R_{I,t})\}} \left[ \text{Var}_t^P(R_{i,t+1}) - \sum_{j=1}^N w_{j,t} \text{Var}_t^P(R_{j,t+1}) \right. \\
\left. - \text{RRA}(E_t^P\{R_{I,t}\}) \text{Var}_t^P(R_{I,t+1}) \frac{E_t^P\{R_{f,t+1}^e\}}{E_t^P\{R_{I,t+1}^e\}} \beta_{i,t}^P \right] \tag{46}
\]

Substituting (45) and (46) into (41) and collecting terms I obtain:

\[
\alpha_i(\theta) \approx -\frac{u'(E_t^P\{R_{I,t}\})}{R_{f,t+1}E_t^P\{u'(R_{I,t})\}} \left[ \frac{\theta}{2} \left( \text{Var}_t^P(R_{i,t+1}) - \sum_{j=1}^N w_{j,t} \text{Var}_t^P(R_{j,t+1}) \right) \right. \\
\left. - \text{RRA}(E_t^P\{R_{I,t}\}) \text{Var}_t^P(R_{I,t+1}) \frac{E_t^P\{R_{f,t+1}^e\}}{E_t^P\{R_{I,t+1}^e\}} \beta_{i,t}^P \right] + \frac{\theta}{2} E_t^P\{R_{I,t+1}\} + \left(1 - \frac{\theta}{2}\right) R_{f,t+1} \tag{47}
\]

Fixed effects assumption requires the expression in (47) to be constant over time. This implies that there should be an approximate linear relation between excess conditional variance and conditional beta with time-varying slope and intercept as functions of \( \theta \) and \( \alpha_i(\theta) \) as follows:

\[
\text{Var}_t^P(R_{i,t+1}) - \overline{\text{Var}}_t^P \approx (\beta_{i,t}^P - 1) s_i(\theta) + a_{i,t}(\theta) \tag{48}
\]

where \( \overline{\text{Var}}_t^P \equiv \sum_{j=1}^N w_{j,t} \text{Var}_t^P(R_{j,t+1}) \),

\[
s_i(\theta) \equiv \text{RRA}(E_t^P\{R_{I,t+1}\}) \text{Var}_t^P(R_{I,t+1}) \left(1 + \left(\frac{2}{\theta} - 1\right) \frac{R_{f,t+1}}{E_t^P\{R_{I,t+1}\}}\right),
\]

\[
a_{i,t}(\theta) \equiv -\frac{2\alpha_i(\theta)}{\theta} \times \frac{R_{f,t+1}^2 E_t^P\{u'(R_{I,t+1})\}}{u'(E_t^P\{R_{I,t+1}\})} = -\frac{2\alpha_i(\theta)}{\theta} f_t.
\]

19
Figure 2: Sufficient conditions for fixed effects
Given common slope $\theta$ in the expected returns equation in the left panel, stocks must satisfy linear constraints on relation between excess conditional variance and beta (in excess of one) shown in the right panel. The constraints’ intercepts and slope depend on the choice of $\theta$ and are not free parameters and the slope must be identical across stocks at any given time while the intercepts are driven by a common factor $f_t$ which is a function of market variance.

Slope coefficient $s_t$ on excess beta must be common for all stocks at any given time while the intercept is multiplicatively separable in stock-specific coefficient and a common factor $f_t$. Figure 2 illustrates the constraint. The left panel shows conditional expected excess return over the index as a linear function of RN variance for some $\theta$. The right panel shows that conditional excess $P$-variance must be linear in conditional excess $P$-beta.

Condition (48) is a $P$-measure counterpart to the two auxiliary assumptions in MW (2018): variance linearization in $Q$-beta in (22) and additive separability of the $Q$-residual variances in (26). When stated in terms of RN moments these conditions are not tied to the observable variables while constraint (48) helps to clarify their roles. The former assumption manifests through the linearity (in $P$-beta) of the systematic portion of excess $P$-variance $(\beta_{i,t}^P - 1)s_t(\theta)$ and the latter is reflected in the multiplicative separability of $a_{i,t}(\theta)$, the idiosyncratic component of excess $P$-variance. While linearization in beta is not required to derive projection equation (32), it is required for stock-specific constant intercepts.

On introspection, it appears to be a very demanding constraint, unlikely to hold generically for stocks with low-to-moderate idiosyncratic risk as conjectured in MW (2018). There must be “precise amount” of idiosyncratic risk so that stock return variances approximately linearize in
$P$-betas at specific common value which depends on $\theta$ and risk aversion. Moreover, idiosyncratic excess variances must be driven by a common factor $f_t$, which turns out to depend on market variance. These implications are shown in more detail next.

2.3.1 Some implications of sufficient conditions

The cross-sectional constraint in (48) requires linearization of excess variances in $P$-betas and it is not guaranteed to have a good fit, in general, because linearization location (slope) depends on $\theta$ and unobservable preference parameters. To see this, consider variance decomposition from the linear projection of return on the index stated in the $P$-moments (analogous to $Q$-measure decomposition in (20)):

$$\text{Var}^P_t (R_{i,t+1}) = (\beta^P_{i,t})^2 \text{Var}^P_t (R_{I,t+1}) + \text{Var}^P_t (\varepsilon_{i,t+1})$$

where $\text{Var}^P_t (\varepsilon_{i,t+1})$ is the idiosyncratic residual $P$-variance. Then, linearize $P$-variances in $P$-beta near $\beta^P_{i,t} = \frac{1}{\theta^P}$ to obtain approximately linear $P$-variance decomposition (recall analogous linearization in $Q$-beta in (22) near $\frac{1}{\theta}$):

$$\text{Var}^P_t (R_{i,t+1}) \approx \left( \frac{2}{\theta^P} \beta^P_{i,t} - \frac{1}{(\theta^P)^2} \right) \text{Var}^P_t (R_{I,t+1}) + \text{Var}^P_t (\varepsilon_{i,t+1}),$$

where linearization error is given by $(\beta^P_{i,t} - \frac{1}{\theta^P})^2 \text{Var}^P_t (R_{I,t+1})$. Using (50) together with constraint $\sum_{i=1}^N w_{i,t} \beta^P_{i,t} = 1$, excess $P$-variance can be written as follows:

$$\text{Var}^P_t (R_{i,t+1}) - \text{Var}^P_t \approx \frac{2}{\theta^P} (\beta^P_{i,t} - 1) \text{Var}^P_t (R_{I,t+1})$$

$$+ \left( \text{Var}^P_t [\varepsilon_{i,t+1}] - \sum_{j=1}^N w_{i,t} \text{Var}^P_t [\varepsilon_{j,t+1}] \right)$$

Imposing linearization constraint (48) on the systematic portion of excess variance, i.e. the first term on the right hand side of (51), implies:

$$s_t(\theta) \approx \frac{2}{\theta^P} \text{Var}^P_t (R_{I,t+1})$$

Approximating in $s_t(\theta)$ by setting $R_{f,t+1}/E_t^P \{R_{I,t+1}\} \approx 1$ gives:

$$\frac{1}{\theta^P} \approx \frac{1}{\theta} RRA_t (E_t^P \{R_{I,t+1}\}) \equiv \frac{RRA_t}{\theta}$$

Thus, the corresponding linearization location of systematic variance in $P$-beta is a multiple of $Q$-beta linearization location $\frac{1}{\theta}$ by the relative risk aversion. If the $Q$-variances are linearized near
\( \theta = 1 \), as chosen in MW, then sufficient conditions require \( P \)-variances to linearize well near \( RRA_t \) which, in general, is not assured and the approximation error is potentially large. The linearization error omitted in the excess variances equation (51) is given by:

\[
\text{Lin. Error}_t = \text{Var}_t^P(R_{I,t+1}) \left[ \left( \beta_{i,t}^P - \frac{1}{\theta^P} \right)^2 - \sum_{j=1}^N w_{j,t} \left( \beta_{j,t}^P - \frac{1}{\theta^P} \right)^2 \right] = (54)
\]

The error may be very significant for \( \theta = 1 \) if \( RRA_t \gg \beta_{i,t}^P \) for all stocks. For the majority of stocks \( \beta_{i,t}^P < 2 \), while asset pricing models typically assume \( RRA_t \gg 2 \) in order to generate reasonable equity premium. Thus, for \( \theta = 1 \), linearization requirement of constraint (48) does not appear to hold generically simply by requiring stocks to have low-to-moderate idiosyncratic risk and hinges on the assumption that risk aversion is low, near the log-utility case. For potentially smaller approximation error, one could set \( \theta = RRA_t \) at average value (or by postulating CRRA utility with some constant risk aversion). However, the risk aversion is not observable and it may change over time, which implies deterioration of sufficient conditions in periods when risk aversion differs substantially from the average. I do not examine further the cross-sectional linear constraints in (48) because they depend on unobservable parameters and therefore are challenging to verify empirically. Instead, I focus on the implications for time-variation of return variances because they are independent of unobservable preference parameters and can be tested.

To explore the implications of constraints (48) for time-variation of return variances, I begin by providing intuition for factor \( f_t \) driving the idiosyncratic components of the excess variances. Taking Taylor expansion of \( u'(R_{I,t+1}) \) evaluated at \( E_t^P \{ R_{I,t+1} \} \) up to the second order and then taking expectation gives:

\[
f_t \approx \frac{R_{f,t+1}^2}{u'(E_t^P \{ R_{I,t+1} \})} \left( u'(E_t^P \{ R_{I,t+1} \}) + \frac{1}{2} u'''(E_t^P \{ R_{I,t+1} \}) \text{Var}_t^P(R_{I,t+1}) \right)
\]

Using the definitions of absolute prudence for expected utility \( u(w) \), \( APR(w) \equiv -u'''(w)/u''(w) \) and absolute risk aversion \( ARA(w) \equiv -u''(w)/u'(w) \) the above can written as:

\[
f_t \approx R_{f,t+1}^2 \left( 1 + \frac{1}{2} ARA(E_t^P \{ R_{I,t+1} \}) APR(E_t^P \{ R_{I,t+1} \}) \text{Var}_t^P(R_{I,t+1}) \right)
\]

Therefore, idiosyncratic excess variance, captured by the term \( a_{i,t}(\theta) \), must be a function of the market variance. This implies from (48) that systematic and idiosyncratic components of the
excess variance are not independent as both are functions of the conditional market variance with the coefficients determined by preferences of the representative agent.

The constraint on one-factor structure in idiosyncratic components of excess $P$-variances is a counterpart of additive separability of the residual variances assumed under the $Q$-measure in (26). One implication of this assumption is that differences of the RN variance across times $t$ and $t'$ are constant for any two stocks $i$ and $j$ with equal conditional $Q$-betas with respect to the inverse SDF return. The betas may be different across time but must be equal across stocks at each time. This can be seen by differencing equations (20) at times $t$ and $t'$ for such stocks:

$$SVIX_{i,t}^2 - SVIX_{j,t}^2 = \phi_i - \phi_j = SVIX_{i,t'}^2 - SVIX_{j,t'}^2,$$

(57)

where $\phi_i$ and $\phi_j$ are stock-specific constant components of the residuals introduced in the additive decomposition assumption (26). Condition (57) requires only additive separability of the residual variances, i.e. it holds without linearization and can be obtained by differencing equation (20) prior to linearization. Therefore, this condition is independent from the slope choice (linearization location). Stated under the $Q$-measure such an assumption is not immediately transparent or testable because $Q$-betas are not observable. However, a similar condition independent of the slope and expressed in terms of $P$-variances is implied by (48) and it can be used to gain additional insights about the structure imposed on time-variation of risk.

Consider differencing equation (48) for a pair of stocks ($i$ and $j$) with equal conditional $P$-betas:

$$Var_P^P(R_{i,t+1}) - Var_P^P(R_{j,t+1}) = a_{i,t}(\theta) - a_{j,t}(\theta) = -\frac{2f_t}{\theta} [\alpha_i(\theta) - \alpha_j(\theta)]$$

(58)

Since $P$-betas are assumed to be the same for $i$ and $j$, the same equation applies to the differences of idiosyncratic components of their conditional variances. This means that, for groups of stocks with equal conditional beta, the differences in idiosyncratic risk of any two stocks within groups must follow one-factor structure with factor $f_t$, i.e. such differences must be perfectly correlated over time for arbitrary pairs of such stocks chosen from any group. Of course, the constituents of the groups may change as their conditional betas vary, but whenever two pairs of stocks ($i,j$) and ($k,l$) have the same conditional betas (within pairs only) at two times $t$ and $t'$, then the ratio of the
differences in variance between the pairs must be constant across two times: 

\[
\frac{\text{Var}_t^P(R_{i,t+1}) - \text{Var}_t^P(R_{j,t+1})}{\text{Var}_t^P(R_{k,t+1}) - \text{Var}_t^P(R_{l,t+1})} = \frac{\text{Var}_{t'}^P(R_{i,t'+1}) - \text{Var}_{t'}^P(R_{j,t'+1})}{\text{Var}_{t'}^P(R_{k,t'+1}) - \text{Var}_{t'}^P(R_{l,t'+1})}.
\]

This constraint is independent of the slope coefficient \( \theta \) chosen in the expected returns equations which is useful for empirical tests in the next subsection.

Asset pricing models typically do not constrain properties of idiosyncratic variance because it does not affect prices or expected returns. In this case, however, the requirement that expected return is expressed as a linear function of RN variance imposes intricate structure linking idiosyncratic and systematic variances in the cross-section and binding their time variation. As the empirical tests of constant intercepts for portfolios suggests, such a tight restriction does not appear to hold in the data for \( \theta = 1 \). Moreover, irrespective of the value of the slope \( \theta \), the idiosyncratic components of excess variances must follow one-factor structure so that variance differences of equal beta stocks satisfy constraint (59). Empirical tests of this restriction, presented in the next subsection, do not support this condition, casting doubt on general feasibility of constant intercept assumption for any value of the slope parameter.

I conclude the analysis of sufficient conditions with a useful corollary which suggests upper or lower bound on the expected return for certain stocks. These bounds are novel and hold regardless of the risk aversion, which distinguishes them from the lower bound analysis in Kadan and Tang (2017). Turn to equation (45) and note that the term multiplying beta deviation from one and the term multiplying on the outside of square brackets are both negative, assuming standard increasing utility of the representative agent. It follows that, for stocks with \( P \)-variance above the average and \( \beta_{i,t}^P \) below one, the equation implies \( \alpha_{i,t} < 0 \) and the opposite is true when \( P \)-variance is below the average and \( P \)-beta is above one. For these two cases, respectively, excess RN variances provide upper or lower bound on the expected excess returns of a stock over the index as follows:

\[
\text{Var}_t^P(R_{i,t+1}) > (<) \sum_{j=1}^N w_{j,t} \text{Var}_t^P(R_{j,t+1}) \quad \text{and} \quad \beta_{i,t}^P < (>) 1
\]

\[
\Rightarrow \alpha_{i,t} < (>) 0
\]

\[
\Rightarrow \frac{E_t^P\{R_{i,t+1}\} - E_t^P\{R_{I,t+1}\}}{R_{f,t+1}} < (>) \left( SVIX^2_{i,t} - SVIX^2_t \right)
\]

\[
(60)
\]

\footnotetext{Note that betas need to be the same only within pairs \((i,j)\) and \((k,l)\) but they can be different across time and pairs, i.e. condition (59) holds as long as \( \beta_{i,t}^P = \beta_{j,t}^P; \beta_{k,t}^P = \beta_{l,t}^P; \beta_{i,t'}^P = \beta_{j,t'}^P; \beta_{k,t'}^P = \beta_{l,t'}^P. \)
Empirically, beta and variance are positively cross-sectionally correlated but the correlation is not perfect, so the bounds may be useful for some stocks. For stocks which have both variance and beta either above or below their weighed averages, the direction of the bound is ambiguous. In these cases some further progress can be made under the assumptions about representative risk aversion as discussed by Kadan and Tang (2017). Empirical analysis of these novel bounds is outside the scope of the present paper but it presents interesting venue for future research.

2.4 Testing one-factor constraint on differences of $P$-variances

The constraint (59) can be tested empirically under the assumptions that historical betas and variances can be used as proxies for the respective conditional moments and that equal betas with respect to a broad market index imply equal betas with respect to the true market portfolio (beta co-monotonicity). Unlike in some of the traditional asset pricing tests, the value of beta is not important here, so long as stocks with equal beta can be identified, because the test itself is based only on variances. To construct the test, I use the sample of returns of the S&P 500 index constituents and the return on the index from January 1996 to June 2018. I obtain daily returns from the Center for Research in Security Prices (CRSP) database.

The proxy for conditional variance is computed as follows. Daily return variances are computed using trailing 60-trading-days window for each trading day of the month. The average of the daily variances within a calendar month $t$ is used as a proxy for conditional variance $Var^P_t(R_{i,t+1})$ in month $t$ and the test is done at monthly frequency. In order to identify stocks suitable for the test, I construct conditional betas with respect to the S&P 500 index return using similar methodology. For each trading day, daily betas are estimated from a linear regression on the index return using trailing 60-trading-days window and the average of betas within a calendar month $\hat{\beta}^P_{i,t}$ is used as a proxy for conditional beta in month $t$.

I construct a collection of pairs of stocks $(i,j)$ which in any two months $t$ and $t'$ have betas

---

8Kadan and Tang (2017) show that, for sufficiently high risk aversion, the covariance term approximated in equation (43) will be negative and in this case equation (11) can be used to construct a lower bound on expected stock return.

9To obtain monthly variance daily variance can be scaled by a factor of 30 but it is not relevant for the test since scaling is identical for all stocks and I use daily variances in all calculations and scale them by $10^4$ to convert from decimal base into percentage base.
within \( e = 0.01 \) of each other relative to their mid-point so that:

\[
\left| \hat{\beta}^P_{i,t} - \hat{\beta}^P_{j,t} \right| \leq \frac{e}{2}, \quad \text{and} \quad \left| \hat{\beta}^P_{i,t'} - \hat{\beta}^P_{j,t'} \right| \leq \frac{e}{2}.
\]

These occurrences are recorded without regard to the order of stocks, i.e. \((i, j)\) and \((j, i)\) satisfying (61) are considered as one pair, and constructed for all available ordered time pairs \(t < t'\). Thus, this set includes all possible stock-time pairs \((i, j, t, t')\) satisfying (61). Stocks must have equal betas in two or more months to be included and for pairs such that beta equality occurs in more than two months all combinations of ordered months are created. That is, if two stocks have near-equal betas in three months \(t_1 < t_2 < t_3\), then three stock-time pairs can be constructed \((i, j, t_1, t_2), (i, j, t_1, t_3), (i, j, t_2, t_3)\). Finally, for each ordered pair of months \(t < t'\), I take all distinct pairs of stocks which have equal betas at \(t\) and \(t'\) and construct all possible 2-pair combinations or quadruples \(((i, j), (k, l), t, t')\), without regard to the order of the pairs so that \(((i, j), (k, l), t, t')\) is the same as \(((k, l), (i, j), t, t')\). This ensures no repetition so that statistical significance is not inflated. These quadruples constitute exhaustive list of possible combinations of stock pairs and times when beta equality occurs within each pair at both times. This large set is used to construct monthly sets of stocks for the tests as described next.

Given a horizon \(h\), for each month \(t\) the set of stocks for the test is constructed for all \(t'\) such that \(t < t' \leq t + h\). That is, for each month \(t\) the set includes all quadruples of stocks which at any month \(t'\) up to \(t + h\) have nearly-equal betas within pairs of the quadruple at \(t\) and again at some future time \(t'\). The tests are reported for three horizons \(h = 3\) (up to one quarter ahead), \(h = 6\) (up to half a year ahead), and for \(h = 12\) (up to one year ahead). This construction of sets is nested for each month, so that longer horizon sets for a given month include all shorter horizons sets for this month, which creates progressively higher number of observations for each month.\(^{10}\) To avoid potential outliers created by small values of the denominators in equation (59) I test an equivalent equation by cross-multiplying the two sides to avoid division operation.

I test the constraint by estimating a linear regression for each month \(t\) for all quadruples of stocks included in the set for a given horizon:

\[
(\Delta_{i,j} \sigma_t^2 \times \Delta_{k,l} \sigma_{t'}^2) = a + b (\Delta_{i,j} \sigma_t^2 \times \Delta_{k,l} \sigma_{t'}^2) + e_{((i,j),(k,l),t,t')} \quad (62)
\]

\(^{10}\)However, across months the sets are quite different because equality of beta in months \(t\) and then again in some \(t < t' \leq t + h\) does not imply equality in months \(t + 1\) and \(t + 1 < t' \leq t + 1 + h\) for the same stocks.
where $\Delta_{i,j} \sigma^2_t$ is the difference in variances of stocks $i$ and $j$ at time $t$, i.e. the numerators and denominators of the ratios in (59). If the constraint (59) holds, the coefficients should satisfy $a = 0$ and $b = 1$. Figures 6, 7, and 8 report, respectively, the results for horizon of three, six, and twelve months. The top row panels show the estimates, the middle row shows the t-statistics for separate tests of $a = 0$ and $b = 1$, and the bottom row shows the $p$-values for the joint restriction $F$-test $(a = 0, b = 1)$ as well as adjusted $R^2$.11 The t-statistics are scaled as log of one-half of the absolute value of the t-statistic so that values above zero approximately correspond to at least 5% significance level.

The intercepts $a$ tend to be larger at the beginning of the sample and become negligible in the second half of the sample. The intercepts $a$ are significantly different from zero at least at 5% confidence in 28, 38, and 41 months out of 266 months for the quarterly, half-annual, and annual horizons, respectively. The slope coefficients $b$ tend to be below one and, in the vast majority of cases, are statistically significantly different from one at least at the 5% level: in 258, 263, and 265 months out of 266, respectively for quarterly, half-annual, and annual horizon. Thus, while the intercepts $a$ appear to be insignificant in many cases and particularly small in the second half of the sample, the restriction $b = 1$ is rejected in the vast majority of months.

The tests for the joint null hypothesis $a = 0$ and $b = 1$, shown in the bottom-left panels of the figures, are not rejected at least at 5% confidence only in 8, 1, and 1 months out of 266 for quarterly, half-annual, and annual horizon, respectively. The progression towards uniform rejection with longer horizons suggests that most of the incidences of failure to reject constraint are driven by the low power of the tests in months with small number of stock pairs with equal beta, concentrated in the early part of the sample. When the tests are conducted on larger sets of securities, by increasing horizon or in later part of the sample, the constraints tend to be overwhelmingly rejected.12 Overall, the results across all horizons are inconsistent with constant intercept assumption for any slope $\theta$, and put in doubt general feasibility of calculating expected returns solely from RN variances.

11The median number (across all months) of distinct quadruples in the sets used for the regression is 3022, 4350, and 6362 for quarterly, half-annual, and annual horizon, respectively. That is, a typical regression is based on quadruples constructed from all unique combinations of about 78, 94, and 113 pairs of stocks which have equal beta at $t$ and again at some other month within horizon $h$ from $t$.

12This is further confirmed in unreported results by including in the test all available observations without constraining the horizon, i.e. the regression is estimated in each month $t$ by including all quadruples with equal betas for any available $t' > t$. The slope coefficients $b$ decline significantly towards zero compared with reported horizons and tests of constraints $b = 1$ and joint constraint $a = 0$ and $b = 1$ are strongly rejected in all months.
2.5 Alternative identification with common intercept

There exists an alternative identification for expected returns equations, but it comes with additional challenges. In some ways, it is harder to implement empirically than the common slope estimation but it is considered here for completeness and because it helps to understand why the forecasting regressions estimated in MW with common zero intercept may improve upon the other predictors.

Suppose that intercepts $c_{i,t}(\theta)$ in (31) are set to a constant $c$. Then, as long as $SVIX_{i,t}^2 \neq SVIX_t^2$, there always exist time-varying functions $\theta_{i,t}(c)$ for the projections with common intercept for such stocks:

\begin{equation}
\alpha_{i,t} = c + \left( \frac{\theta_{i,t}(c)}{2} - 1 \right) \left( SVIX_{i,t}^2 - \overline{SVIX}_t^2 \right) \tag{63}
\end{equation}

\begin{equation}
\frac{E_{i}^P \{ R_{i,t+1} \} - E_{i}^P \{ R_{I,t+1} \}}{R_{f,t+1}} = c + \frac{\theta_{i,t}(c)}{2} \left( SVIX_{i,t}^2 - \overline{SVIX}_t^2 \right) \tag{64}
\end{equation}

This is the case shown on Figure 3 for two stocks for $c = 0$. When the intercept is set to a common value, each stock has a unique slope for its projection onto the blue plane of expected returns and RN variances. While doing so identifies the slope, each line applies only to a subset of stocks in the gray plane lying on that line (or close to it) as opposed to having a common but indeterminate slope as was previously shown in Figure 1 on the right panel. In general, the slopes are time-varying so if they are assumed to be constant for estimation, then the constraint analogous to (48) but with common intercept and stock-specific slopes must also hold. Thus, over time, stocks must move only along the concentric red lines shown on Figure 3, then the projection equations will have constant slope and common intercept. Similarly to the previous case, stock-specific constant slopes can exist at most for one value of the common constant intercept or else differencing equation (64) implies constant excess RN variances and constant expected returns.

When slopes $\theta_i(c)$ are assumed to be constant over time for some value of the intercept $c$, the challenge of this identification strategy is that it requires figuring out which stocks are on (or near) a given equation line so that one can run a regression to estimate slope $\theta_i$ for this group. To identify such a group with a common slope one has to find stocks which satisfy linear relations as required by the constraint in (48) (with straightforward variables swap as $\alpha_i(\theta) = c$ and $\theta = \theta_i(c)$). That is, if the conditional variance and beta in this constraint are estimated for each stock, one has to find a group of stocks which is on (or near) a line for some arbitrary value $s_t = s_i$ as shown on the
Figure 3: Identification with common intercept

Expected returns equations can be identified by assuming common intercept and different slopes for subsets of stocks. The figure shows two stocks as blue dots on the gray plane connected to their projections on the blue plane by the dotted lines. The red lines in the gray plane go through the origin and the stocks. The projections of the red lines onto the blue plane are marked as black lines which go through the origin and the stocks’ projection points. Forcing the intercept to zero for all stocks uniquely identifies the slope of the projections but their slopes are stock-specific and generally time-varying.

right panel of Figure 4 summarizing this case. For every \( \theta_i \) there exist some stocks such that their expected returns are determined as linear function of the excess RN variance as shown on the left panel. For \( \theta_i \) to be stock-specific constant, these stocks must also have a linear relation between excess variance and beta with some slope \( s_t(\theta_i) = s_i \) and zero intercept according to (48) as shown on the right panel. Assuming groups of stocks satisfying a given linear relation in (48) are identified in the right panel, the expected returns equation can be estimated cross-sectionally within each group to find the corresponding \( \theta_i \) in the left panel.\(^{13}\)

Note that one can always find a number \( s_i \) such that (48) holds with zero intercept for a group

\(^{13}\) A caveat applies here because \( s_t(\theta) \) is generally time-varying and its mapping into a particular \( \theta \) is not constant. Thus, a group of stocks with the same slope of linear relation on the right panel may correspond to different values of \( \theta \) at different times because unobservable conditional quantities in \( s_t(\theta) \) change. In other words, if we find a group of stocks with a given slope in the excess variance-beta space (right panel) such that \( s_t(\theta) = s \) then \( \theta_t(s) \equiv s_t^{-1}(s) \neq s_{t+1}^{-1}(s) \equiv \theta_{t+1}(s) \) in general. Therefore, application out-of-sample would require additional assumption that this mapping is sufficiently slowly moving so that \( \theta_t(s) \) can be used to forecast returns next period.
**Figure 4: Common zero intercept**

For a group of stocks such that their conditional excess $P$-variance and conditional beta are linearly related with some slope and zero intercept on the right panel, there is a linear relation between their expected returns and excess RN variances with some slope $\theta_i$ and zero intercept. If this group is identified, then $\theta_i$ can be estimated from the cross-sectional regression within the group. The slope $\theta_i$ is specific to the group and the group constituents may change over time.

The conditional expected excess return can be expressed as:

$$E_t\{R_{i,t+1} - R_{I,t+1}\} = \theta_i (\beta_{P_i,t} - 1)$$

and the conditional excess $P$-variance as:

$$\text{Var}_P t (R_{i,t+1}) - \text{Var}_P t$$

of some stocks with approximately equal ratios of excess variance to excess beta. That is, for any line going through the origin in the right panel in Figure 4 there will always be some stocks near the line if the cross-section is large. The exact value of the slope $s_i$ does not matter at this stage: as long as stocks have the the same ratio of conditional excess variance to beta, the prediction of the model is that the same stocks must have expected return which is linear in the excess RN variance with some unknown slope $\theta_i$, as shown in the left panel. In a sufficiently rich cross section, stocks satisfying these type of constraints for some range of values $s_i$ can always be found by dividing the cross section into small groups with approximately equal ratio of excess variance to excess beta.\(^{14}\)

In other words, the cross section of stocks can always be divided into small sectors around lines with some slopes $s_i$ on Figure 4 in the right panel.\(^{15}\)

---

\(^{14}\)For stocks with beta equal to one this ratio is not defined and they constitute a special group. For them, conditional excess variance must be zero according to (48). Additionally, based on (46), their excess $SVIX_i^2$ is also equal to zero and therefore their expected return must be equal to that of the index. Since such stocks are located at the origin on both panels of figure 4 they can be combined for estimation of $\theta_i$ along arbitrary line.

\(^{15}\)There is an additional quantitative restriction $s_i(\theta_i) = s_i$ which can be tested using the slopes $\theta_i$ and $s_i$ estimated in the process. Note from (48) that for two different stocks (groups) $i$ and $j$ the difference $s_i(\theta_i) - s_i(\theta_j) = s_i - s_j$ depends on $\theta_i$, $\theta_j$ and a common multiplicative term. For any four stocks the ratio of the differences $s_i(\theta_i) - s_i(\theta_j)$ for two distinct pairs of stocks only depends on their respective $\theta$'s, a restriction which can be tested.
Empirical application of this method faces two principal problems. First, there is no guarantee that the same constraint line maps into the same slope over time (see fn. 13). Second, there is no single value of $\theta$ which applies to all stocks in the cross-section. Such a method also requires precisely the inference about conditional physical distribution moments which the original formula attempts to circumvent. Even though the final equation involves only RN variances as determinants of the expected returns, finding the relevant constituents requires estimation of conditional $P$-moments.

For the forecasting regression, MW use common zero intercept and slope of $\frac{1}{2}$. In general, equation (64) implies that this is a mis-specified regression, but for some stocks it may be correctly specified because they approximately satisfy constraint (48) with zero intercept and $\theta = 1$. However, the subsample of such stocks is unknown and its constituents likely change over time. Nonetheless, MW (2018) show that RN variances are somewhat better predictors of the expected returns compared with several alternative variables. It is possible that forward looking information from RN variances outweighs specification error. This points to a potential benefit from refining this methodology by estimating different slope coefficients for groups with common ratio of excess variance to excess beta. This requires estimates of conditional $P$-variance and $P$-beta from historical data under the additional assumption that stock-specific slopes are constant over time. Perhaps a better alternative is to use equation (29). This equation has a known coefficient of unity on excess RN variances for all stocks instead of the multiplicity of unknown coefficients approximately applicable to subsets of stocks. Thus, it avoids specification error at the expense of the additional assumptions and estimation. I leave formal empirical analysis of these interesting issues for future research.

3 Concluding remarks

In this paper I investigate theoretically the link between expected return and RN variance. I show that linear equations connecting these two moments are not identified and form a continuum with infinitely many slope-intercept combinations for each stock. This indeterminacy is a generic property arising from the projection of the well-identified two-factor equation onto one-factor equation. Identification requires setting common slope or intercept at some arbitrary value while imposing stock-specific constant constraint on the other coefficient. Empirical test of constant intercepts assumption for portfolios is strongly rejected for slope of one-half. The conditions for RN vari-
ance to be a sufficient statistic for expected returns turn out to be very restrictive for time- and cross-sectional variation of total and idiosyncratic risk. The test for one-factor structure of variance differences for individual stocks is rejected and suggests that it may be infeasible to suppress time-variation of the intercepts for any value of the slope parameter. The alternative identification with common intercept and multiple slopes for sub-sets of stocks requires estimation of the conditional physical distribution moments as well as the additional constraints for the assumption that slopes are stock specific constants.

While these results do not support the estimation of expected returns solely from RN variances, they do point to the benefits of integrating $Q$-side moments with traditional methods. The equations used in my analysis are based on Martin (2017) and rely only on basic no-arbitrage conditions. They combine information from RN variances and conditional $P$- covariances and have well identified coefficients. With additional assumptions about SDF and traditional inference from historic data such equations may provide more accurate estimation of expected returns than the methods which exclusively rely either on physical or RN distribution moments. I also show that, for some stocks, RN variances can be used to construct upper or lower bounds on expected returns. These bounds are novel and hold irrespective of risk aversion in the underlying economy. The behavior of these bounds and the benefits of alternative estimation techniques combining RN and physical moments represent potentially interesting venues for future research.

Although the exposition here is initially motivated by MW’s methodology, the findings are general. Overall spirit of the results is consistent with the debate in the literature about the recovery theorem which refers to computing physical probability distribution from the RN probability distribution proposed by Ross (2015). Ross shows that under certain strong conditions one can construct physical transition probabilities solely from the RN counterparts. The constraint required in the recovery theorem implies strong restrictions on the SDF, as pointed out in Borovička, Hansen and Scheinkman (2016). The latter paper shows that recovery of the physical probabilities without assumptions about preferences is limited to the SDFs with a constant martingale component. My theoretical results highlight the strength of assumptions needed to connect the first and the second moments of returns under the RN and physical probability distributions while the empirical results suggest that these assumptions are not supported for stock returns.
References


Kadan, Ohad, and Xiaoxiao Tang, 2017, A bound on expected stock returns, working paper, University of Washington St. Louis.


Martin, Ian, and Christian Wagner, 2016, What is the expected return on a stock?, working paper London School of Economics.


Table 1: Summary statistics for the index portfolio $p = I$ and for quintile portfolios $p = \{1, 2, 3, 4, 5\}$ sorted on average RN variance during every two years. The returns, RN variances and intercepts $\hat{c}_{p,t}(\theta = 1)$ are constructed for maturities of $h = \{30, 91, 182, 365\}$ days. Portfolios are either value weighted (Panel A) with weights based on average market capitalization in a given year or equally weighted (Panel B). $N$ is the average number of stocks in the portfolio and $T$ is the number of non-overlapping time periods available for each portfolio-maturity combination.

<p>| $h$ | $p$ | $R_{p,t}$ | $\text{SVIX}<em>{p,t}$ | $\hat{c}</em>{p,t}(\theta)$ |
|-----|-----|----------|----------------|----------------|----------|</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Mean</th>
<th>St.dev.</th>
<th>Mean</th>
<th>St.dev.</th>
<th>Mean</th>
<th>St.dev.</th>
<th>$N$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>I</td>
<td>1.012</td>
<td>0.046</td>
<td>0.009</td>
<td>0.007</td>
<td>479</td>
<td>264</td>
<td></td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>I</td>
<td>1.033</td>
<td>0.081</td>
<td>0.027</td>
<td>0.017</td>
<td>479</td>
<td>88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>182</td>
<td>I</td>
<td>1.062</td>
<td>0.105</td>
<td>0.049</td>
<td>0.028</td>
<td>477</td>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>365</td>
<td>I</td>
<td>1.130</td>
<td>0.175</td>
<td>0.101</td>
<td>0.055</td>
<td>471</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>1.011</td>
<td>0.033</td>
<td>0.004</td>
<td>0.003</td>
<td>0.002</td>
<td>0.029</td>
<td>98</td>
<td>264</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>1.012</td>
<td>0.045</td>
<td>0.006</td>
<td>0.004</td>
<td>0.002</td>
<td>0.018</td>
<td>99</td>
<td>264</td>
</tr>
<tr>
<td>30</td>
<td>3</td>
<td>1.013</td>
<td>0.053</td>
<td>0.008</td>
<td>0.006</td>
<td>0.002</td>
<td>0.017</td>
<td>97</td>
<td>264</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>1.013</td>
<td>0.067</td>
<td>0.012</td>
<td>0.010</td>
<td>-0.0002</td>
<td>0.028</td>
<td>94</td>
<td>264</td>
</tr>
<tr>
<td>30</td>
<td>5</td>
<td>1.012</td>
<td>0.098</td>
<td>0.024</td>
<td>0.022</td>
<td>-0.007</td>
<td>0.062</td>
<td>88</td>
<td>264</td>
</tr>
<tr>
<td>91</td>
<td>1</td>
<td>1.028</td>
<td>0.053</td>
<td>0.011</td>
<td>0.006</td>
<td>0.002</td>
<td>0.056</td>
<td>99</td>
<td>88</td>
</tr>
<tr>
<td>91</td>
<td>2</td>
<td>1.034</td>
<td>0.077</td>
<td>0.017</td>
<td>0.010</td>
<td>0.005</td>
<td>0.030</td>
<td>99</td>
<td>88</td>
</tr>
<tr>
<td>91</td>
<td>3</td>
<td>1.034</td>
<td>0.090</td>
<td>0.024</td>
<td>0.016</td>
<td>0.002</td>
<td>0.030</td>
<td>97</td>
<td>88</td>
</tr>
<tr>
<td>91</td>
<td>4</td>
<td>1.040</td>
<td>0.122</td>
<td>0.033</td>
<td>0.021</td>
<td>0.004</td>
<td>0.049</td>
<td>94</td>
<td>88</td>
</tr>
<tr>
<td>91</td>
<td>5</td>
<td>1.039</td>
<td>0.178</td>
<td>0.068</td>
<td>0.050</td>
<td>-0.014</td>
<td>0.111</td>
<td>88</td>
<td>88</td>
</tr>
<tr>
<td>182</td>
<td>1</td>
<td>1.052</td>
<td>0.078</td>
<td>0.021</td>
<td>0.010</td>
<td>0.004</td>
<td>0.076</td>
<td>98</td>
<td>44</td>
</tr>
<tr>
<td>182</td>
<td>2</td>
<td>1.065</td>
<td>0.102</td>
<td>0.032</td>
<td>0.018</td>
<td>0.012</td>
<td>0.041</td>
<td>99</td>
<td>44</td>
</tr>
<tr>
<td>182</td>
<td>3</td>
<td>1.062</td>
<td>0.125</td>
<td>0.043</td>
<td>0.024</td>
<td>0.003</td>
<td>0.048</td>
<td>96</td>
<td>44</td>
</tr>
<tr>
<td>182</td>
<td>4</td>
<td>1.072</td>
<td>0.158</td>
<td>0.065</td>
<td>0.038</td>
<td>0.002</td>
<td>0.069</td>
<td>94</td>
<td>44</td>
</tr>
<tr>
<td>182</td>
<td>5</td>
<td>1.070</td>
<td>0.209</td>
<td>0.119</td>
<td>0.070</td>
<td>-0.027</td>
<td>0.134</td>
<td>87</td>
<td>44</td>
</tr>
<tr>
<td>365</td>
<td>1</td>
<td>1.092</td>
<td>0.111</td>
<td>0.045</td>
<td>0.024</td>
<td>-0.008</td>
<td>0.114</td>
<td>98</td>
<td>22</td>
</tr>
<tr>
<td>365</td>
<td>2</td>
<td>1.132</td>
<td>0.174</td>
<td>0.068</td>
<td>0.039</td>
<td>0.019</td>
<td>0.069</td>
<td>98</td>
<td>22</td>
</tr>
<tr>
<td>365</td>
<td>3</td>
<td>1.129</td>
<td>0.197</td>
<td>0.085</td>
<td>0.046</td>
<td>0.008</td>
<td>0.057</td>
<td>95</td>
<td>22</td>
</tr>
<tr>
<td>365</td>
<td>4</td>
<td>1.164</td>
<td>0.263</td>
<td>0.136</td>
<td>0.084</td>
<td>0.016</td>
<td>0.103</td>
<td>92</td>
<td>22</td>
</tr>
<tr>
<td>365</td>
<td>5</td>
<td>1.183</td>
<td>0.397</td>
<td>0.234</td>
<td>0.129</td>
<td>-0.014</td>
<td>0.243</td>
<td>86</td>
<td>22</td>
</tr>
</tbody>
</table>

Panel A: Value weighted

Panel B: Equally weighted
Table 2: Estimates of the regression test (37) for equality of intercepts across time. The regressions are estimated using intercepts for quintile portfolios $p = \{1, 2, 3, 4, 5\}$ sorted on average RN variance during every two years. The intercepts are constructed for maturities of $h = \{30, 91, 182, 365\}$ days. Portfolios are either value weighted (Panel A) with weights based on average market capitalization in a given year or equally weighted (Panel B). For each time $t$ the regression includes intercepts within 4 observations from time $t$. The last column shows $p$-value of the joint $F$-test for $a = 0$ and $b = 1$.

<table>
<thead>
<tr>
<th>$h$</th>
<th>$p$</th>
<th>$a$</th>
<th>$a$ t-stat.</th>
<th>$b$</th>
<th>$b$ t-stat.</th>
<th>$R^2$</th>
<th>$N$</th>
<th>$P(&gt; F)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>1</td>
<td>0.002</td>
<td>2.734</td>
<td>0.049</td>
<td>1.590</td>
<td>0.002</td>
<td>1046</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>0.002</td>
<td>4.067</td>
<td>0.002</td>
<td>0.076</td>
<td>0.0000</td>
<td>1046</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>3</td>
<td>0.001</td>
<td>2.728</td>
<td>0.069</td>
<td>2.251</td>
<td>0.005</td>
<td>1046</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>-0.0004</td>
<td>-0.477</td>
<td>0.004</td>
<td>0.135</td>
<td>0.0000</td>
<td>1046</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>5</td>
<td>-0.007</td>
<td>-3.588</td>
<td>0.081</td>
<td>2.618</td>
<td>0.007</td>
<td>1046</td>
<td>0</td>
</tr>
<tr>
<td>91</td>
<td>1</td>
<td>0.003</td>
<td>1.064</td>
<td>-0.014</td>
<td>-0.261</td>
<td>0.0002</td>
<td>342</td>
<td>0</td>
</tr>
<tr>
<td>91</td>
<td>2</td>
<td>0.006</td>
<td>3.321</td>
<td>0.058</td>
<td>1.071</td>
<td>0.003</td>
<td>342</td>
<td>0</td>
</tr>
<tr>
<td>91</td>
<td>3</td>
<td>0.002</td>
<td>1.045</td>
<td>0.007</td>
<td>0.127</td>
<td>0.0000</td>
<td>342</td>
<td>0</td>
</tr>
<tr>
<td>91</td>
<td>4</td>
<td>0.004</td>
<td>1.367</td>
<td>-0.022</td>
<td>-0.403</td>
<td>0.0005</td>
<td>342</td>
<td>0</td>
</tr>
<tr>
<td>91</td>
<td>5</td>
<td>-0.015</td>
<td>-2.418</td>
<td>0.023</td>
<td>0.421</td>
<td>0.001</td>
<td>342</td>
<td>0</td>
</tr>
<tr>
<td>182</td>
<td>1</td>
<td>0.007</td>
<td>1.093</td>
<td>0.014</td>
<td>0.174</td>
<td>0.0002</td>
<td>166</td>
<td>0</td>
</tr>
<tr>
<td>182</td>
<td>2</td>
<td>0.013</td>
<td>4.069</td>
<td>-0.102</td>
<td>-1.308</td>
<td>0.010</td>
<td>166</td>
<td>0</td>
</tr>
<tr>
<td>182</td>
<td>3</td>
<td>0.003</td>
<td>0.708</td>
<td>-0.159</td>
<td>-2.083</td>
<td>0.026</td>
<td>166</td>
<td>0</td>
</tr>
<tr>
<td>182</td>
<td>4</td>
<td>0.001</td>
<td>0.106</td>
<td>-0.009</td>
<td>-0.116</td>
<td>0.0001</td>
<td>166</td>
<td>0</td>
</tr>
<tr>
<td>182</td>
<td>5</td>
<td>-0.029</td>
<td>-3.218</td>
<td>-0.056</td>
<td>-0.724</td>
<td>0.003</td>
<td>166</td>
<td>0</td>
</tr>
<tr>
<td>365</td>
<td>1</td>
<td>0.002</td>
<td>0.155</td>
<td>-0.084</td>
<td>-0.767</td>
<td>0.008</td>
<td>78</td>
<td>0</td>
</tr>
<tr>
<td>365</td>
<td>2</td>
<td>0.027</td>
<td>3.511</td>
<td>-0.100</td>
<td>-0.966</td>
<td>0.012</td>
<td>78</td>
<td>0</td>
</tr>
<tr>
<td>365</td>
<td>3</td>
<td>0.005</td>
<td>0.789</td>
<td>0.003</td>
<td>0.030</td>
<td>0.0000</td>
<td>78</td>
<td>0</td>
</tr>
<tr>
<td>365</td>
<td>4</td>
<td>0.007</td>
<td>0.593</td>
<td>-0.071</td>
<td>-0.642</td>
<td>0.005</td>
<td>78</td>
<td>0</td>
</tr>
<tr>
<td>365</td>
<td>5</td>
<td>-0.031</td>
<td>-1.158</td>
<td>-0.124</td>
<td>-1.157</td>
<td>0.017</td>
<td>78</td>
<td>0</td>
</tr>
</tbody>
</table>

Panel A: Value weighted

<table>
<thead>
<tr>
<th>$h$</th>
<th>$p$</th>
<th>$a$</th>
<th>$a$ t-stat.</th>
<th>$b$</th>
<th>$b$ t-stat.</th>
<th>$R^2$</th>
<th>$N$</th>
<th>$P(&gt; F)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>1</td>
<td>0.005</td>
<td>4.990</td>
<td>0.067</td>
<td>2.162</td>
<td>0.004</td>
<td>1046</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>0.005</td>
<td>8.133</td>
<td>0.029</td>
<td>0.947</td>
<td>0.001</td>
<td>1046</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>3</td>
<td>0.003</td>
<td>8.479</td>
<td>0.061</td>
<td>1.979</td>
<td>0.004</td>
<td>1046</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>0.002</td>
<td>4.880</td>
<td>0.005</td>
<td>0.178</td>
<td>0.0000</td>
<td>1046</td>
<td>0</td>
</tr>
<tr>
<td>30</td>
<td>5</td>
<td>-0.007</td>
<td>-3.940</td>
<td>0.036</td>
<td>1.147</td>
<td>0.001</td>
<td>1046</td>
<td>0</td>
</tr>
<tr>
<td>91</td>
<td>1</td>
<td>0.012</td>
<td>3.879</td>
<td>-0.022</td>
<td>-0.405</td>
<td>0.0005</td>
<td>342</td>
<td>0</td>
</tr>
<tr>
<td>91</td>
<td>2</td>
<td>0.015</td>
<td>7.680</td>
<td>0.006</td>
<td>0.104</td>
<td>0.0000</td>
<td>342</td>
<td>0</td>
</tr>
<tr>
<td>91</td>
<td>3</td>
<td>0.009</td>
<td>6.317</td>
<td>0.046</td>
<td>0.843</td>
<td>0.002</td>
<td>342</td>
<td>0</td>
</tr>
<tr>
<td>91</td>
<td>4</td>
<td>0.007</td>
<td>4.905</td>
<td>-0.059</td>
<td>-1.108</td>
<td>0.004</td>
<td>342</td>
<td>0</td>
</tr>
<tr>
<td>91</td>
<td>5</td>
<td>-0.016</td>
<td>-3.138</td>
<td>-0.043</td>
<td>-0.802</td>
<td>0.002</td>
<td>342</td>
<td>0</td>
</tr>
<tr>
<td>182</td>
<td>1</td>
<td>0.027</td>
<td>4.754</td>
<td>-0.059</td>
<td>-0.747</td>
<td>0.003</td>
<td>166</td>
<td>0</td>
</tr>
<tr>
<td>182</td>
<td>2</td>
<td>0.029</td>
<td>7.480</td>
<td>-0.037</td>
<td>-0.469</td>
<td>0.001</td>
<td>166</td>
<td>0</td>
</tr>
<tr>
<td>182</td>
<td>3</td>
<td>0.020</td>
<td>6.761</td>
<td>-0.041</td>
<td>-0.533</td>
<td>0.002</td>
<td>166</td>
<td>0</td>
</tr>
<tr>
<td>182</td>
<td>4</td>
<td>0.011</td>
<td>3.945</td>
<td>-0.004</td>
<td>-0.045</td>
<td>0.0000</td>
<td>166</td>
<td>0</td>
</tr>
<tr>
<td>182</td>
<td>5</td>
<td>-0.038</td>
<td>-4.221</td>
<td>-0.094</td>
<td>-1.204</td>
<td>0.009</td>
<td>166</td>
<td>0</td>
</tr>
<tr>
<td>365</td>
<td>1</td>
<td>0.047</td>
<td>3.510</td>
<td>-0.097</td>
<td>-0.856</td>
<td>0.010</td>
<td>78</td>
<td>0</td>
</tr>
<tr>
<td>365</td>
<td>2</td>
<td>0.053</td>
<td>6.869</td>
<td>-0.141</td>
<td>-1.286</td>
<td>0.021</td>
<td>78</td>
<td>0</td>
</tr>
<tr>
<td>365</td>
<td>3</td>
<td>0.039</td>
<td>6.305</td>
<td>-0.068</td>
<td>-0.614</td>
<td>0.005</td>
<td>78</td>
<td>0</td>
</tr>
<tr>
<td>365</td>
<td>4</td>
<td>0.024</td>
<td>3.884</td>
<td>-0.002</td>
<td>-0.014</td>
<td>0.0000</td>
<td>78</td>
<td>0</td>
</tr>
<tr>
<td>365</td>
<td>5</td>
<td>-0.053</td>
<td>-2.549</td>
<td>-0.110</td>
<td>-1.011</td>
<td>0.013</td>
<td>78</td>
<td>0</td>
</tr>
</tbody>
</table>

Panel B: Equally weighted
The figure shows the estimated intercepts from the expected return equations \( \hat{c}_{p,t}(\theta) \) computed in equation (36) for \( \theta = 1 \). Intercepts are constructed for value-weighted quintile portfolios sorted on average RN variance over the two-year periods. Each panel shows the series for a given quintile portfolio \( p = \{1, 2, 3, 4, 5\} \) and maturity \( h = \{30, 91, 182, 365\} \) days.

**Figure 5: Time series of intercepts**
Figure 6: Regression tests of the variance constraint equation: one quarter horizon
The figure presents the results of the test for one-factor structure of the return variances. The top row of panels shows the estimates of the coefficients $a$ and $b$ from regression (62) estimated for each month. The middle row shows $t$-statistics for tests $a = 0$ (left panel) and $b = 1$ (right panel). The bottom row shows $p$-values of the $F$-test for the joint restriction $a = 0$ and $b = 1$ (left panel) and the adjusted $R^2$ of the regressions (right panel). The regressions are based on sets of stocks for one quarter horizon.

![Figure 6: Regression tests of the variance constraint equation: one quarter horizon](image-url)
Figure 7: Regression tests of the variance constraint equation: six-month horizon
The figure presents the results of the test for one-factor structure of the return variances. The top row of panels shows the estimates of the coefficients $a$ and $b$ from regression (62) estimated for each month. The middle row shows $t$-statistics for tests $a = 0$ (left panel) and $b = 1$ (right panel). The bottom row shows $p$-values of the $F$-test for the joint restriction $a = 0$ and $b = 1$ (left panel) and the adjusted $R^2$ of the regressions (right panel). The regressions are based on sets of stocks for six-month horizon.
Figure 8: Regression tests of the variance constraint equation: one year horizon
The figure presents the results of the test for one-factor structure of the return variances. The top row of panels shows the estimates of the coefficients $a$ and $b$ from regression (62) estimated for each month. The middle row shows $t$-statistics for tests $a = 0$ (left panel) and $b = 1$ (right panel). The bottom row shows $p$-values of the $F$-test for the joint restriction $a = 0$ and $b = 1$ (left panel) and the adjusted $R^2$ of the regressions (right panel). The regressions are based on sets of stocks for one year horizon.
Appendix A  Alternative derivation and indeterminacy

The alternative derivation of the expected return formula from the earlier working paper MW (2016) does not use Q-beta linearization and yet it produces similar indeterminacy of coefficients as the one discussed in the main text. I first reproduce the alternative method, which is using an algebraic identity, and then show that this method also implies continuum of equations with arbitrary coefficients. Consider the following identity from MW (2016):

\[ R_i R_j = \frac{1}{2} \left[ R_i^2 + R_j^2 - (R_i - R_j)^2 \right] \]  

(A.1)

and use equation (9) with \( j = \text{"g"} \) to obtain:

\[
E_P \{ R_{i,t+1} \} = \frac{1}{R_{f,t+1}^2} E_Q \{ R_{i,t+1} R_{g,t+1} \} = \frac{1}{2} \frac{R_{g,t+1}}{R_{f,t+1}} \left[ \frac{R_{i,t+1}^2 + R_{g,t+1}^2 - (R_{i,t+1} - R_{g,t+1})^2}{R_{f,t+1}^2} \right] 
\]  

(A.2)

From this, using the properties of RN measure in (6), after some manipulation derive:

\[
\frac{E_P \{ R_{i,t+1} \} - R_{f,t+1}}{R_{f,t+1}} = \frac{1}{2} Var_Q \left[ \frac{R_{i,t+1}}{R_{f,t+1}} \right] + \frac{1}{2} Var_Q \left[ \frac{R_{g,t+1}}{R_{f,t+1}} \right] - \frac{1}{2} Var_Q \left[ \frac{R_{i,t+1} - R_{g,t+1}}{R_{f,t+1}} \right]
\]

(A.3)

MW (2016) then define \( \Delta_{i,t} \) and assume that it is additively separable in asset- and time-specific terms as follows:

\[
\Delta_{i,t} \equiv -\frac{1}{2} Var_Q \left[ \frac{R_{i,t+1} - R_{g,t+1}}{R_{f,t+1}} \right] = \alpha_i + \lambda_t
\]

(A.4)

One additional (normalization) assumption is that \( \sum_{i=1}^{N} w_{i,t}\alpha_i = 0 \), that is the cross sectional mean of \( \alpha_i \)'s is folded into \( \lambda_t \). From (A.3) the same steps with adding for index portfolio and then subtracting from individual expected return formula result in the same equation (27) with \( \theta = 1 \).

This alternative derivation of the expected return formula shares similar indeterminacy of coefficients with the published version of MW (2018). To see this, note that for any real number \( \theta > 0 \) the identity (A.1) can be extended as follows:

\[
R_i R_j = \frac{1}{2} \left[ \theta R_i^2 + \frac{1}{\theta} R_j^2 - (\sqrt{\theta} R_i - \frac{1}{\sqrt{\theta}} R_j)^2 \right]
\]

(A.5)

While equation (A.1) corresponds to \( \theta = 1 \), the above identity leaves \( \theta \) unspecified. The only relevant property for it to work is that coefficients on each return must be reciprocal to each other.

\( ^{16} \)For \( \theta < 0 \) replace the third term in square brackets, including the “−” in front, with \( +(\sqrt{|\theta|} R_i + \frac{1}{\sqrt{|\theta|}} R_j)^2 \).
I can then retrace the steps used previously to obtain equation (27). The new version of equation (A.3) becomes:

\[
E_t^P \{ R_{i,t+1} \} - R_{f,t+1} = \frac{\theta}{2} Var_t^Q \left[ \frac{R_{i,t+1}}{R_{f,t+1}} \right] + \frac{1}{2\theta} Var_t^Q \left[ \frac{R_{g,t+1}}{R_{f,t+1}} \right] - \frac{1}{2} Var_t^Q \left[ \frac{\sqrt{\theta} R_{i,t+1} - \frac{1}{\sqrt{\theta}} R_{g,t+1}}{R_{f,t+1}} \right]
\]

(A.6)

I redefine \( \Delta_{i,t} \), \( \alpha_{i} \) and \( \lambda_{t} \) from (A.4) with analogous additive separability assumption as follows:

\[
\Delta_{i,t}(\theta) \equiv -\frac{1}{2} Var_t^Q \left( \frac{\sqrt{\theta} R_{i,t+1} - \frac{1}{\sqrt{\theta}} R_{g,t+1}}{R_{f,t+1}} \right) = \alpha_{i}(\theta) + \lambda_{t}(\theta)
\]

and, as before, impose \( \sum_{i=1}^{N} w_{i,t} \alpha_{i}(\theta) = 0 \) without loss of generality. From here it is straightforward to follow the same steps as before to obtain equation (27).

**Appendix B Covariance approximation**

This appendix reproduces the approximation of the conditional covariance \( Cov_t^P (m_{t+1} R_{i,t+1}, R_{i,t+1}) \) from Kadan and Tang (2017) and extends it to the case of similar covariance for excess returns \( Cov_t^P (m_{t+1} R_{e,t+1}, R_{e,t+1}) \). Kadan and Tang assume a representative agent economy with a homogenous utility function \( u(w) \) and the relative risk aversion \( RRA(w) \equiv -w \frac{u''(w)}{u'(w)} \). They also assume that the index portfolio is the market portfolio and that the SDF can be represented as:

\[
m_{t+1} = \frac{1}{\lambda_{t}} u'(R_{I,t+1}) \text{, where } \lambda_{t} = R_{f,t+1} E_t^P \{ u'(R_{I,t+1}) \}.
\]

(B.1)

Consider the first order Taylor approximation at \( R_{i,t+1} \approx E_t^P \{ R_{I,t+1} \}, R_{I,t+1} \approx E_t^P \{ R_{I,t+1} \} \), i.e. near the market expected return for both returns:

\[
u'(R_{I,t+1}) R_{i,t+1} \approx u'(E_t^P \{ R_{I,t+1} \}) E_t^P \{ R_{I,t+1} \} + u'(E_t^P \{ R_{I,t+1} \}) (R_{i,t+1} - E_t^P \{ R_{I,t+1} \})
\]

\[+ E_t^P \{ R_{I,t+1} \} u''(E_t^P \{ R_{I,t+1} \}) (R_{i,t+1} - E_t^P \{ R_{I,t+1} \})
\]

(B.2)

\[17\text{To derive it, I use the property } E_t^Q \left[ \frac{R_{i,t+1} + 1}{R_{f,t+1} + 1} \right] = 1 \text{ and } \frac{1}{2} \left[ \theta + \frac{1}{\theta} - (\sqrt{\theta} - \frac{1}{\sqrt{\theta}})^2 \right] = 1.
\]

\[18\text{Kadan and Tang use a more general multi-variate approximation with all returns near the market expected return: } R_{i,t+1} \approx E_t^P \{ R_{I,t+1} \}, i \in \{ 1, ..., N \}. \text{ But given that the weighted approximation points sum up to } E_t^P \{ R_{I,t+1} \} \text{ approximating the index return directly results in equivalent expressions.}
\]
Then the covariance is approximated as follows:

\[
Cov_t^P \left( \frac{u'(R_{I,t+1})}{\lambda_t} R_{i,t+1}, R_{i,t+1} \right) \approx \frac{u' (E_t^P \{ R_{I,t+1} \})}{R_{f,t} E_t^P \{ u'(R_{I,t+1}) \}} \left[ Var_t^P (R_{i,t+1}) \right. \\
+ \frac{u'' (E_t^P \{ R_{I,t+1} \})}{u' (E_t^P \{ R_{I,t+1} \})} E_t^P \{ R_{I,t+1} \} Cov_t^P (R_{i,t+1}, R_{I,t+1}) \\
= \frac{u' (E_t^P \{ R_{I,t+1} \})}{R_{f,t} E_t^P \{ u'(R_{I,t+1}) \}} \left[ Var_t^P (R_{i,t+1}) \right. \\
- RRA \left( E_t^P \{ R_{I,t+1} \} \right) Var_t^P (R_{I,t+1}) \beta_{i,t}^P 
\]

where I used the fact that \( E_t^P \{ R_{I,t+1} \} \) is conditionally constant and applied standard definitions of beta and relative risk aversion. This is the first approximation used in the text.

Now consider similar Taylor expansion for the term with excess return \( R_{i,t+1}^e \equiv R_{i,t+1} - R_{f,t+1} \), again approximated at the expected return on the index (for both the index and stock return):

\[
\begin{align*}
    u'(R_{I,t+1}) R_{i,t+1}^e & \approx u'(E_t^P \{ R_{I,t+1} \}) E_t^P \{ R_{I,t+1}^e \} + u'(E_t^P \{ R_{I,t+1} \}) (R_{i,t+1}^e - E_t^P \{ R_{I,t+1}^e \}) \\
    & + E_t^P \{ R_{I,t+1}^e \} u''(E_t^P \{ R_{I,t+1} \}) (R_{i,t+1}^e - E_t^P \{ R_{I,t+1} \})
\end{align*}
\]

Then, the covariance using excess returns is approximated as follows:

\[
Cov_t^P \left( \frac{u'(R_{I,t+1})}{\lambda_t} R_{i,t+1}^e, R_{i,t+1}^e \right) \approx \frac{u' (E_t^P \{ R_{I,t+1} \})}{R_{f,t} E_t^P \{ u'(R_{I,t+1}) \}} \left[ Var_t^P (R_{i,t+1}) \right. \\
+ \frac{u'' (E_t^P \{ R_{I,t+1} \})}{u' (E_t^P \{ R_{I,t+1} \})} E_t^P \{ R_{I,t+1} \} Cov_t^P (R_{i,t+1}, R_{I,t+1}) \\
= \frac{u' (E_t^P \{ R_{I,t+1} \})}{R_{f,t} E_t^P \{ u'(R_{I,t+1}) \}} \left[ Var_t^P (R_{i,t+1}) \right. \\
- RRA \left( E_t^P \{ R_{I,t+1} \} \right) \left( \frac{E_t^P \{ R_{I,t+1} \}}{E_t^P \{ R_{I,t+1} \}} \right) Var_t^P (R_{I,t+1}) \beta_{i,t}^P
\]

where I used the fact that \( Cov_t^P \left( R_{i,t+1}^e, R_{i,t+1}^e \right) = Var_t^P (R_{i,t+1}) \). This is the second approximation used in the text.

**Appendix C  Sample construction for empirical tests**

To test for constant intercept assumption in the expected returns equation I construct estimates of the intercepts in equation (36). The data for returns is obtained from CRSP and the data for option
prices is obtained from Option Metrics, both accessed through WRDS. The original sample includes daily data from January 1, 1996 to December 31, 2017 and includes stocks that were at any time during this window among constituents of the S&P500 index. The list of index constituents is from CRSP and the header information is matched with Option Metrics, first on CUSIP and then a few remaining unmatched securities are matched on ticker symbol to achieve nearly complete match.\footnote{Fewer than 20 permanent numbers from CRSP cannot be matched to a security ID in Option Metrics using this method, out of slightly over 1000 stocks that were in the index at any time during the sample years.}

To construct measures for RN variances I use implied premiums for puts and calls from volatility surface files in Option Metrics with standardized maturities of 30, 91, 182, and 365 days.

I apply several filters to the options prices data. First, there must be at least three available implied strikes for puts and calls with valid implied premium. Second, the range of puts and calls implied strikes must overlap. Third, the close price of the stock must be within the range for puts and calls. For the tests I use monthly frequency data and select from each month in the sample the first date with available option prices and returns. I use the daily CRSP data to construct cumulative returns with a given horizon up to the maturity of the option.\footnote{If the end day of the maturity horizon falls on the weekend, then the cumulative return to the previous business day is scaled by the ratio of the option maturity to the actual number of calendar days until the last business day with available returns. Typically this results in actual calendar days being less by one or two days relative to standardized maturity, so some of these factors are slightly larger than one.}

For the tests I use monthly frequency data and select from each month in the sample the first date with available option prices and returns. I use the daily CRSP data to construct cumulative returns with a given horizon up to the maturity of the option.\footnote{MW also filter a small number of stock-months based on monotonicity of the computed RN variance measures. The filters I use appear to remove such instances and the resulting RN variances are all monotonic in maturity.}

For the RN variance measures I use implied premiums for puts and calls from volatility surface files in Option Metrics with standardized maturities of 30, 91, 182, and 365 days.

To construct measures for RN variances I use implied premiums for puts and calls from volatility surface files in Option Metrics with standardized maturities of 30, 91, 182, and 365 days.

I apply several filters to the options prices data. First, there must be at least three available implied strikes for puts and calls with valid implied premium. Second, the range of puts and calls implied strikes must overlap. Third, the close price of the stock must be within the range for puts and calls. For the tests I use monthly frequency data and select from each month in the sample the first date with available option prices and returns. I use the daily CRSP data to construct cumulative returns with a given horizon up to the maturity of the option.\footnote{If the end day of the maturity horizon falls on the weekend, then the cumulative return to the previous business day is scaled by the ratio of the option maturity to the actual number of calendar days until the last business day with available returns. Typically this results in actual calendar days being less by one or two days relative to standardized maturity, so some of these factors are slightly larger than one.}

For the tests I use monthly frequency data and select from each month in the sample the first date with available option prices and returns. I use the daily CRSP data to construct cumulative returns with a given horizon up to the maturity of the option.\footnote{MW also filter a small number of stock-months based on monotonicity of the computed RN variance measures. The filters I use appear to remove such instances and the resulting RN variances are all monotonic in maturity.}

For stocks with available options data, the RN variance measures are computed by numerically integrating using the standard spanning formula (e.g. Carr and Madan (2001) and Bakshi and Madan (2000)). Theoretical value of the RN variance expectation is given by the following integral:

\[
SVIX^2_{t,t} \equiv E_t^Q \left\{ \frac{(R_{i,t\tau} - R_{f,t\tau})^2}{R^2_{f,t\tau}} \right\} = \frac{2}{R^2_{f,t\tau}S^2_t} \int_0^\infty \min\left[C_t(K,\tau), P_t(K,\tau)\right] dK, \tag{C.1}
\]

where \(S_t\) is the time-\(t\) stock price, and \(C(K,\tau)\) and \(P(K,\tau)\) are premiums for European call and put on the stock, respectively, for strike \(K\) and maturity \(\tau\).

To compute the integral (C.1) numerically, I define a fine grid of 1000 points on moneyness spanning from 0.2 to 2.5. For grid points within the range of implied strikes provided by Option Metrics, I use cubic spline interpolation to approximate the value of the premiums. For grid points
outside of the available implied strikes, I use Black-Scholes formula to compute the premium with
volatility set to the implied volatility provided for the lowest (highest) point for puts (calls). That is,
outside of the volatility surface data, the implied volatility is assumed to be constant corresponding
to the nearest available OTM option. I treat options as European due to the adjustment to premium
computed in Option Metrics volatility surface files.

The index portfolio \( I \) is constructed using either equal weights or the average market capital-
ization of the available stocks over a given year. To construct portfolios for the tests I assign stocks
into quintiles based on the average \( \text{SVIX}_{i,t}^2 \) over every two years and aggregate returns and RN
variances either equally weighted or value weighed using average market capitalization during the
year. Each of the four maturities has five quintile portfolios and two different weighting schemes
for portfolio and the index. This results in 40 different monthly series of estimated intercepts as
defined in equation \( 36 \). The monthly series are used entirely while the remaining maturities are
sampled to avoid overlap: 91 days maturities are sampled at the beginning of calendar quarter
months (January, April, July, October); 182 days maturities are sampled in January and July; 365
days maturities are sampled in January. Table 1 provides summary statistics for the portfolios and
the index and Figure 5 shows time-series of the intercepts for each portfolio-maturity combination
for the value-weighted case.

In the construction of the final sample for the regression tests \( 37 \) I also limit the number of
times each intercept is used to four. At every time \( t \) the intercepts from up to period \( t' = t + 4 \) are
included in the regression equation on the right hand side. As a result, each intercept enters four
times as the left-hand side variable and four times as the right hand side variable. The total number
of available observation for the intercepts regression is \( 4T - 10 \) where \( T \) is the number of available
periods over the 22 year sample window (264 months, 88 quarters, 44 half-years)