SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2010-9

Pitfalls in VAR based return decompositions: A clarification

Tom Engsted, Thomas Q. Pedersen and Carsten Tanggaard

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Pitfalls in VAR based return decompositions: A clarification*

Tom Engsted, Thomas Q. Pedersen, and Carsten Tanggaard†

Abstract

Based on Chen and Zhao's (2009) criticism of VAR based return decompositions, we explain in detail the various limitations and pitfalls involved in such decompositions. First, we show that Chen and Zhao's interpretation of their excess bond return decomposition is wrong: the residual component in their analysis is not 'cashflow news' but 'interest rate news' which should not be zero. Consequently, in contrast to what Chen and Zhao claim, their decomposition does not serve as a valid caution against VAR based decompositions. Second, we point out that in order for VAR based decompositions to be valid, the asset price needs to be included as a state variable. In parts of Chen and Zhao's analysis the price does *not* appear as a state variable, thus rendering those parts of their analysis invalid. Finally, we clarify the intriguing issue of the role of the residual component in equity return decompositions. In a properly specified VAR, it makes no difference whether return news and dividend news are both computed directly or one of them is backed out as a residual.

Keywords: Return variance decomposition, news components, VAR model, information set, predictive variables, redundant models.

JEL Codes: C32, G12, G17.

---

*February 2010. We gratefully acknowledge the detailed comments from John Campbell on earlier versions of this paper. We alone are responsible for remaining errors.

†CREATES, University of Aarhus, Building 1326, DK-8000 Aarhus C. E-mail: tengsted@creates.au.dk (Tom Engsted); tqpedersen@creates.au.dk (Thomas Q. Pedersen); ctanggaard@creates.au.dk (Carsten Tanggaard). CREATES is funded by the Danish National Research Foundation.

1 Introduction

What causes asset prices to fluctuate? Is it mostly news about future cash-flows or news about future discount rates (risk premia)? Since the seminal studies by Campbell (1991) and Campbell and Ammer (1993), it has become standard to use VAR based return variance decompositions to answer this, and related, questions.[1] The methodology developed by Campbell and his coauthors (which builds on Campbell and Shiller's (1988) log-linear return approximation) is ingenious and has substantially increased our understanding of the underlying drivers of asset market fluctuations. However, there are a number of limitations and pitfalls involved in such VAR based decompositions, and although Campbell and Ammer (1993) and Campbell, Lo, and MacKinlay (1997) did caution against uncritical interpretation of VAR based decompositions, subsequent empirical studies often have not recognized the limitations and pitfalls involved and have computed invalid decompositions. In a recent comprehensive study, Chen and Zhao (2009) discuss some of these limitations. In particular, they show that the decompositions can be quite sensitive to the variables included in the VAR, and the results are highly dependent on the predictive variables capturing the time-varying nature of expected returns. Of crucial importance here seems to be the treatment of cashflow news as a residual to be backed out from an identity.

In this paper we provide further discussion of the various limitations and pitfalls involved in VAR based return decompositions. In particular, we take issue with some of the claims and analyses made by Chen and Zhao (2009). First, we argue that Chen and Zhao's interpretation of their decomposition for bond returns is wrong. Using excess bond returns they find a substantial residual component which they denote 'cashflow news' in their decomposition, and they claim that this is an error because for Treasury bonds nominal casflows are fixed; thus, cashflow news must be zero. However, we point out that since Chen and Zhao make the decomposition for *excess* bond returns, i.e. bond return in excess of a short term nominal interest rate, what they call 'cashflow news' is in fact 'nominal interest rate news', and this component should not necessarily be zero. Thus, in contrast to what Chen and

[1] Such return variance decompostions have also been used in studies of monetary policy effects on stock returns (Patelis, 1997; Bernanke and Kuttner, 2005), intertemporal CAPM's (Campbell and Vuolteenaho, 2004), international stock market linkages (Ammer and Mei, 1996; Engsted and Tanggaard, 2004), and in many other areas (see section 1.3.1 in Chen and Zhao (2009) for further references).

Zhao claim, their bond return decomposition does not document counterintuitiveness or unreliability of the VAR based decomposition approach.

Second, we point out a crucial aspect of VAR based decompositions neglected by Chen and Zhao, namely that in order for the decomposition to be valid, the asset price needs to be incuded as a state variable in the VAR. This is sometimes forgotten in empirical studies. In parts of Chen and Zhao's (2009) analysis the price does *not* appear as a state variable, thus rendering those parts of their analysis invalid.

Third, and related to the first point, an important element in most applications of the decomposition for equities is the treatment of cashflow news as a residual; dividend growth does not appear as a separate state variable in the VAR. This has caused confusion in the literature about the possible overstatement of the importance of the cashflow news component. Insufficient return predictability by the state variables is often conjectured to lead to an upward biased cashflow news component due to this component's residual determination. However, as pointed out recently by Campbell, Polk, and Vuolteenaho (2010), if one abstracts from the approximation error in the underlying log-linear return approximation, in a VAR that contains the dividend-price ratio as a state variable, it does not matter for the return decomposition whether the cashflow news component or the discount rate news component is obtained as a residual. What matters is the choice of predictor variables (in addition to the dividend-price ratio) to include in the VAR. Campbell, Polk, and Vuolteenaho are not completely clear in explaining this insight, so we attempt to provide a full and detailed clarification of this point. In particular, through a small empirical analysis using US stock market data, we show that in a properly specified VAR model the only difference between backing out the cashflow news component or the discount rate news component is the approximation error.

Finally, due to the conjectured problems of treating the cashflow news component as a residual, Chen and Zhao (2009) suggest to model cashflows directly along with returns. We point out, however, that in a properly specified VAR, it makes no difference whether return news and dividend news are both computed directly or one of them is backed out as a residual. The reason it makes a difference in Chen and Zhao's analysis is because they use *excess* stock returns instead of just returns (which implies that part of their "model noise" component captures real interest rate news), and because they impose arbitrary - but untested - zero restrictions in their VAR model.

The main conclusion from our analysis is that while Chen and Zhao are correct in arguing that the VAR based return decomposition approach is somewhat sensitive to the state variables included in the VAR, their claims about the "unreliability" and "counterintuitiveness" of the approach are unjustified. Used in a proper way, the approach is a very useful and informative method for analyzing the movements and pricing of asset returns.

We emphasize that some of the points we make are not new and can be found in the existing literature. However, it is our impression that many empirical researchers (ourselves included) have not been fully aware of the - in some cases quite intriguing - pitfalls and limitations of VAR based return decompositions and, hence, we believe that a detailed and comprehensive explanation for them will be useful.

The rest of the paper is organized as follows. Section 2 describes the basic return decomposition and the associated VAR methodology. Then, in section 3, we explain in detail the various pitfalls involved in this methodology. We discuss decompositions for both bond and stock returns, and we provide a small empirical example illustrating our main points. Section 4 offers some concluding remarks. The Appendix contains some technical details.

2 VAR based return variance decomposition

In this section we describe the decomposion for stock returns because most of our subsequent discussion will focus on equities. In sub-section 3.1 in the next section (and in footnote 9 in sub-section 3.2) we consider the decomposition for bonds.

2.1 The return decomposition

The stock return decomposition is derived from Campbell and Shiller's (1988) log-linear return approximation

$$r_{t+1} \approx k + \rho p_{t+1} + (1-\rho)d_{t+1} - p_t, \tag{1}$$

where r_{t+1} is log return from t to $t+1$, p_t and d_t are log prices and log dividends at time t, respectively, ρ is a constant slightly less than one, and k is a linearization constant. Equation (1) can be written as $p_t \approx k + \rho p_{t+1} + (1-\rho)d_{t+1} - r_{t+1}$, and solving this equation forward for p_t and imposing the no-bubble transversality condition $\lim_{j\to\infty} \rho^j p_{t+j} = 0$, gives (abstracting from the fact that the relation only holds approximately)

$$p_t = \sum_{i=0}^{\infty} \rho^j [(1-\rho)d_{t+1+i} - r_{t+1+i}] + \frac{k}{1-\rho}. \tag{2}$$

An equation similar to (2) holds for p_{t+1}, where all variables are leaded one period. Now, by inserting these present value expressions for p_t and p_{t+1} into the conditional expectations version of (1), $E_t p_t = p_t = k + \rho E_t p_{t+1} + (1 - \rho)E_t d_{t+1} - E_t r_{t+1}$, we obtain after some rewriting the following expression for unexpected one-period returns (c.f. Campbell, 1991):

$$r_{t+1} - E_t r_{t+1} = (E_{t+1} - E_t) \sum_{i=0}^{\infty} \rho^i \Delta d_{t+1+i} - (E_{t+1} - E_t) \sum_{i=1}^{\infty} \rho^i r_{t+1+i}. \tag{3}$$

The term $(E_{t+1} - E_t)$ represents the change in expectation due to new information that arrives between time t and $t+1$. Thus, the two components on the right-hand side of (3) represent 'news' about future dividend growth and returns, respectively. In the literature these two components are sometimes referred to as 'cashflow news' and 'discount rate news'. The equation says that an unexpected positive stock return must be due to either positive news about future dividend growth or negative news about future returns, or both.[2]

It is important to realize that - given there are no bubbles - (3) holds as a dynamic approximate *identity*. It is derived from the log-linear return approximation, which is derived from the way stock returns are defined. If there is no bubble, and dividend growth is fixed, a rise in future returns can only come about by a decrease in stock prices today, i.e. a negative current return. Similarly, if future returns are fixed, higher future dividend growth must imply a higher level of stock prices, i.e. a positive current return. Denote by η_{t+1} the return innovation component on the left-hand side of (3), and denote by $\eta_{d,t+1}$ and $\eta_{r,t+1}$, respectively, the two news components on the right-hand side. Then (3) can be written as

$$\eta_{t+1} = \eta_{d,t+1} - \eta_{r,t+1}. \tag{4}$$

Note that in the above derivation it is assumed that $E_t p_t = p_t$, i.e. p_t is in the information set at time t. This is an important aspect to remember when

[2]For $i = 0$, the first term on the right-hand side of (3) is $\Delta d_{t+1} - E_t \Delta d_{t+1}$. This is, strictly speaking, not a 'news' component but an 'innovation' component, i.e. unexpected one-period dividend growth.

setting up the VAR model to be used in estimating the innovation and news components (see below).

2.2 The VAR model

Following Campbell (1991) and Campbell and Ammer (1993), many papers have used (4) in combination with a VAR model to estimate the fraction of unexpected return volatility attributable to news about future dividend growth and news about future returns, respectively. In most papers the VAR model includes returns, r_t, as the first variable, and a number of predictive variables for returns as the other variables. A crucial such predictor variable is the dividend-price ratio, $d_t - p_t$, which is stationary in case of no bubbles, and which in the return predictability literature has been found to be an important predictor of future returns, c.f. e.g. Cochrane (2008). Write the VAR model as follows (where all variables are measured in deviation from their unconditional means, and where we for simplicity assume a first-order system):

$$\begin{bmatrix} r_{t+1} \\ d_{t+1} - p_{t+1} \\ x_{t+1} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} r_t \\ d_t - p_t \\ x_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{r,t+1} \\ \varepsilon_{dp,t+1} \\ \varepsilon_{x,t+1} \end{bmatrix}. \tag{5}$$

x_t is an arbitrary stationary predictor variable, taken to be a scalar for simplicity, a_{11},...,a_{33} are VAR parameters, and $\varepsilon_{j,t+1}$ is the VAR innovation associated with the j'th equation. This VAR system can be written in compact form as $\mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t + \boldsymbol{\varepsilon}_{t+1}$. From this system the VAR estimate of $\mathbf{z}_{t+1} - E_t\mathbf{z}_{t+1}$ is $\boldsymbol{\varepsilon}_{t+1}$. Similarly, the VAR estimate of $(E_{t+1} - E_t)\mathbf{z}_{t+1+i}$ is $\mathbf{A}^i\boldsymbol{\varepsilon}_{t+1}$. Therefore, if we define a selection vector $\mathbf{e1}' = (1 \;\; 0 \;\; 0)$ that picks out the first element of the VAR, $r_{t+1} = \mathbf{e1}'\mathbf{z}_{t+1}$, the VAR estimates of the return innovation and return news components in (4) are

$$\eta_{t+1} = \mathbf{e1}'\boldsymbol{\varepsilon}_{t+1}, \tag{6}$$

$$\eta_{r,t+1} = \mathbf{e1}'\rho\mathbf{A}(\mathbf{I} - \rho\mathbf{A})^{-1}\boldsymbol{\varepsilon}_{t+1}, \tag{7}$$

where the VAR residuals and parameter estimates are inserted in $\boldsymbol{\varepsilon}_{t+1}$ and $\mathbf{A}$, respectively. Given the two components in (6) and (7), the final news component $\eta_{d,t+1}$ is computed from the identity (4):

$$\eta_{d,t+1} = \eta_{t+1} + \eta_{r,t+1}. \tag{8}$$

To measure the relative importance of return news and dividend news in explaining the variability of return innovations, usually the variances and covariances of each of the components (6)-(8) are computed. From (4) it follows that $\mathrm{Var}(\eta_{t+1}) = \mathrm{Var}(\eta_{d,t+1}) + \mathrm{Var}(\eta_{r,t+1}) - 2\mathrm{Cov}(\eta_{d,t+1}, \eta_{r,t+1})$. Thereby the fraction of return innovation variance explained by, say, dividend news is $\mathrm{Var}(\eta_{d,t+1})/\mathrm{Var}(\eta_{t+1})$. Due to the covariance term such variance ratios may be difficult to interpret, so instead one can orthogonalize the components using a Cholesky decomposition. The drawback of orthogonalizing, however, is that the Choleskey decomposition is not independent of the ordering of the variables and it is not clear whether returns or dividends should be ordered first. In the literature one can find both unorthogonalized and orthogonalized decompositions, and with various orderings of the variables.

As seen from the above, the return innovation component and the return news component are computed directly from the VAR estimates, while the dividend news component is obtained indirectly as a residual. This is the procedure most often used in the literature. When working with monthly or quarterly data, an advantage of this procedure is that one does not have to deal with the strong seasonal variation that characterizes dividend growth. However, there are several limitations and pitfalls in using (4) and (5) in the manner described above to decompose unexpected stock returns into news about future dividends and returns.

3 The pitfalls

3.1 The first pitfall: Chen and Zhao's (2009) bond return decomposition

Chen and Zhao (2009) conduct a comprehensive analysis of the limitations of VAR based return decompositions. In the first part of their analysis they focus on bond returns. Hence, we begin by discussing their bond return decomposition. Then, in the subsequent sections we focus more on equities.

With reference to Campbell and Ammer's (1993) equation (A4), Chen and Zhao (2009) write up the following equation

$$e_{t+1} - \left(-(E_{t+1} - E_t) \sum_{j=1}^{N} r_{t+1+j} \right) = 0. \quad (9)$$

e_{t+1} is the bond return innovation, i.e. the left-hand side of (3) where r_{t+1} is now the one-period bond return, and where the cashflow news component is zero because bonds have fixed nominal cashflows. The summation in (9) goes to N because the bond has N periods to maturity, and there is no discounting with ρ because it is a zero-coupon bond (meaning that cashflows are zero).[3] With reference to (9), Chen and Zhao state that "the difference between unexpected bond return and DR [discont rate] news (i.e. CF [cashflow] news by definition) must be zero."

Next, Chen and Zhao employ a VAR system with excess bond return as the first element and state variables like the term spread, the real interest rate, inflation, and the credit spread as the other elements. They estimate the VAR using Treasury bond data from Ibbotson covering the period 1926-2002. From the estimated VAR they compute the return innovation component and the DR news component directly, and then infer the CF news component as a residual by subtracting the first two components from each other, c.f. equation (9). Using this procedure, they find that the variance of the CF news component is larger than or at least as large as the variance of the DR news component. Since the CF news component should be zero, they claim that their analysis has shown that the usefulness of the VAR based decomposition approach is seriously limited. They attribute the substantial CF news component to model misspecification and insufficient bond return predictability by the state variables.

However, there is a flaw in Chen and Zhao's interpretation of their results. In Chen and Zhao's empirical analysis they work with *excess* bond returns, i.e. the bond return in excess of a short term nominal interest rate. However, equation (9) holds for returns, *not* excess returns. Campbell and Ammer (1993) write up the equation for excess bond returns (equation (3) in Campbell and Ammer's paper) which shows that excess bond return innovations are the sum of inflation news, real interest rate news, and excess

[3] Chen and Zhao's (2009) notation is unclear. To be precise, if the zero-coupon bond has N periods to maturity, then the summation in (9) should go to N-1 instead of N. In addition, the returns in the summation should not denote future one-period returns on bonds with unchanging maturity, but future one-period returns on a bond whose maturity shrinks over time. See equation (A4) in Campbell and Ammer (1993).

bond return news (all multiplied by minus one). The first two of these news components combined give news about future nominal interest rates. Denote by $y_{1,t+1}$ the one-period nominal interest rate, and $r_{N,t+1} - y_{1,t+1}$ the excess return on the N-period bond. Then Campbell and Ammer's equation (3) implies the following decomposition for excess return innovations, $e_{N,t+1} \equiv (r_{N,t+1} - y_{1,t+1}) - E_t(r_{N,t+1} - y_{1,t+1})$:

$$e_{N,t+1} - \left(-(E_{t+1} - E_t) \sum_{j=1}^{N-1} y_{1,t+1+j} - (E_{t+1} - E_t) \sum_{j=1}^{N-1} (r_{N-j,t+1+j} - y_{1,t+1+j}) \right) = 0 \tag{10}$$

From footnote 3 in Campbell and Ammer (1993) it is clear that in obtaining their equation (3), i.e. our equation (10), they assume that the one-period nominal interest rate used to define excess returns is known at time t such that $E_t y_{1,t+1} = y_{1,t+1}$. This implies that excess return innovations, $e_{N,t+1}$, are identical to return innovations, $r_{N,t+1} - E_t r_{N,t+1}$. Thus, by construction (10) is a decomposition for both return and excess return *innovations*. This may be the reason why Chen and Zhao do not distinguish between returns and excess returns in their discussion and application of (9). However, by comparing (9) and (10), we see that an extra *news* component appears when decomposing excess returns instead of just returns, namely news about future nominal interest rates. The intuition is as follows: for fixed future *excess* returns, news that future nominal interest rates will be higher must imply that future returns must also be higher. Since the bond expires at par at time $t+N$, this means that the current bond price will have to drop, thereby giving a negative return innovation.

Equation (10) shows that when the excess bond return news component is subtracted from the excess bond return innovation component, which is what Chen and Zhao do to obtain the socalled 'CF news' component, then in fact what that component captures is not cashflow news but nominal interest rate news. Naturally, that news component will not be zero; in fact, nominal interest rates have varied a lot - and unpredictably - over time, in part due to large unpredictable movements in inflation. Campbell and Ammer (1993) document that inflation news accounts for a substantial part of the variability of excess bond return innovations. This is fully consistent with Chen and Zhao's finding of a substantial residual news component in their excess bond return decompositon.

It is not clear why Chen and Zhao in their theoretical discussion focus on bond returns but then in the empirical analysis use *excess* returns. The only hint they give is on p.5227 where they write: "We have used excess bond returns, which is decomposed into the DR and CF news components. This procedure is different from that of Campbell and Ammer (1993), who decompose bond *yield* (instead of actual returns) into three DR news components. The procedure we use is appropriate for our purpose because it exactly matches what has been done to the equity returns in Campbell and Vuolteenaho (2004a): the realized excess returns are used, and the DR and CF news are separated. The consistent treatment on both equity and bond makes the comparison meaningful." [4]

From the first part of this statement one gets the impression that Campbell and Ammer's decomposition is for the bond yield *and not the bond return.* But this is not correct. It is clear from Campbell and Ammer's equation (3), i.e. our equation (10), that their decomposition is for the one-period excess return innovation on an N-period zero-coupon bond. It is true that the one-period return innovation is equivalent to the bond yield innovation (see Campbell and Ammer's equation (6)), but the decomposition is still for the *return* innovation (which - due to the equivalence between return innovation and yield innovation - can be reinterpreted as a decomposition for the bond yield innovation). From the second part of Chen and Zhao's statement one can infer that they use *excess* bond returns in order to make a consistent treatment of stocks and bonds, and decompositions for stocks are usually done in terms of excess returns. However, it is not clear why Chen and Zhao want to compare their bond return decomposition with decompositions for stocks. The point of using bond returns is to focus on an asset where we know that the CF news component is zero. Finding a significant CF news component would indicate problems for the VAR based methodology. For making this point there is no need to make a comparison to stock return decompositions. And, as we have seen, using *excess* bond returns blurs the interpretation of the residual news component.

Thus, both Campbell and Ammer's analysis and Chen and Zhao's analysis give variance decompositions for excess bond return innovations, and both analyses indicate that nominal interest rate news is a major determinant of the variability of unexpected excess returns. Since Chen and Zhao neglect the interest rate news component, they misinterpret this result as implying

[4] "Campbell and Vuolteenaho (2004a)" in the quote from Chen and Zhao (2009) corresponds to Campbell and Vuolteenaho (2004) in our paper.

that the VAR based decomposition methodology is flawed.

3.1.1 Additional comments

There are additional reasons for being cautious when interpreting Chen and Zhao's results and comparing these to the results reported by Campbell and Ammer. Apart from differences in sample periods and frequencies, the main difference between the two analyses is that while Chen and Zhao estimate the excess bond return news component directly by including excess bond returns as a state variable in the VAR, Campbell and Ammer back out this component as a residual. The reason Campbell and Ammer use this approach is: "This choice of residual is forced on us because we cannot directly measure the sequence of excess returns on the bond as its maturity shrinks over its remaining life." (Campbell and Ammer, 1993, p.14).

Chen and Zhao do not use the actual returns from a single bond whose maturity shrinks over time. Instead they use the Ibbotson time series for intermediate-term bond returns where the underlying bond portfolio is rebalanced every period to secure that the maturity of the portfolio is consistently "intermediate-term". This choice, however, has implications for how to apply (9) or (10) - which only hold exactly for N-period zero-coupon bonds - in generating the DR and CF news components.

First, since the Ibbotson bond portfolio consists of coupon bonds, the future returns in (9) and (10), r_{t+1+j} and $(r_{N-j,t+1+j} - y_{1,t+1+j})$, and the future interest rates in (10), $y_{1,t+1+j}$, need to be discounted by a discount rate, ρ, slightly less than one, and even taking discounting into account (9) and (10) now only hold approximately, see e.g. Campbell, Lo, and MacKinlay (1997, pp.406-408). For a coupon bond selling close to par it is suitable to set ρ equal to $1/(1+C)$, where C is the nominal coupon. Chen and Zhao do not state the value of ρ in their application, although they claim that because ρ is close to one it makes little difference to the results whether or not discounting is imposed.

Second, as Campbell, Polk, and Vuolteenaho (2010, footnote 3) have pointed out, using bond portfolio data implies that there should be some 'cashflow news' variability in Chen and Zhao's decomposition because of the rebalancing of the portfolio: the coupon rates of the included bonds vary over time. Thereby the discount rate ρ becomes time-varying. Chen and Zhao (2009, footnote 20) argue - in response to Campbell, Polk, and Vuolteenaho's criticism - that this effect is very minor and that it is unlikely to be the main cause driving their results.

Third, the individual bonds in the portfolio do not have the exact same maturity; thus N is not uniquely defined. Chen and Zhao do not state what value of N they use in their application.

Finally, it is clear from (10) that as j increases, the excess return forecasts should be for bonds with shrinking maturity, and for j close to N the forecasts are for very short term bonds. However, Chen and Zhao's generated excess return forecasts are for a bond portfolio with the same maturity ($\approx N$) for all j. With reference to Campbell and Ammer (1993), Chen and Zhao in their footnote 11 state that "there might be some modeling noise because of the change of bond maturity from one month to the next. Such noise is unlikely to be the main driver of the results." Chen and Zhao's reference to Campbell and Ammer is misplaced here. Campbell and Ammer (1993, footnote 6) refer to the fact that the difference between an n-period yield and an $(n-1)$-period yield is very small for large n. This is not the situation in Chen and Zhao's analysis, where there might be a big difference between return forecasts for intermediate-term bonds and return forecasts for short-term bonds.[5]

All these qualifications notwithstanding, and even if Chen and Zhao are correct in stating that their effects are negligible, the bottom line is that since Chen and Zhao use *excess* returns, their decomposition for Treasury bonds does not provide evidence against the VAR based decomposition approach because the residual news component in their decomposition does not, as they erroneously claim, measure 'cashflow news' (which should be zero); instead the residual news component measures nominal interest rate news which naturally should not be zero. Thus, Chen and Zhao's bond return example does not serve as a valid caution against VAR based return decompositions.

[5] Engsted and Tanggaard (2001) derive a variance decomposition for returns from a *consol*, i.e. an infinite-maturity coupon bond. Naturally, such a bond's maturity does not shrink over time and, hence, in contrast to Campbell and Ammer (1993) it is then possible to include bond returns directly as a state variable in the VAR and to directly compute the bond return news component. However, similarly to Chen and Zhao's (2009) analysis, Engsted and Tanggaard's (2001) analysis suffers from the problem mentioned by Campbell, Polk, and Vuolteenaho (2010), namely that coupon rates are not constant. This introduces an extra model noise component in their analysis.

3.2 The second pitfall: prices in the information set

We now return to the decomposition for equities (though, see footnote 9). The second pitfall involved in VAR estimation of the return decomposition (4) is concerned with the state variable $d_t - p_t$. As already mentioned, in most applications this variable is included as a predictor for future stock returns. However, in some cases $d_t - p_t$ is excluded from the system, which is not legitimate unless p_t is included in some other form instead. This follows from one of the assumptions made in deriving (4) from (1), namely that $E_t p_t = p_t$; that is, p_t is in the information set at time t. If this is not the case, then $E_t p_t = p_t +$ 'projection error', and consequently an extra error term (in addition to the approximation error) would be added to (4). Since (1) can be rewritten in ex ante form as $E_t(d_t - p_t) = d_t - p_t \approx \rho E_t(d_{t+1} - p_{t+1}) + E_t r_{t+1} - E_t \Delta d_{t+1} - k$, given that $d_t - p_t$ is in the time t information set, including $d_t - p_t$ is equivalent to including p_t.[6] In practice $d_t - p_t$ is most often included rather than just p_t because $d_t - p_t$ is more likely to be stationary than p_t.

Researchers are not always aware of the necessity of including prices in the information set and specify a VAR model that does not include $d_t - p_t$, or some other variable involving p_t such that p_t can be isolated, in the state vector. For example, in columns 2, 6, and 7 in Chen and Zhao's (2009) Table 2, none of the state variables involve p_t. Thus, the decompositions in those columns are invalid.[7] Interestingly, the results in columns 2, 6, and 7, are those for which the dividend news component, which is backed out as a residual, gets relatively most importance compared to the return news component (we discuss these results further in the next sub-section). Chen and Zhao (2009, p.5228) state, regarding their choice of state variables, that "The first variable [returns] is necessary to decompose returns, while the others are optional." The second part of this statement is not correct. In addition to returns, the state vector *has* to include a variable involving p_t.[8]

In some recent studies, e.g. Campbell and Vuolteehano (2004) and Campbell, Polk, and Vuolteenaho (2010), the dividend-price ratio is replaced by

[6] Including the dividend-price ratio in level instead of log form, i.e. D_t/P_t, as in e.g. Campbell (1991), is also valid since knowing D_t/P_t is equivalent to knowing $d_t - p_t$.

[7] The same holds for panels A and B4 in Chen and Zhao's (2009) Table 4.

[8] Note that although p_t figures through the state variable $r_t \equiv \log(P_t + D_t) - \log(P_{t-1}) = p_t - p_{t-1} + \log(1 + \exp(d_t - p_t))$, this does not mean that knowing r_t is equivalent to knowing p_t. Thus, having return as a state variable is not sufficient for $E_t p_t$ to equal p_t.

the price-earnings ratio (defined as the log ratio of a price index to a 10-year moving average of past earnings). This is - in theory - not legitimate unless earnings are also included in such a way that prices can be recovered from all the variables. Campbell and his coauthors do not directly discuss this issue, but from pp.333-334 in Campbell, Polk, and Vuolteenaho (2010) one can infer that they have done robustness analyses replacing the price-earnings ratio with the dividend-price ratio but that the latter turns out to be less powerful as a return predictor than the former. Naturally, in practice in a specific sample other valuation ratios may turn out to be better predictors than the theoretically correct dividend-price ratio thus justifying using the other ratios. In any case, valuation ratios like price-earnings are highly correlated with the dividend-price ratio. Another suitable predictor, that in *practice* may be superior to the dividend-price ratio, is the book-to-market ratio, as used by e.g. Chen and Zhao (2009, Table 2, columns 5 and 8) and Campbell, Polk, and Vuolteenaho (2010) in parts of their robustness analysis.

Campbell and Ammer (1993, p.12) write: "The dividend-price ratio is included as a good forecaster of stock returns, and also because *the return decomposition holds only conditional on an information set that includes the stock price itself.*" (Italics added). Campbell (1991) excludes the dividend-price ratio in a minor part of his analysis in order to check the robustness of his results, and he finds that the results are quite sensitive to the inclusion of this variable. This is not surprising in light of the discussion above. Excluding $d_t - p_t$ from the information set, and not including some other variable that involves p_t, effectively means that investors are projecting the *level* of the current price based on only knowing r_t and x_t, and their lags in higher-order systems (in Campbell, 1991, x_t is the 'relative bill rate'). Naturally this may lead to substantial projection errors.[9]

[9] In this sub-section we have discussed the importance of including the stock price in the information set when doing return decomposions for equities. Naturally, a similar point holds for bonds. In bond return decompositions the bond price needs to be included in the VAR information set in order for the decomposition to be valid. For zero-coupon bonds the price of an n-period bond is equal to minus n times the log n-period yield. Thus, including the n-period yield, or the spread between the n-period and 1-period yields (as in Campell and Ammer, 1993), is a valid way of including prices in the information set.

3.3 The third pitfall: the residual news component

The third pitfall concerns the choice of component to be treated as a residual. As explained above, usual practice in the literature has been to compute the return innovation and the return news component directly while obtaining the dividend news component as a residual from the identity (4). This has led to some confusion in the literature about the role of return predictability from the state variables and its possible effects in overstating the importance of the residual component. In this section our aim is to clarify exactly the role of the residual news component. We show that if the VAR is properly specified, the results are insensitive to whether return news or dividend news is treated as a residual, from which it also follows that there are no gains in directly modeling both news components.

Given the VAR model (5), of special interest is the degree of predictability of r_{t+1} from the state variables r_t, $d_t - p_t$ and x_t. Consider the extreme case where these variables are not able to capture any of the time-varying nature of returns. Then the first row of the $\mathbf{A}$ matrix will be a zero vector, and $\eta_{r,t+1}$ in (7) will be 0 for all t. Thus, from (8) all variability of return innovations will be attributed by construction to news about future dividends, simply because the dividend news component is treated as a residual. From this, a natural conjecture is that if the news component treated as a residual (in this case dividend news) is found to be substantial while the remaining news component (in this case return news) is small, then it does not necessarily imply that most of the variability of the stock market is due to dividend news, but could simply be due to the chosen VAR predictor variables not being able to capture the time-varying nature of returns. In Campbell and Ammer's (1993, p.12) words: "When several components of an asset price are variable ... the VAR results must be interpreted more cautiously, as giving a variance decomposition conditional on whatever information is included in the system. *In practice it seems likely that the VAR results will tend to overstate the importance of whichever component is treated as a residual*". (Italics added).[10] Similarly, Chen and Zhao (2009) state in their abstract: "Many studies directly estimate the DR news but back out the CF news as

[10] When only one component of the asset price is time-varying, the dividend-price ratio will summarize all the information that investors have about that component (c.f. Campbell and Shiller (1988) and Campbell and Ammer (1993)), and there is no need to include additional information variables in the system. In that case a zero first row in the $\mathbf{A}$ matrix *must* imply time-varying expected dividend growth, and all return innovation variability will correctly be attributed to dividend news. If both expected dividend growth

a residual. We argue that this approach has a serious limitation because the DR news cannot be accurately measured due to the small predictive power, and the CF news, as a residual, inherits the large misspecification error of the DR news."

Thus, one might think that the results are sensitive to which component is treated as a residual, dividend news or return news. However, Campbell, Polk, and Vuolteenaho (2010, p.322) state that this is not the case: "The approximate identity linking returns, dividends, and stock prices, $r_{t+1} \approx k + \rho p_{t+1} + (1-\rho)d_{t+1} - p_t$, can be rewritten as $r_{t+1} \approx k - \rho(d_{t+1} - p_{t+1}) + (d_t - p_t) + \Delta d_{t+1}$. Thus, a VAR that contains $r_{t+1}, (d_{t+1} - p_{t+1})$, and an arbitrary set of other state variables is equivalent to a VAR that contains $\Delta d_{t+1}, (d_{t+1} - p_{t+1})$, and the same set of other state variables. The two VARs will generate exactly the same news terms. Of course, the news terms are sensitive to the other state variables in the VAR system. Therefore, the important decision in implementing this methodology is not the decision to forecast returns or cash flows, but the choice of variables to include in the VAR".

Taken at face value, the above statement from Campbell, Polk and Vuolteenaho implies that a VAR for e.g. r_{t+1}, $d_{t+1} - p_{t+1}$, and some arbitrary state variable, x_{t+1}, will generate the same news terms as a VAR for Δd_{t+1}, $d_{t+1} - p_{t+1}$, and x_{t+1}. But this is not completely accurate. In order for the two VARs to generate identical news terms, the underlying information sets have to be identical. The approximate identity does *not* link r_{t+1} and $d_{t+1} - p_{t+1}$ in the first system to Δd_{t+1} and $d_{t+1} - p_{t+1}$ in the second system. Consider the system (5). Inserting from the first two equations in (5) into the approximate identity, gives (abstracting from k)

$$\begin{aligned}
\Delta d_{t+1} &= r_{t+1} - (d_t - p_t) + \rho(d_{t+1} - p_{t+1}) \\
&= (a_{11} + \rho a_{21})r_t + (a_{12} + \rho a_{22} - 1)(d_t - p_t) \\
&\quad +(a_{13} + \rho a_{23})x_t + (\varepsilon_{r,t+1} + \rho \varepsilon_{dp,t+1}),
\end{aligned}$$

which can be regarded as one equation in a comprehensive VAR system for Δd_{t+1}, $d_{t+1} - p_{t+1}$, x_{t+1}, *and* r_{t+1}. Thus, in the second system r_t needs to be included as an additional predictor variable and, similarly, in the first system

and expected returns vary over time, a zero first row in the **A** matrix would correspond to a particular correlation structure between the included relevant predictive variables, i.e. $(r_t, d_t - p_t, x_t)$ and the omitted relevant predictive variables.

Δd_t needs to be included as an additional predictor variable, in order for the two systems to generate identical news terms. Note that the comprehensive system for Δd_{t+1}, $d_{t+1} - p_{t+1}$, x_{t+1}, and r_{t+1}, is redundant: knowing the last three equations one can infer the first from the approximate identity and, similarly, knowing the first three equations one can infer the fourth.[11]

The above point is a reminiscence of Cochrane's (2008, p.1540) observation that due to the approximate identity (1), a VAR model for r_{t+1}, Δd_{t+1}, and $d_{t+1} - p_{t+1}$, is redundant in the sense that "one can infer the data, coefficients, and error of any one equation from those of the other two." Cochrane illustrates the redundancy in a restricted VAR model for r_{t+1}, Δd_{t+1}, and $d_{t+1} - p_{t+1}$, in which only $d_t - p_t$ figures as a common predictive variable, i.e. $a_{11} = a_{13} = a_{21} = a_{23} = 0$, and no x_{t+1} variable, in (5). In that special case, because of the identical information set (consisting of only $d_t - p_t$), from the estimated coefficients and residuals in the system for r_{t+1} and $d_{t+1} - p_{t+1}$, the coefficients and residuals in the equation for Δd_{t+1} can be inferred using the approximate identity, and apart from approximation error those coefficients and residuals would be identical to those obtained by estimating the Δd_{t+1} equation directly. Hence, whether one computes the dividend news component directly using the estimated equation for Δd_{t+1}, or indirectly as a residual using (4) and the estimated equations for r_{t+1} and $d_{t+1} - p_{t+1}$, makes no difference. The presence of the dividend-price ratio is crucial in establishing the exact equivalence of these two approaches, but if $d_{t+1} - p_{t+1}$ is replaced by some other ratio involving prices, e.g. the price-earnings ratio, that captures predictability in a similar manner as the dividend-price ratio, the differences between the two approaches will be very minor.

The bottom line is that what is important is not which news component is treated as a residual, but to include in the VAR the dividend-price ratio (or another valuation ratio highly correlated with the dividend-price ratio) and possibly additional state variables that capture the main part of the predictive variability of returns/dividend growth not captured by the dividend-price

[11] John Campbell has pointed out to us that an alternative way to get identical news terms in the two systems (r_{t+1}, $d_{t+1} - p_{t+1}$, x_{t+1}) and (Δd_{t+1}, $d_{t+1} - p_{t+1}$, x_{t+1}) is to include an extra lag of the dividend-price ratio, $d_{t-1} - p_{t-1}$, as a predictor variable. In the first system this implies that r_t, $d_t - p_t$, and $d_{t-1} - p_{t-1}$ appear among the predictor variables, and a linear combination of these three give Δd_t through the aproximate identity so that dividend growth is effectively also part of the information set. Similarly, in the second system Δd_t, $d_t - p_t$, and $d_{t-1} - p_{t-1}$ appear among the predictor variables, and a linear combination of these give r_t through the approximate identity.

ratio. Chen and Zhao (2009) show that in practice the relative importance of the two news components is very sensitive to the choice of predictor variables. This also holds if one disregards the invalid decompositions (c.f. our discussion in section 3.2) in Chen and Zhao's analysis, as seen by comparing columns 1, 3, 4, 5, and 8 in Chen and Zhao's Table 2. Regarding the invalid decompositions in columns 2, 6, and 7 in Chen and Zhao's Table 2, because prices do not appear in the state vector, the 'projection error' in $E_t p_t = p_t$ + 'projection error' becomes large, c.f. section 3.2, and this large 'projection error' then becomes part of the residual news component. This explains why the cashflow news component is relatively much larger in those three columns compared to the other columns.

The confusion in the literature about the role of the residual news component, has led some (see sub-section 3.3.1 below) to suggest to model directly both news components, instead of treating one of them as a residual, by including both returns and dividends directly in the VAR. However, it should be clear from the discussion above that such an approach will not lead to news terms that are different from those obtained in a properly specified system in which one of the terms is backed out residually. And, as we now illustrate, modeling both news components directly may quickly run into yet another pitfall.

In order to obtain both $\eta_{r,t+1}$ and $\eta_{d,t+1}$ directly, augment the VAR model (5) with an equation for Δd_t. Assume that this variable is the second element of the system, while r_t is the first element, as before, and $d_t - p_t$ and x_t are now the third and fourth elements, respectively. Then $\eta_{r,t+1}$ is still computed as in (7) (where an extra 0 is added to the $\mathbf{e1}'$ vector), but now $\eta_{d,t+1}$ is computed as

$$\eta_{d,t+1} = \mathbf{e2}'(\mathbf{I} - \rho\mathbf{A})^{-1}\boldsymbol{\varepsilon}_{t+1}, \tag{11}$$

where $\mathbf{e2}' = (0 \;\; 1 \;\; 0 \;\; 0)$.

If the VAR variables are not able to capture the time-varying properties of returns, i.e. the first row of $\mathbf{A}$ is a zero vector, then all the variability of return innovations will still be attributed to the 'dividend news' component, *also in the case of completely unpredictable dividend growth.* This can be seen from (11) which implies that if the second row of $\mathbf{A}$ is also a zero vector, then $\eta_{d,t+1}$ will *not* be equal to 0; instead, $\eta_{d,t+1} = \mathbf{e2}'\boldsymbol{\varepsilon}_{t+1}$. Thus, in this case return innovations are just equal to dividend growth innovations.

The important thing to notice here is that when the VAR model includes all three variables, r_t, Δd_t, and $d_t - p_t$, i.e. $\mathbf{z}_t = (r_t, \Delta d_t, d_t - p_t, x_t)'$, the sum of $-\eta_{r,t+1}$ and $\eta_{d,t+1}$ will always add up to η_{t+1} (apart from the approximation error), also when none of the news components are obtained as a residual from the identity (4). The reason is as follows: by rewriting (1) into

$$r_{t+1} = (d_t - p_t) - \rho(d_{t+1} - p_{t+1}) + \Delta d_{t+1}, \tag{12}$$

and projecting onto $\mathbf{z}_t$, we get $\mathbf{e1}'\mathbf{A}\mathbf{z}_t = \mathbf{e3}'\mathbf{z}_t - \rho\mathbf{e3}'\mathbf{A}\mathbf{z}_t + \mathbf{e2}'\mathbf{A}\mathbf{z}_t$, where $\mathbf{e3}' = (0 \quad 0 \quad 1 \quad 0)$. Thus, VAR forecasts by construction satisfy (12) [and thereby (1)], and since (4) is derived from (1) it follows that VAR generated innovation and news components from a model that includes all three variables, r_t, Δd_t, and $d_t - p_t$, will automatically satisfy the identity (4), apart from the approximation error. Again, the underlying reason for this result is the redundancy of the VAR model. Knowing two of the three equations for r_{t+1}, Δd_{t+1}, and $d_{t+1} - p_{t+1}$, one can infer the third. Nothing is gained by modeling both returns and dividend growth in a system that also contains the dividend-price ratio.[12]

The redundancy of the VAR model will show up clearly as an econometric problem in models with more than one lag. Assume a second-order system; among the regressor variables would then be r_t, $d_t - p_t$, $d_{t-1} - p_{t-1}$, and Δd_t. But these four variables are linearly related by the approximate identity (12). Thus, there would be almost perfect multicollinearity and the VAR parameter estimates would be meaningless. The only reason the estimation does not completely break down is because of the approximation error in (12).

3.3.1 Two examples of directly modeling cashflow news

In an early application of the VAR based return decomposition, Campbell and Mei (1993, section 2.4) calculate an alternative direct cashflow news measure and compare it with the residually obtained measure. Campbell and Mei

[12] The special case referred to above with the first and second rows of $\mathbf{A}$ being zero vectors is, of course, unrealistic. If there is no bubble, then either r_{t+1} or Δd_{t+1} (or both) will be predictable. In theory, however, with the given choice of state variables, zero vectors in the first two rows of $\mathbf{A}$ could happen with a particular correlation structure between $(r_t, \Delta d_t, d_t - p_t, x_t)$ and the omitted relevant variables. The important point here is that even in this extreme case the sum of $-\eta_{r,t+1}$ and $\eta_{d,t+1}$ will add up to η_{t+1}. Even with a rational bubble in stock prices, the two news components will add up to the innovation component. A bubble implies an explosive log dividend-price ratio with an autoregressive coefficient of $1/\rho > 1$, and a zero $d_t - p_t$ coefficient in the equations for r_{t+1} and Δd_{t+1}, in accordance with Cochrane (2008, section 4.1). See the Appendix for details.

compute a variance decomposition for excess stock returns in terms of news about future dividends, real interest rates, and excess returns, c.f. equation (4) in Campbell and Mei (1993) (which is equivalent to equation 1 in Campbell and Ammer, 1993). Since the decomposition is for *excess* returns, an extra news component, "real interest rate news", appears. Campbell and Mei's benchmark VAR contains excess returns, real interest rates, the dividend-price ratio, and additional predictor variables. From that system, excess return news and real interest rate news are computed directly while dividend news is backed out residually. In section 2.4 in their paper, Campbell and Mei then estimate an equation for cashflows in which dividend growth depends on the state variables from the benchmark VAR. From the dividend growth innovations and VAR innovations a direct cashflow news component can then be computed and compared with the indirectly computed cashflows news component. In their application, Campbell and Mei find that the directly and indirectly computed cashflow news components are very highly correlated and that none of the subsequent results depend in a qualitatively important way on which component is used. Since the dividend-price ratio is a state variable in the VAR, and since the information set in the dividend growth equation is identical to the information set in the VAR, it is not surprising that Campbell and Mei find it unimportant whether dividend news is computed directly or backed out residually. As we have shown above, with identical information sets (that include the dividend-price ratio) it is only the approximation error in the underlying log-linear return approximation that can make the results differ.

Chen and Zhao (2009, section 4.1.1) also propose to model dividend growth directly in order to get a direct estimate of the dividend news component, and in contrast to Campbell and Mei (1993), Chen and Zhao find that their results are highly dependent on whether dividend news is computed directly or indirectly. However, there are two aspects of Chen and Zhao's analysis that complicate the interpretation of their results. First, they adopt a separate VAR system for the dividend growth rate because "the state variables that predict equity returns do not necessarily predict dividend growth rate." (Chen and Zhao, 2009, p.5236). In their footnote 23 they correctly say that "this can be regarded as a single VAR system with equity return and dividend growth rate included, and with parameter restrictions." The VAR for dividend growth includes the dividend-price ratio as a predictor variable. Thus, their two VAR systems are effectively one restricted system with returns, dividend growth, and the dividend-price ratio among the state variables. However, because of the particular zero-restrictions that they impose, the information set in the system used to compute dividend news is *not* identical to the information set used to compute return news. This explains part of the difference in results depending on whether dividend news is modeled directly or indirectly.

Second, as in Campbell and Mei (1993), Chen and Zhao (2009) use *excess* returns instead of just returns (similar to what they do in their bond return example, c.f. section 3.1). However, in contrast to Campbell and Mei, Chen and Zhao neglect the "real interest rate news" component that needs to be added to the return decomposition when excess returns are used instead of returns. This implies that the "residual" component in their equation (23) captures not just "model noise", as indicated by Chen and Zhao, but also "real interest rate news". This provides another reason for why the "residual" news component plays such an important role in Chen and Zhao's analysis. In fact, Campbell and Mei (1993) show that real interest rate variation has a significant impact on variance decompositions of excess stock returns, and thus cannot just be neglected. Had Chen and Zhao included real interest rates in their VAR model, and computed the real interest rate news component, and had they specified the VAR without arbitrary (and untested) zero-restrictions, then the "residual" component would play a very limited role because that component would then only capture approximation error.[13] In any case, Chen and Zhao's (2009, p.5238) claim that their exercise "points to the difficulty of drawing meaningful conclusions using the residual-based return decomposition approach" seems unjustified. What their exercise shows is that in order to draw meaningful conclusions using the residual-based decomposition approach, one needs to specify a proper VAR system and remember to compute all relevant news terms.

3.3.2 An empirical illustration

To illustrate some of the above points, we carry out a small empirical analysis. Table 1 reports variance decompositions for US stock returns. The data are annual real log returns, log real dividend growth, and the log dividend-price ratio, computed from the S&P price index and associated dividends from

[13] Campbell and Vuolteenaho (2004) and Campbell, Polk, and Vuolteenaho (2010) also disregard real interest rates in their decomposition of excess stock returns. However, in the first of these papers the authors explicitly assume that the real interest rate is constant (Campbell and Vuolteenaho, 2004, p.1263).

1871 to 2008 (from Robert Shiller's webpage, www.robertshiller.com). The first three rows in the table show VAR parameter estimates for the three-dimensional system (r_{t+1}, Δd_{t+1}, $d_{t+1}-p_{t+1}$) and associated return innovation variance decompositions computed in two different ways, either by backing out dividend news as in row 1 (i.e. using only the estimates in rows 1 and 3, and then computing $\eta_{d,t+1}$ from equation (4)), or by backing out return news as in row 2 (i.e. using only the estimates in rows 2 and 3, and then computing $\eta_{r,t+1}$ from equation (4)).[14] We see that the decompositions are very similar. In both decompositions the dividend news component accounts for 47% of the variance of return innovations, while the return news component accounts for either 35% or 39%; accordingly, the remaining covariance component accounts for 18% or 14% depending on the decomposition.

The only cause for these slight differences is the approximation error in the underlying log-linear return approximation. To show this, rows 8 to 10 in Table 1 report similar decompositions as in rows 1 to 3, the only difference being that actual log returns, r_{t+1}, are replaced by approximate log returns computed as $r_{t+1}^{approx} = k + \rho p_{t+1} + (1-\rho)d_{t+1} - p_t$ from equation (1), implying no approximation error by construction. As seen, the two decompositions, where either dividend news or return news is backed out residually, are now identical.

In rows 4 to 7 in Table 1 we report decompositions for two different two-dimensional VAR systems: one for r_{t+1} and $d_{t+1} - p_{t+1}$ where dividend news is backed out residually (rows 4 and 5), and one for Δd_{t+1} and $d_{t+1} - p_{t+1}$ where return news is backed out residually (rows 6 and 7). We see that now the differences between the two decompositions are much larger than the differences between the decompositions in rows 1 and 2. This shows that Campbell, Polk and Vuolteenaho (2010) are not entirely correct in stating that a VAR for r_{t+1}, $(d_{t+1}-p_{t+1})$ will generate the same news terms as a VAR for Δd_{t+1}, $(d_{t+1} - p_{t+1})$. The reason is that in addition to the approximation error, the two VAR systems do not have identical information sets because the approximate identity does not link r_{t+1} and $d_{t+1}-p_{t+1}$ in the first system to Δd_{t+1} and $d_{t+1} - p_{t+1}$ in the second system.

The results in Table 1 also show how the decompositions are sensitive

[14] In the system composed of rows 2 and 3 in Table 1, the dividend news component is computed directly, as in equation (11), and the return innovations are computed from the dividend growth innovations and dividend-price ratio innovations as $\varepsilon_{r,t+1} = \varepsilon_{d,t+1} - \rho\varepsilon_{dp,t+1}$, which follows from how the identity (1) links the VAR errors (c.f. Cochrane, 2008). The return news component is then backed out residually from (4).

to which state variables are included in addition to the dividend-price ratio. By comparing the system composed of rows 2 and 3 with the smaller system composed of rows 6 and 7, we see that the decompositions are quite different in the two systems. In the latter system both individual news components are much smaller - and the covariance term is much larger - than in the former system. The reason for these differences is not due to a difference in which component is treated as a residual, because in both systems return news are backed out residually. The reason is that in the first system (rows 2 and 3) r_t figures as a predictor variable which it does not in the second system, and r_t has strong predictive ability for especially Δd_{t+1}, as seen by the coefficient estimate of 0.275 with a standard error of 0.072 and an R^2 value of 25.1% in row 2, compared to an R^2 of only 10.6% in row 6. These results are in accordance with the results reported by Chen and Zhao (2009) in their Table 2.[15]

As we have seen, in a properly specified VAR, return variance decompositions are independent of which news component is treated as a residual, from which it also follows that nothing is gained by modeling both news components directly. The reason is that the approximate identity (1) makes a system that contains both returns, dividend growth, and the dividend-price ratio redundant. In VAR models with more than one lag this redundancy will manifest itself as near-perfect multicollinearity among the regressor variables. On the S&P data, a regression of r_t onto $d_t - p_t$, $d_{t-1} - p_{t-1}$, and Δd_t, gives an R^2 of 0.9995, and if we add an extra lag on the predictor variables in the first three rows in Table 1, the VAR parameter estimates become "wild" and/or have huge standard errors, which is clearly an indication of near-perfect multicollinearity. Naturally, such a system is econometrically meaningless.

[15] The results reported in Table 1 are much more supportive of dividend growth being predictable than the results reported by Cochrane (2008). The reasons for these differences are that we use a longer sample period, we use S&P data instead of CRSP data, and we include additional predictor variables in addition to the dividend-price ratio. Our results imply a much larger dividend news component than Cochrane's results. See also Chen (2009) and Engsted and Pedersen (2010) for recent detailed investigations of return and dividend predictability.

4 Concluding remarks

In this paper we have attempted to explain in detail the often quite subtle limitations and pitfalls involved in the VAR based return variance decompositions pioneered by Campbell (1991) and Campbell and Ammer (1993). This methodology constitutes an important and informative framework to analyze asset market fluctuations, and it has been applied in numerous subsequent studies in several different areas within financial and monetary economics. However, if care is not taken in specifying the underlying VAR system, invalid decompositions will result and the methodology will appear more fragile than it actually is.

A recent example of a paper that in some cases uses improper VAR systems to compute return decompositions, is Chen and Zhao (2009). In the present paper we have shown that because Chen and Zhao are not sufficiently careful in setting up their VAR systems, their claims about the "unreliability" and "counterintuitiveness" of the decomposition approach are unjustified.

Three points are especially important to remember when using the VAR based decomposition methodology. First, if the decomposition is for *excess* returns an extra news component appears in addition to cashflow news and excess return news, namely interest rate news: nominal interest rate news in the case of nominal bonds, and real interest rate news in the case of stocks. Second, in order for the decomposition to be valid the asset price needs to be included as a state variable in the VAR. Third, if the VAR system is properly specified it makes no difference whether cashflow news or discount rate news is backed out residually, and it makes no difference whether both news components are computed directly or one of them is backed out residually.

Used in a proper way, the VAR based return decomposition approach is a very useful and informative method for analyzing the movements and pricing of asset returns.

5 Appendix

In this appendix we derive the parameter restrictions that a rational bubble implies for a VAR model that includes both returns and dividend growth in adition to the dividend-price ratio (c.f. footnote 12). Write the VAR as follows

$$
\begin{bmatrix} r_{t+1} \\ \Delta d_{t+1} \\ d_{t+1} - p_{t+1} \\ x_{t+1} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} r_t \\ \Delta d_t \\ d_t - p_t \\ x_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{r,t+1} \\ \varepsilon_{d,t+1} \\ \varepsilon_{dp,t+1} \\ \varepsilon_{x,t+1} \end{bmatrix},
$$

or, in compact form, $\mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t + \boldsymbol{\varepsilon}_{t+1}$. By defining the three selection vectors, $\mathbf{e1} = (1 \;\; 0 \;\; 0 \;\; 0)'$, $\mathbf{e2} = (0 \;\; 1 \;\; 0 \;\; 0)'$, and $\mathbf{e3} = (0 \;\; 0 \;\; 1 \;\; 0)'$, and projecting onto $\mathbf{z}_t$, we get from equation (12):

$$
\begin{aligned}
\mathbf{e1}'\mathbf{A}\mathbf{z}_t &= \mathbf{e3}'\mathbf{z}_t - \rho\mathbf{e3}'\mathbf{A}\mathbf{z}_t + \mathbf{e2}'\mathbf{A}\mathbf{z}_t \quad \Rightarrow \\
\mathbf{e1}'\mathbf{A} &= \mathbf{e3}' - \rho\mathbf{e3}'\mathbf{A} + \mathbf{e2}'\mathbf{A}.
\end{aligned} \tag{A1}
$$

r_t, Δd_t, and x_t are stationary variables, but with a rational bubble $d_t - p_t$ is explosive such that $a_{13} = a_{23} = a_{31} = a_{32} = a_{34} = a_{43} = 0$. Thus, from (A1) we get the following additional parameter restrictions

$$
\begin{aligned}
a_{11} &= a_{21} \\
a_{12} &= a_{22} \\
a_{33} &= \frac{1}{\rho} > 1 \\
a_{14} &= a_{24}
\end{aligned}
$$

Note that with a bubble the explosive autoregressive coefficient for $d_t - p_t$ is closely connected to the linearization parameter $\rho < 1$, in accordance with Cochrane (2008, section 4.1). Strictly speaking, under a bubble the approximation error in (12) is unbounded because the log dividend-price ratio is explosive. However, in the literature bubbles are often discussed within the log-linear framework, e.g. Cochrane (2008, section 4.1). Engsted, Pedersen, and Tanggaard (2010) analyze in more detail the properties of the log-linear approximation in the presence of bubbles.

6 References

Ammer, J. and J. Mei (1996): Measuring international economic linkages with stock market data. *Journal of Finance* 51, 1743-1763.

Bernanke, B.S. and K.N. Kuttner (2005): What explains the stock market's reaction to Federal Reserve policy? *Journal of Finance* 60, 1221-1257.

Campbell, J.Y. (1991): A variance decomposition for stock returns. *Economic Journal* 101, 157-179.

Campbell, J.Y. and J. Ammer (1993): What moves the stock and bond markets? A variance decomposition for long-term asset returns. *Journal of Finance* 48, 3-37.

Campbell, J.Y., A.W. Lo, and A.C. MacKinlay (1997): *The Econometrics of Financial Markets*. Princeton University Press.

Campbell, J.Y. and J. Mei (1993): Where do betas come from? Asset price dynamics and the sources of systematic risk. *Review of Financial Studies* 6, 567-592.

Campbell, J.Y., C. Polk, and T. Vuolteenaho (2010): Growth or glamour? Fundamentals and systematic risk in stock returns. *Review of Financial Studies* 23, 305-344.

Campbell, J.Y. and R.J. Shiller (1988): The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1, 195-228.

Campbell, J.Y., and T. Vuolteenaho (2004): Bad beta, good beta. *American Economic Review* 94, 1249-1275.

Chen, L. (2009): On the reversal of return and dividend predictability: A tale of two periods. *Journal of Financial Economics* 92, 128-151.

Chen, L., and S. Zhao (2009): Return decomposition. *Review of Financial Studies* 22, 5213-5249.

Cochrane, J.H. (2008): The dog that did not bark: A defense of return predictability. *Review of Financial Studies* 21, 1533-1575.

Engsted, T., and T.Q. Pedersen (2010): The dividend-price ratio does predict dividend growth: International evidence. *Journal of Empirical Finance* (forthcoming).

Engsted, T., T.Q. Pedersen, and C. Tanggaard (2010): The log-linear return approximation with and without bubbles. *Work in progress*, University of Aarhus.

Engsted, T. and C. Tanggaard (2001): The Danish stock and bond markets: Comovement, return predictability and variance decomposition. *Journal of Empirical Finance* 8, 243-271.

Engsted, T. and C. Tanggaard (2004): The comovement of US and UK stock markets. *European Financial Management* 10, 593-607.

Patelis, A.D. (1997): Stock return predictability and the role of the monetary sector. *Journal of Finance* 52, 1951-1972.

		r_t	Δd_t	$d_t - p_t$	R^2	$\frac{\mathrm{Var}(\eta_r)}{\mathrm{Var}(\eta)}$	$\frac{\mathrm{Var}(\eta_d)}{\mathrm{Var}(\eta)}$	$\frac{\text{-2Cov}(\eta_r,\eta_d)}{\mathrm{Var}(\eta)}$
1	r_{t+1}	0.109	-0.217	0.078	0.057	0.352	0.467	0.181
		(0.100)	(0.139)	(0.037)				
2	Δd_{t+1}	0.275	0.091	-0.055	0.251	0.394	0.466	0.140
		(0.072)	(0.085)	(0.021)				
3	$d_{t+1} - p_{t+1}$	0.166	0.327	0.895	0.773			
		(0.116)	(0.177)	(0.042)				
4	r_{t+1}	0.098		0.080	0.036	0.357	0.498	0.146
		(0.099)		(0.037)				
5	$d_{t+1} - p_{t+1}$	0.182		0.895	0.773			
		(0.114)		(0.042)				
6	Δd_{t+1}		0.117	-0.084	0.106	0.275	0.384	0.342
			(0.091)	(0.029)				
7	$d_{t+1} - p_{t+1}$		0.343	0.880	0.777			
			(0.180)	(0.043)				

		r_t^{approx}	Δd_t	$d_t - p_t$	R^2	$\frac{\mathrm{Var}(\eta_r)}{\mathrm{Var}(\eta)}$	$\frac{\mathrm{Var}(\eta_d)}{\mathrm{Var}(\eta)}$	$\frac{\text{-2Cov}(\eta_r,\eta_d)}{\mathrm{Var}(\eta)}$
8	r_{t+1}^{approx}	0.114	-0.220	0.084	0.062	0.383	0.462	0.155
		(0.099)	(0.141)	(0.037)				
9	Δd_{t+1}	0.276	0.083	-0.057	0.260	0.383	0.462	0.155
		(0.071)	(0.084)	(0.021)				
10	$d_{t+1} - p_{t+1}$	0.168	0.315	0.894	0.780			
		(0.113)	(0.177)	(0.042)				

Notes: The table reports VAR parameter estimates (with heteroscedasticity-consistent standard errors in parentheses) and associated return variance decompositions. η, η_r, and η_d are return innovation, return news, and dividend news, respectively. r_{t+1}^{approx} is calculated as $r_{t+1}^{approx} = k + \rho p_{t+1} + (1-\rho)d_{t+1} - p_t$, with $\rho = 0.96$ (obtained as $(1 + e^{\overline{d-p}})^{-1}$). The data are annual real S&P data from 1871-2008 (from Robert Shiller's webpage).

Table 1: Variance decompositions for US stock returns

Research Papers 2010

CREATES
Center for Research in Econometric Analysis of Time Series

2009-55:	Michael Jansson and Morten Ørregaard Nielsen: Nearly Efficient Likelihood Ratio Tests for Seasonal Unit Roots
2009-56:	Valeri Voev: On the Economic Evaluation of Volatility Forecasts
2009-57:	Jesper Rangvid, Maik Schmeling and Andreas Schrimpf: Global Asset Pricing: Is There a Role for Long-run Consumption Risk?
2009-58:	Olaf Posch: Risk premia in general equilibrium
2009-59:	Christian M. Dahl and Emma M. Iglesias: Modelling the Volatility-Return Trade-off when Volatility may be Nonstationary
2009-60:	Ole E. Barndorff-Nielsen, José Manuel Corcuera and Mark Podolskij: Limit theorems for functionals of higher order differences of Brownian semi-stationary processes
2010-01:	Anders Bredahl Kock and Timo Teräsvirta: Forecasting with nonlinear time series models
2010-02:	Gunnar Bårdsen, Stan Hurn and Zoë McHugh: Asymmetric unemployment rate dynamics in Australia
2010-03:	Jesper Rangvid, Maik Schmeling and Andreas Schrimpf: Cash Flow-Predictability: Still Going Strong
2010-04:	Helle Bunzel and Walter Enders: The Taylor Rule and "Opportunistic" Monetary Policy
2010-05:	Martin M. Andreasen: Non-linear DSGE Models and The Optimized Particle Filter
2010-06:	Søren Johansen and Bent Nielsen: Discussion of The Forward Search: Theory and Data Analysis by Anthony C. Atkinson, Marco Riani, and Andrea Ceroli
2010-07:	Giuseppe Cavaliere, Anders Rahbek and A.M.Robert Taylor: Bootstrap Sequential Determination of the Co-integration Rank in VAR Models
2010-08:	Peter R. Hansen and Asger Lunde: Estimating the Persistence and the Autocorrelation Function of a Time Series that is Measured with Error
2010-09:	Tom Engsted, Thomas Q. Pedersen and Carsten Tanggaard: Pitfalls in VAR based return decompositions: A clarification